NFT for Eternity

Hadar Y. Jabotinsky
Hadar Jabotinsky Center for Interdisciplinary Research of Financial Markets, Crises and Technology; School of Law, Zefat Academic College

Michal Lavi
Hadar Jabotinsky Center for Interdisciplinary Research of Financial Markets, Crises and Technology; School of Law, Zefat Academic College

Follow this and additional works at: https://repository.law.umich.edu/mjlr

Part of the Computer Law Commons, First Amendment Commons, and the Internet Law Commons

Recommended Citation
Available at: https://repository.law.umich.edu/mjlr/vol56/iss3/5

https://doi.org/10.36646/mjlr.56.3.NFT

This Article is brought to you for free and open access by the University of Michigan Journal of Law Reform at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in University of Michigan Journal of Law Reform by an authorized editor of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.
ABSTRACT

Non-fungible tokens (NFTs) are unique tokens stored on a digital ledger – the blockchain. They are meant to represent unique, non-interchangeable digital assets, as there is only one token with that exact data. Moreover, the information attached to the token cannot be altered as on a regular database. While copies of these digital items are available to all, NFTs are tracked on blockchains to provide the owner with proof of ownership. This possibility of buying and owning digital assets can be attractive to many individuals.

NFTs are presently at the stage of early adoption and their uses are expanding. In the future, they could become a fundamental and integral component of tomorrow’s web. NFTs bear the potential to become the engine of speech: as tokenized expressions cannot be altered or deleted, they enable complete freedom of expression, which is not subject to censorship. However, tokenized speech can also bear significant costs and risks, which can threaten individual dignity and the public interest. Anyone can tokenize a defamatory tweet, a shaming tweet, or a tweet that includes personal identifying information and these tokenized expressions can never be deleted or removed from the blockchain, risking permanent damage to the reputations of those involved. Even worse, anyone can tokenize extremist political views, such as alt-right incitement, which could ultimately result in violence against minorities, and infringe on the public interest.

To date, literature has focused on harmful speech that appears on dominant digital platforms, but has yet to explore and address the benefits, challenges and risks of tokenized speech. Such speech cannot be deleted from the web in the same way traditional internet intermediaries currently remove content. Thus, the potential influence of NFTs on freedom of expression remains unclear. This Article strives to fill the gap and contribute to literature in several ways. It introduces the idea of owning digital assets by using NFT technology, surveys the main uses of tokenizing digital
assets and the benefits of such practices. It aims to raise awareness of the potential of tokenized speech to circumvent censorship and to act as the engine of freedom of expression. Yet it also addresses the challenges and risks posed by tokenized speech. Finally, it proposes various solutions and remedies for the abuse of NFT technology, which may have the potential to perpetuate harmful speech. As we are well aware of the challenges inherent in our proposals for mitigation, this Article also addresses First Amendment objections to the proposed solution.

TABLE OF CONTENTS

INTRODUCTION................................................................. 830

I. NFT TOKENS: CREATING UNIQUE RIVAL ASSETS............... 841
   A. What Are NFTs? How Do They Work?.......................... 841
   B. This is Mine! Main Uses of NFT Tokens......................... 844
      1. Crypto Art and Historical Documentation of Inventions........ 844
      2. Collectible Game Cards and Virtual Goods.............. 847
      3. Markets for Physical Items/ Distribution of Resources and Ticket Brokers 850
      4. The Music and Film Industries – Ownership and Rights Management........... 851
   C. The Future: Tokenizing Speech - Sharing NFTs as an Integral Part of the Internet......................... 853

II. PUBLIC - PRIVATE AND PRIVATE ONLINE CENSORSHIP AND THE NFT ENGINE OF SPEECH ........................................ 854
   A. Motivations for Removing Content ................................ 855
      1. Removal of Harmful Content to Comply with the Law........ 855
      2. Public- Private Cooperation in Content Removal........... 859
      3. Removal of Content that Does not Comply with the Terms of Service or the Intermediary's Agenda .... 860
         (a) Categorical Classification that Overlooks Nuances of Context .................. 862
         (b) Removing Content That Is Inconsistent with the Intermediary's Agenda ......... 866
      4. False Positives................................................... 867
   B. CDA Section 230’s Immunity – No Intermediary Liability for Content Removal, No Liability for Failing to Remove Content ... 869
   C. A World Without Censorship – NFTs as The Engine of Speech .. 871
III. THE FLIP SIDE OF NFTS: PERMANENCY OF ABUSIVE SPEECH ......872
   A. Tokenized Speech Abused – A Roadmap of Aggravated Harm ... 875
      1. Harm to Reputation .................................................. 875
         a. Defamation ..................................................... 875
         b. Shaming ....................................................... 876
      2. Harm to Public Interest – The Case of Fake News.... 878
      3. Privacy Harm.......................................................... 880
         a. The Problem of Doxing- Violence, Economic
            Harm and Emotional Distress ......................... 880
         b. Sexual Privacy and Intimacy......................... 881
      4. Extremist Speech that Encourages Physical
         Violence, arm to Autonomy and Democracy ........ 885
         a. Hate Speech .................................................. 885
         b. Incitement to Terrorism ......................... 886

IV. TOWARDS A COMPREHENSIVE FRAMEWORK OF SOLUTIONS AND
    REMEDIES FOR ABUSIVE TOKENIZED SPEECH ...................... 889
   A. Safety-by- Design - Ex Ante Prevention ..................... 889
      1. Minimizing What is Stored on the Blockchain .... 891
      2. Prescreening Tokens and Preventing NFTs
         Containing Unprotected Speech from Entering
         the Blockchain ................................................... 889
   B. Ex Post Mitigation- Obscuring Harmful NFTs ............. 901
      1. “Superseding” Old, Inaccurate Data with New Data
         on the Blockchain ................................................ 901
      2. Harnessing Online Intermediaries to Obscure Abusive
         Tokens on the Level of The World Wide Web ........ 903
   C. Prosecution or Lawsuits Against Abusive Token Owners –
      Remedies ............................................................. 907
      1. Difficulties and Obstacles to Direct Lawsuits ......... 908
         a. Harmful Expressions Do Not Necessary Violate
            the Law ..................................................... 908
         b. The Problem of Recourse, Reluctance to File Legal
            Action and Collection of Monetary Remedies ....... 909
         c. The Problem of Anonymity and the Procedural
            Difficulties of Unmasking the Violator ............ 910

CONCLUSION ........................................................................ 914
INTRODUCTION

On March 5, 2021, about five years after tweeting his first tweet on Twitter, former Twitter CEO Jack Dorsey put the tweet up for sale as an NFT, a non-fungible token.1 The tweet, which stated: “just setting up my twttr”, was offered on a platform called “Valuables” and was sold for about USD 2.9 Million to Mr. Sina Estavī.2 The buyer compared the tweet to the “Mona Lisa” and believed it was a wise investment.3

Shortly afterwards, on March 25, a New York Times column on NFTs written by Kevin Roose was converted into an NFT4 and auctioned on a marketplace named Foundation,5 with all proceeds going to the New York Times Neediest Cases Fund. The profile of the auction’s winner that bought the NFT containing the column “was linked to a Twitter profile belonging to a Dubai-based music production company, and to an Instagram account identified as that of Farzin Fardin Fard”.6 It is however not clear if the winner is Mr. Fard or some other individual or multiple people. The winner bought the column for the sum of USD 560,000 and added it to his NFT collection.7

Non-fungible tokens (NFTs) are unique tokens commonly produced using the ERC-721 standard, a free open token standard that describes how to build non-fungible tokens.8 The tokens allow their buyers to become owners of scarce and unique non-interchangeable digital assets, as the tokens link to a unique asset, that cannot be duplicated or replaced.9 The uniqueness of the NFT is inherent in the fact that no two

2. VALUABLES BY CENT, https://v.cent.co/ (last visited Mar. 28, 2023) [https://perma.cc/j3LSL-QRXe].
4. See Kevin Roose, Buy This Column on the Blockchain!, N.Y. TIMES (Jun. 30, 2021), https://nyti.ms/3ehoMWH [https://perma.cc/SE7E-RLMQ].
7. See id.
tokens have the exact same data. Tokens containing the digital assets are stored on a technological digital ledger – the blockchain – and can be transferred within it. The blockchain connects users to each other through a chain of digital blocks. A public blockchain is usually maintained by its peers and is therefore considered a distributed ledger without a central authority. The programs that create and transfer NFTs are called smart contracts. These smart contracts are an additional layer of technology, basically comprised of command instructions such as “if X happens, execute Y”. As such, these contracts are automatically executable. The blockchain thereby enables the transfer of value or assets without the help of intermediaries. It is, in its core, a system based on trust: in the technology of the blockchain and the peer-to-peer network that allows secured cryptographically record keeping, guaranteeing authenticity and uniqueness of assets. NFTs offer full, real, digital ownership over the assets attached to the tokens. As the asset is attached to a token, when the asset is sold, the token it is attached to is transferred on the distributed ledger technology (“DLT”) and the transaction is recorded in a transparent manner. This is a way to ensure that all users of the DLT can keep track of registered transactions and to provide the owner with proof of ownership. NFTs are ba-
sically a way to allow the exchange of digital assets in return for more tangible ownership.

The benefits offered by blockchain technology make it possible to reduce the costs, risks, constraints, and fraud associated with traditional trading systems. This technology can also help improve security, reliability of originality verification, and traceability of digital assets.\(^\text{18}\) Given that “NFT” was the Collins English Dictionary’s 2021’s word of the year,\(^\text{19}\) and that we hear this buzzword over and over again, it stands to reason that NFTs are worth noticing. This relatively new technology enables engagements that were previously impossible, disrupts traditional business models, and can be the impetus for new business models and markets, such as the market for digital artwork, collectibles, and “emotional assets” or “investments of passion.”\(^\text{20}\) In this way, NFTs can promote efficiency and innovation, as they offer entrepreneurs and artists an additional means of financial reward for their digital work.\(^\text{21}\)

As our Article will further demonstrate,\(^\text{22}\) NFT’s can have many uses. These tokens, however, are primarily used in the industries of art, music, and gaming.\(^\text{23}\) For example, perhaps the most famous NFT sale occurred when Mike Winkelmann — the digital artist known as Beeple — put an NFT digital art photo titled “Everydays: The First 5000 Days” up

---


\(^\text{19}\) See David Shariatmadari, Get Your Crypto at the Ready: NFTs are Big in 2021, COLLINS LANGUAGE LOVERS BLOG (Nov. 24, 2021), https://blog.collinsdictionary.com/language-lovers/get-your-crypto-at-the-ready-nfts-are-big-in-2021/ [perm.cc/CBB5-HV6H].


\(^\text{22}\) See discussion infra Part I.2.

\(^\text{23}\) See Fairfield, supra note 8, at 1273; see also Brian Elzweig & Lawrence J. Trautman, When Does a Non-Fungible Token (NFT) Become a Security?, GA. ST. L. REV. 295, 295 (2023) (explaining that there are many uses of NFTs such as arts, sports and securities, and the preferred way of regulation should depend on its use since “it is the particular use of a given NFT that will determine its appropriate regulatory regime, since it may take the form of a collectible, data associated with a physical item, financial instrument, or a permanent record associated with a person, such as marriage license, or property deed”).
for auction and sold it for USD 69 Million.\textsuperscript{24} One might wonder – why would anyone pay such a large sum for a photo which can easily be downloaded by anyone in a Google search? The answer relates to human psychology – people have an urge to own things and derive some sort of mental benefit from their possessions. In fact, this even plays a role in defining identity.\textsuperscript{25} Studies show that possession not only contributes to the sense of self but is also perceived as an extension of the self.\textsuperscript{26} Thus, buyers of NFTs know that they have become the owner of an authentic asset (as an NFT cannot be forged) and in turn are rewarded by the satisfaction they get from owning an original.\textsuperscript{27} Such ownership also helps the owner gain endorsement of the author and social recognition.\textsuperscript{28}

NFTs are also starting to gain popularity in solving practical legal issues such as copyright. Registering copyrights on the blockchain helps ensure global protection for the copyrighted item on a system which is both transparent and open.\textsuperscript{29} Moreover, NFTs allow efficient management of rights in the music industry as they enable a Tokenized Music License (“TML”) and can be an efficient tool for distributing royalties for use of musical works or other copyrighted items.\textsuperscript{30} For example, a new blockchain music streaming platform named ROCKI, recently launched “royalty income right music” NFTs.\textsuperscript{31} Soon, other streaming platforms


\textsuperscript{26} See Russell W. Belk, Possessions and the Extended Self, 15 J. CONSUMER RSCH. 139 (1988).

\textsuperscript{27} See Leena Kim, WTF Is an NFT—and Why Should We Care?, TOWN & COUNTRY (Mar. 25, 2021), [https://bit.ly/3uxPinn [perma.cc/RAE2-zK7P]].

\textsuperscript{28} See Brian L. Frye, After Copyright: Punning NFTs in a Clout Economy, 45 COLUM. J. L. & ARTS 341, 348–49 (2022).

\textsuperscript{29} Felipe Erazo, Italian Copyright Agency Selects Algorand to Create Over Four Million NFTs to Represent Author Rights, BITCOIN.COM (Mar. 26, 2021), [https://news.bitcoin.com/italian-copyright-agency-selects-algorand-to-create-over-four-million-nfts-to-represent-author-rights/ [perma.cc/LN8X-7GLU]].


\textsuperscript{31} Rocki, ROCKI, the New Music Streaming Platform on the Blockchain Just Launched Their First Royalty Income Right Music NFT (Nonfungible Token) for a Song, Which Sold for a Record $40 ETH ($244,800 USD at the Time of Writing), CISION (Dec. 16, 2022), [https://prn.to/seyD1fk [https://perma.cc/SI93-BHSF]].
followed suit, as they discovered the potential of NFTs in royalty collection. 32

The gaming industry is another prominent NFT user. Cryptokitties is one example. 33 This is a blockchain-based game played on the Ethereum blockchain. Players can buy and sell pictures of cats tied to a particular NFT token. Each cat has unique features and a unique digital identity. 34 The game allows users to “breed, trade, and play with the virtual cats in a rich online community.” 35

NFTs are the driving force behind a new wave of crypto adoption. 36 People who were never interested in cryptocurrencies as a financial asset discovered a sudden interest in NFTs because they are interested in the assets attached to them. 37 Just like any new technology, one can expect the use of NFTs to expand beyond early adopters. 38 Given past patterns of technological adoption, the late majority and the laggers can be expected to quickly follow suit, 39 and NFT use is likely to thrive.

Wide adoption might happen sooner than we think. This is because blockchain-based social media platforms and blockchain-based infrastructures are developing as well. 40 One such example is the blockchain-based social media platform known as “Steemit.” 41 This platform is operated by a decentralized community in which users create content. The platform rewards creators of popular posts with its own generated cryptocurrencies. 42 If more mainstream platforms such as Facebook

35. See Mofokeng & Matima, supra note 20, at 14.
36. See Cooper Turley, If You Haven’t Followed NFTs, Here’s Why You Should Start, TECHCRUNCH (Feb. 27, 2021), https://tcrn.ch/2RKqOQ [https://perma.cc/TWN7-9Y4H].
37. Peter Carstairs & Lucy Sanderson, Non-Fungible Tokens and Digital Assets – What’s the Deal?, MINTER ELLISON (Oct. 6, 2021), https://www.minterellison.com/articles/non-fungible-tokens-and-digital-art-whats-the-deal [https://perma.cc/KWV8-MELS] (“Because NFTs are non-fungible, and because they are stored in a blockchain (more on this below), this has enabled an entirely new market in digital artworks to be created whereby digital works can be readily bought and sold.”).
39. Currently the early adopters already use NFTs and we are at the early majority stage of adoption. See ul Hassan et. al. supra note 14, at 13.
and Twitter follow, most of us would come to encounter and use decentralized blockchain platforms on a daily basis. Moreover, NFTs are expected to be integrated with the internet via API (“Application Programming Interface”). This set of definitions and protocols allows computers, products, or services to connect and communicate with other products and services. The integration of NFTs with the internet is expected to accelerate as programmers develop more APIs that can be connected and comprehended by smart contracts on a blockchain. For example, in the future, ticketing companies might create NFT concert tickets. Buying such a ticket would allow the owner to keep the digital ticket as a souvenir, add an attestation on social media to connect with other attendees, and of course sell it in the future if its value increases as a collectible. DLT-based systems are likely to become a substantial component of the architecture of tomorrow’s Web 3.0, a decentralized blockchain based world wide web free of intermediation, “leaping “forward to open, trustless and permissionless networks”. If this indeed occurs, it would bring NFT use to the forefront of the digital stage.

NFTs open up a new world for legal literature and thought as well. Yet there is an absence of core research on the relevant legal issues pertaining to NFTs. Most studies on NFTs focus on smart contracts, property rights, and ownership. The literature currently neglects to address the expressive values of NFTs and the unique free speech implications of these tokens. NFTs are different from other types of speech published on internet websites and social networks. Both the NFT and the transactions associated with it are freely available on the

---


45. Ul Hassan et. al, supra note 14, at 11 (explaining that API can be programmed in a way that could allow blockchain infrastructure to “talk” with the internet infrastructure).


47. See Bobby Allyn, People are Talking about Webs. Is it the Internet of the Future or just a Buzzword? NPR (Nov. 21, 2021), https://www.npr.org/2021/11/21/105698346/webs-internet-jargon-or-future-vision [https://perma.cc/9yNC-SZAT].


digital platform for anyone to see, but the information attached to the token cannot be altered like in a regular database or like on an intermediated internet platform. This is because the blockchain has no single central authority which manages the system. The blockchain's distributed ledger system is jointly administered by a network of communication endpoints, known as “nodes”, for transmission or redistribution of data. The connectedness of the entire chain of blocks protects the blockchain and the tokens it transfers. The permanency of the information on NFT tokens, the inability to remove it from the public blockchain, and the fact that the blockchain never forgets, all substantially limit the ability of governments and private actors to censor speech. Off the blockchain, currently most private actors respond to government regulations and guidelines and cooperate with their takedown requests. In addition, most social media platforms also respond to takedown requests by private individuals, by removing legitimate content as a result of collateral censorship. Thus, they silence opposing opinions or expressions that are unsuitable to the agendas of governments and private individuals. Another aspect of removing content relates to content that does not comply with the platform’s terms of service. Yet sometimes platforms even remove content in what seems to be an arbitrary move. NFTs restrict the power of state actors and private intermediates to remove content. Thus, they can help foster a vibrant

52. Fairfield, supra note 8, at 1270.
53. See Zeynep Tufekci, Twitter and Tear Gas: The Power and Fragility of Networked Protest Ch. 2 (2017) (addressing governmental censorship).
56. Such censorship happens collateral censorship "when a (private) intermediary suppresses the speech of others in order to avoid liability" for such speech. See Felix T. Wu, Collateral Censorship and the Limits of Intermediary Immunity, 87 NOTRE DAME L. REV. 293, 295–96 (2011).
57. See id. at 323–24; infra II.A.3 (as the article explains later this happens because of categorical classification that overlooks nuances of context).
marketplace of ideas as they allow the free flow of information. In other words, NFTs have the potential to become the *engine of speech*.

The permanency of tokenized speech, however, jeopardizes efforts to combat harmful speech. If, for example, John Doe tokenizes a defamatory post on Jane Doe, or tokenizes a tweet with private information about her, and then becomes the owner of the tweet, it would be visible to all blockchain users. Such private information can include her home address, phone number, social security number, credit card, bank account numbers and more. Doxing can also expose her to physical danger, stalking, and on and offline harassment. Due to the decentralized structure of the blockchain and the lack of dependency on a central server, Jane would not be able to report it to a central intermediary and request to take it down. As mentioned, the structure of the blockchain system does not allow alteration of the information published on it. 58 A tokenized tweet could also contain information that can shame Jane in public and harm her dignity and reputation.

The potential harm of tokenized speech extends beyond personal reputations and can encroach on the public interest. Firstly, such permanent, tokenized speech is likely to infringe on the dignity of minorities and other marginalized populations. As such populations suffer from a higher-than-average degree of shaming, perpetuating speech that shames these groups could increase socio-economic gaps. 59 Secondly, it could cause data pollution that would contaminate the flow of accurate information to the general public and even erode democracy, especially when a false rumor on a public official is tokenized. Worst of all, when incitement or hate speech is tokenized, it could even result in violence outside the digital sphere.

In addition, even though the possibility to keep data permanent and visible on the blockchain can mitigate governmental and private censorship, those very same features of permanency and visibility can backfire. For example, they might enable oppression by totalitarian governments, which might track and collect tokenized information on protestors and sentence them to prison. 61

---

58. See Tatar at al., *supra note 51*, at 3.
60. See Omri Ben-Shahar, *Data Pollution*, 11 J. Legal Analysis 104, 105, 112–13 (2019) (treating “fake news” as “data pollution” that disrupts social institutions and public interests in a similar manner to environmental pollution).
61. It should be noted that even in democracies, governments might use records collected on opponents to blackmail them. See Neil Richards, *Why Privacy Matters* 132–133 (2021) (describing how the U.S. NSA wanted to surveil “radicalizers” who are not terrorists but merely radical critics of U.S policy. Richards explains that this practice “raises troubling questions about the govern-
As NFTs become more integrated with social media platforms, tokenized speech is expected to gain more visibility and attract more attention.\textsuperscript{62} As a result, tokenized speech can exacerbate harm. Moreover, such tokens allow eternal memory of the tokenized speech, which in theory could live forever. In other words, tokenized harmful speech could be shackled to unwitting victims, and could have a long-term impact on their ability to live in peace.

Tokenized NFT speech changes the balance between free speech and the individual rights to dignity, reputation and privacy. This leaves a lacuna which the law has yet to address. Currently, tokenized speech remains forever on the blockchain, without any ability to remove it, and thus victims are likely to be left without any remedy. This Article aims to fill this void by proposing various solutions that would be available to victims in order to mitigate the harm caused by tokenized speech. It proposes \textit{ex ante} solutions that should be adopted by NFT marketplaces and intermediaries that allow integration of NFTs in their infrastructure, in order to decrease visibility of the harmful token and mitigate any damage. It also proposes \textit{ex post} solutions by developing remedies through direct legal action against the owner of the harmful tokenized speech, \textit{inter alia} in order to minimize the profitability of owning tokenized harmful speech. Keeping these goals in mind, the Article is divided into the following parts:

Part I demonstrates the new possibility to fully own digital property. It explains what non-fungible tokens (NFTs) are and surveys their features. It illustrates the distinctions between NFTs and fungible tokens, as each type of token possesses a different value. It also elucidates the benefits of the decentralized blockchain-based system in providing transparency and protection against fraud. Subsequently, it provides an overview of the primary applications of NFTs in art, music, collectibles, gaming, and other industries,\textsuperscript{63}—demonstrating that NFTs provide a way of achieving the previously impossible, thus allowing for innovation and growth.\textsuperscript{64}

Part II focuses on online content moderation. It explains that on the internet intermediaries control the flow of information. “While it seems as if everyone ‘can publish freely and instantly online,’ many intern-

---

\textsuperscript{62} See Sul Hassan et. al, supra note 14 at 11.

\textsuperscript{63} Fairfield, supra note 8, at \textit{78}1273.

\textsuperscript{64} See MATT HOUGAN \\ 
termediaries in fact ‘actively curate the content’ that their users post on their platforms”. 65 As private entities, intermediaries are not considered public forums and thus are not subject to judicial scrutiny under the First Amendment. 66 Section 230(c) of the Communications Decency Act (“CDA”) titled “protection for Good Samaritan private blocking and screening of offensive material” encourages online intermediaries to use their editorial discretion and grants them immunity from liability for editorial decisions. 67 Intermediaries do indeed remove harmful speech from their platforms. However, in some cases they may also remove legitimate content because it is not compatible with their agendas, or due to algorithmic misclassification, 68 or a the content was flagged as “false positive” for harmful content. 69 Moreover, they can remove content in cooperation with government requests 70 or even due to requirements by authoritarian governments. 71 Subsequently, the Article puts the spotlight on one prominent NFT feature—the fact that the blockchain never forgets, and the information carried by a token can

67. Communication Decency Act, 47 U.S.C. § 230(c); DANIELLE KEATS CITRON, THE FIGHT FOR PRIVACY, PROTECTING DIGNITY, IDENTITY AND LOVE IN THE DIGITAL AGE 86 (2022) (explaining that Section 230(c) of the CDA originally aimed “to incentivize private efforts aimed at combating ‘offensive’ material.” Section 230(c) (1) addresses immunity for under removal of content and Section 230(c) (2) conversely addresses the immunity for over removal of content by intermediaries.) [hereinafter CITRON, THE FIGHT FOR PRIVACY].
70. Shadmy & Shany, supra note 55.
71. THE NATL. BUREAU OF ASIAN RSCH, SOCIAL MEDIA PLATFORMS AND AUTHORITARIAN CENSORSHIP IN ASIA (Dec. 23, 2020) https://bit.ly/34aFAMn [https://perma.cc/GB3C-87R7] (“Authoritarian governments absolutely do try to find ways to enforce censorship on platforms, with or without the witting help of the platforms themselves... Platforms have an incentive to avoid challenging local authorities or offending cultural sensitivities, even when that leads to suppression of political speech.”).
never be deleted or altered. The fact that NFTs cannot be altered is a

game changer for speech and offers great promise to the area of free-

dom of expression, as the unique features of NFTs can remove barriers
to the flow of information caused by private censorship. In other words,
NFTs can be an engine of speech.

Part III focuses on the flip side of tokenized speech. It explains that
the very same features that make NFTs an engine of speech can also in-
crease the potential of harmful tokenized speech to cause harm. In con-
trast to regular user-generated content in an intermediated system,
where intermediaries moderate through moderators or algorithmic en-
forcement, the blockchain is truly decentralized and has no central in-
termediaries. Moreover, the features of NFTs— their place of trade on
the blockchain, combined with the blockchain’s eternal memory - limit
enforcement options against tokenizing harmful expressions. Subse-
quently, this part maps out the main types of harmful speech that can
be tokenized and the types of harm that can be caused, such as digni-
tary harm, infringement of the public interest, and even violence and
physical harm. It also surveys the types of remedies that are available
to, and applied by, online intermediaries, which are not available on the
blockchain. This roadmap illustrates how NFTs can exacerbate the se-
verity of damage caused by various types of harmful content and takes
the first step towards developing a framework for mitigating this harm.

Part IV proposes solutions to mitigate the problem of misuse of
NFTs to perpetuate harmful expression. The first type of solution is
aimed at NFT marketplaces and proposes that these marketplaces
should adopt a safety-by-design concept. In other words, it proposes a
 technological design that would mitigate the damage caused by tokens
that contain harmful information on victims ex ante.\footnote{On the concept
of mitigating the harm of dissemination of defamation by design see
Lavi, Publish, supra note 65, at 494. For expansion on a similar concept in the
realm protection of privacy see CHRI$ JAY HOOFNAGLE, FEDERAL TRADE
COMMISSION PRIVACY LAW AND POLICY 190–91 (Cam-
bidge Univ. Press, 2016).}

For example, the NFT marketplace could pre-screen the content of
the speech before tokenizing it, or choose to store it outside the blockchain
and only include a reference that points to it on the blockchain.\footnote{ul
Hassan et al., supra note 14, at 17; Fairfield, supra note 8, at 1272 (“[A]n NFT
stands for ownership of something not directly stored on the blockchain . . . .”)}

Such solutions could be adopted voluntarily by NFT marketplaces, or imposed on them through
regulation. The second type of solution focuses on obscuring harmful
speech. It is aimed mainly at internet intermediaries that can reduce
the visibility of tokenized expressions that have been integrated on the
internet. The third type of solution is directed at the original seller of
the token containing harmful speech and the buyers of such tokens. It argues that in some circumstances, especially when the token contains unprotected speech, the identities of the token's original creator and subsequent buyers should be unmasked. This would allow the victim to file an action against them in court. Courts could apply the remedy of compensation and obligate the original creator and the buyers to compensate victims or apply a remedy of disgorgement. Under such remedy, creators and buyers would be deprived of profits gained from commercialization and sale of the token. Part IV also addresses First Amendment concerns with respect to the constitutionality of the proposed solutions.

I. NFT Tokens: Creating Unique Rival Assets

People want to own things and have a deep passion for possessions that also plays a role in defining their identity. People can own different types of property. One type can be classified as “essentials”, i.e. things people consume and use which are necessary for survival. However, people also surround themselves with unreplaceable emotional goods that relate to their lives and experiences. These items carry with them sentimental and personal value. The fact that emotions are attached to possessions makes them pleasurable to consume. In the past, sentimental assets were unique and rivalrous. However, the shift to the digital era made it harder to preserve these features as “[e]very file sent is sent by making a copy.” The value attached to unique assets has vanished almost completely in the digital environment. The development of non-fungible tokens (NFTs), new unique blockchain-based tokens that have been in use since 2018 which can serve as verifiable proof of origin, has the potential to change this and resurrect the concept of unique, rivalrous assets in the digital sphere.

A. What Are NFTs? How Do They Work?

NFTs are unique tokens meant to represent digital assets that people can buy and sell. These tokens are characterized by special features that differentiate them from fungible tokens, or currencies. First, as each token carries specific data that is unique to it, NFTs are not inter-

74. See Hood, supra note 25, at 56; Rossenpavlov, supra note 25.
75. Fairfield, supra note 8, at 1264.
76. Id.
77. Id.
changeable. Thus, in contrast to fungible tokens such as USD or fungible cryptocurrencies, where any unit can be exchanged with another without affecting the holder, NFTs cannot be replaced with other non-fungible tokens of the same type. Second, they are unique and not uniform. Each token is distinct no one token is like another. The uniqueness of NFTs makes them valuable. Third, they cannot be divided as the basic unit is one token only.

How does one create an NFT? Blockchain technology enabled the creation of NFTs and storage of their value, as tokens are stored on the blockchain and their uniqueness can be certified. The original premise of the blockchain is to establish trust in a network, circumventing the need for any sort of third managing parties. There are NFT exchanges, where anyone can create an account, name and offer an NFT for sale, either for a fixed price or in an auction. Because the blockchain eliminates the need for intermediaries and relies on transparency, it has been dubbed a “trust machine.” The blockchain is maintained by an online peer-to-peer network. It utilizes a distributed ledger technology that tracks transactions, maintains a complete history of verified transactions and contains “algorithms that ensure consistency of data across storage locations.” Thus, ownership is listed on the blockchain. To prevent fraud, blockchain technology uses a consensus mechanism

---

78. Elrom, supra note 10, at 470.
79. Mofokeng & Matima, supra note 20, at 12; Diana Qiao, This is Not a Game: Blockchain Regulation and Its Application to Video Games, 40 N. Ill. U. L. REV. 187 (2020) (Each token carries unique information and varying levels of rarity, meaning that the value of one NFT is not equivalent to the value of another.).
80. Mofokeng & Matima, supra note 20, at 11 (“A crypto collectable is non-fungible, a cryptographically unique, non-replicable digital asset ...); ul Hassan et. al, supra note 14, at 6 (“NFTs is that every token is associated with a unique identifier, rendering each token unique to its respective owner.”). See also Lawrence J. Trautman, Virtual Art and Non-Fungible Tokens, 50 HOFSTRA L. REV. 361, 364 (2021) (“NFT technology leverages digital uniqueness in a way that makes a new social phenomenon possible. There is only one Mona Lisa in the Louvre: owning a copy doesn’t provide the same thrill.”).
81. Fairfield, supra note 8, at 1263 (“What makes these assets valuable is that they are one-of-a-kind.”).
82. Thippa Reddy Gadekallu, Thien Huynh-The, Weizheng Wang, Gokul Yenduri, Pasika Ranaweera, Quoc-Viet Pham, Daniel Benevides da Costa & Madhusanka Liyanage, Blockchain for the Metaverse: A Review, FUTURE GENERATION COMPUT. SYST., June 2023, at 404 (“[E]ach NFTs are non-interchangeable and cannot be divided.”).
83. See ul Hassan et. al, supra note 14, at 2.
84. See Frye, Death of Art, supra note 49, at 4.
which makes attempts to falsify the ledger tremendously expensive, or too risky.\textsuperscript{88} This structure makes the blockchain tamper-proof.\textsuperscript{89}

The verification and permanency of information on the blockchain is made possible thanks to a “mathematical relationship called a hash, and a consensus mechanism for verifying hashes.”\textsuperscript{90} Each block on the blockchain includes, among other data, a hash which can be attached to anything: pictures, text, or lists of transactions and more.\textsuperscript{91} The hash is what makes it possible to display who owns what.\textsuperscript{92} Once a transaction is made, it is recorded on a block on the blockchain. Each new block identifies the previous one through its unique hash.\textsuperscript{93} The process of “mining” creates these new blocks. In this process, several computers compete to solve a mathematical problem, and the computer that wins is given the opportunity to form a block and be rewarded (usually by receiving crypto).\textsuperscript{94} Hashing is the process that converts a grouping of digital data to a single “number,” a hash that serves as the unique identifier for the source data or a digital fingerprint of the source data that cannot be tampered with.\textsuperscript{95} This number is then used as the base for the mathematical problem presented to the next miner.\textsuperscript{96} When adding a subsequent block, it also mathematically references the hash from the previous block, and so on.\textsuperscript{97} In order to alter one block, it is necessary to change every block that came before. Such connectivity of the entire chain protects the blockchain, enables it to remember, and makes it resistant to fraud and censorship.\textsuperscript{98}

Distributed apps (“Dapps”) and software known as smart contracts can be programmed by developers to run on the blockchain.\textsuperscript{99} Ethereum, which runs a self-executing software programming language, is one example.\textsuperscript{100} Dapps can establish their own token systems by creating a smart contract. Such tokens “consist of a hash of the token’s transaction history, and a series of basic standard functions and features.”\textsuperscript{101} In the case of NFTs, the tokens consists of a hash contain-

\begin{footnotesize}
\begin{enumerate}
\item Fairfield, supra note 8, at 1269.
\item ESIOM, supra note 10, at 471.
\item Fairfield, supra note 8, at 1269.
\item Id. at 1270.
\item See id.
\item Id.
\item Id.
\item Id.
\item See Tonya M. Evans, Cryptokitties, Cryptography, and Copyright, 47 AIPLA Q. J. 219, 239 (2019).
\item Fairfield, supra note 8, at 1278.
\item Id.; Evans, supra note 95.
\item Fairfield, supra note 8, at 1279; Evans, supra note 95, at 239.
\item Fairfield, supra note 8, at 1280; Evans, supra note 95, at 240.
\item Evans, supra note 95., at 239.
\item Fairfield, supra note 8, at 1272.
\end{enumerate}
\end{footnotesize}
ing the transaction history of the token, some basic standard features and functions and a URL which allows tracking the file related to the token (the image or text the token represents). The most prominent standard to create an NFT is the ERC-721 standard, which produces unique, non-fungible tokens. Gains or losses in the value of each and every token are different, as every token is unique.

B. This is Mine! Main Uses of NFT Tokens

1. Crypto Art and Historical Documentation of Inventions

NFTs are a novel medium to distribute and exhibit digital art. They are a game changer, as they connect artworks to a digital file that only one person can own, thereby facilitating proof of ownership of unique, original digital artworks. Tokenizing art thus makes it possible for the owner to use their token and display it in various online social spaces, or even in real life utilizing novel physical display methods. One method of display, which allows the artwork to adorn the owner's physical living or working space, uses special screens that resemble picture frames (such as Samsung's "The Frame"). If the art is displayed in public, this also fulfills the owner's need to gain recognition as the owner.

Verifying the artwork on the blockchain helps authenticate the artwork with a unique identification code that distinguishes it from other tokens and protects creators. This protection is achieved in ways which were previously impossible through traditional copyright law and technological protection measures, by allowing proof of ownership through

102. Id.
103. Id.; Evans, supra note 95, at 248 ("ERC-721 is a finalized coding standard interface for non-fungible tokens, referred to early on as deeds, evidencing ownership of both wholly digital crypto assets and physical assets represented in token form (i.e., tokenized assets).”).
104. Fairfield, supra note 8, at 1283.
106. Frye, Death of Art, supra note 49, at 4 (criticizing the concept of NFT: “The problem with “owning” a “unique” copy of a digital artwork is there’s nothing to own”). See id. at 9 (“NFTs mean the art market doesn’t need art anymore”).
108. Id.
109. Fairfield, supra note 8, at 1274 (“The desire to display motivates people to buy digital NFT art just as the desire for recognition motivates displays on social media or in a video game.”).
transfer of the token. This new technological development allows artists to maximize profits and render intermediaries (such as art galleries, art dealers and online platforms) redundant. Buying NFTs also empowers buyers, who gain different types of value from the purchase: (1) The first buyer gains satisfaction from financing a special art project, (2) all buyers gain pride and recognition from their connection to a work of art, (3) buyers become the owner of a tradeable asset which can easily be traded globally. However, as Professor Frye explains, purchasing NFTs is different from purchasing ownership in copyrighted items. As a buyer, one is basically buying unique access to a digital file. In other words, buyers of an NFT own only the NFT. As Frye explains, it is ownership without control of the art itself. And yet, many people buy and sell NFTs anyway, perhaps because as mentioned, NFTs are emotional goods and fulfill the individual’s inherent need to possess As such goods exist and are valued in a social context, ownership helps the owner gain social recognition, endorsement of the author, and of society.

The hype of tokenizing artwork is driven by the buyer’s motivation to gain recognition as the owner of the artwork. Thus, creators who possess the copyright for a work of art (such as a painting, photograph, or video) can gain huge sums of money from selling the tokenized crypto art. For example, Mike Winkelmann, the digital artist known as Beeple,

110. Evans, supra note 95, at 259.
111. Trautman, supra note 80, at 415.
112. See id. at 364, 372 (“Collecting art is social.”).
114. Frye, Death of Art, supra note 49, at 3 (“Only the owner of the NFT can access or transfer that particular unit of data. Essentially, an NFT is a unique “digital object” that someone can own, sell, or buy.”).
115. Id. at 6 (“You aren’t buying a unique digital artwork, because it’s impossible. There’s no such thing. And even if the digital artwork associated with the NFT is protected by copyright, you aren’t buying copyright ownership or any kind of copyright interest. All you own is the NFT.”).
116. See Frye, After Copyright, supra note 49.
117. Frye, Death of Art, supra note 48, at 7.
118. See supra § 1. See also Frye, Death of Art, supra note 49, at 9 (“It doesn’t really matter what you’re buying, all that matters is the demand to own it.”).
119. Trautman, supra note 80, at 364 (“humans value rarity and uniqueness particularly in a social context.”).
120. Frye, After Copyright, supra note 49, at 341–42. (“I call this ‘ownership,’ because it consists of ‘clout,’ rather than control. NFT owners don’t need copyright, because ownership depends on the endorsement of the author, rather than control of the use of the work.”).
121. Fairfield, supra note 8, at 26 (“The desire to display motivates people to buy digital NFT art just as the desire for recognition motivates displays on social media or in a video game.”).
sold his NFT digital artwork in an auction for USD 69,000,000. The artwork “Sophia the Robot”, which was produced by collaboration of Sophia, the humanoid robot, and Andrea Bonaceto, an Italian artist, was tokenized and sold for USD 688,888. The image of Zoe Roth, taken in January 2005 by her father when she was four years old, was sold as an NFT on April 2021 for a whopping USD 473,000. In the photograph, Zoe is standing in front of a burning house. The picture became an internet viral meme synonymous with disaster. A viral video from 2007, showing the baby Charlie Davis biting his big brother's finger and laughing, was tokenized and sold for USD 760,999. Even though the video was sold as an NFT on the blockchain, it remains on YouTube for public access.

The phenomenon of tokenized art is expanding to inventions and tokens containing documents with historical value. For example, UC Berkeley recently offered the opportunity to purchase an NFT of the patent disclosures at the heart of two Nobel Prize-winning inventions from the university’s research labs, as well as an NFT that links to digitized internal forms and correspondence that document the initial research findings that led to highly significant biomedical breakthroughs: CRISPR-Cas9 gene editing and cancer immunotherapy. The NFT was sold for nearly USD 55,000. UC Berkeley intends to use the profits to fund research.

Similarly, Tim Berners-Lee, the founder of the World Wide Web, auctioned the original code for the web as an NFT, “[t]he NFT includes original time-stamped files containing the source code written by Berners-Lee, an animated visualization of the code, a letter written by Berners-Lee on the code and its creation, and a digital ‘poster’ of the full

code”.130 All of the components are digitally signed by Berners-Lee.131 Auction publications announced: “[e]ver thought about what it would be like to own the World Wide Web? Now you sort of can — well, a digital representation of its source code anyway,”132 this NFT of was ultimately sold for USD 5,400,000.133

Another landmark auction is of the first text message that a Vodafone programmer Neil Papworth sent to Richard Jarvis, then a director of Vodafone, 29 years ago. Vodafone has turned the first text message into an NFT that was sold at a Paris auction house for about U.S.D 150,000. The company will donate the revenues to the United Nations Refugee Agency.134

Such recent landmark auctions and token purchases can raise awareness among art buyers, creators, and investors of the market for unique digital art represented by NFTs.135

2. Collectible Game Cards and Virtual Goods

Markets for collectibles, such as stamps, rely mainly on the rarity of goods.136 By utilizing the features of the blockchain and the ERC-721 standard, NFTs enable a kind of rarity for digital assets which was not previously possible, as without NFTs digital assets can easily be replicated.137 Thus, people may buy collectible NFTs as they would buy traditional collectibles. They value collectibles not only because they gain satisfaction from owning and displaying them, or due to the potential to profit from selling them at a higher price,138 but also, and even mainly,

131. Id.
135. See Trauman, supra note 80, at 363.
136. See Jonathan E. Hughes, Demand for Rarity: Evidence from a Collectible Good, 70 I. INDUS. ECON. 147, 160 (2022) https://bit.ly/2TyxtCB [https://perma.cc/86MM-zDE9] (explaining that rarity is a prominent reason for collecting collectibles); Ezekiel & Mofokeng, supra note 20, at 2 (“[P]eople have traditionally placed their money in collectibles that were scarce or rare . . . ”).
137. See Mofokeng & Matima, supra note 20, at 4–5.
138. Id. at 4 (referring to displaying and gaining profits from trading as central reasons for owning collectibles).
because of the social context of collecting and trading. This same context also allows the use of collectible NFTs in the gaming industry.\textsuperscript{139}

One of the first examples of such a collectible, NFT blockchain-based, unique and tradable digital asset is Cryptokitties.\textsuperscript{140} This game emerged at the end of 2017 utilizing the Ethereum blockchain-based platform. In Cryptokitties, tokens are attached to animated pictures of cats and are thereby secured against replication. Now, these pictures can only be transferred with their owner’s permission.\textsuperscript{141} Players “buy and sell cats, breed them to make a new cat or rent them out to breed...Each cat has a unique digital identity, and from that it gets its unique features, like green eyes and spots”.\textsuperscript{142} Such features affect their price. Cryptokitties gained popularity because of its social context. It became a game for a rich, online community of users which interacted in breeding, trading, and playing with virtual cats, creating an entirely new market.\textsuperscript{143} Such a market streams revenue from trading Cryptokitties NFTs. The volume of sales testifies to the development of this new market, as over a period of approximately two years, there have been 538,043 sales of 431,680 unique Cryptokitties, for a total of USD 27,233,277.\textsuperscript{144}

Other collectible NFTs were developed for more idealistic goals. For example, “Honu” was developed with the objective of using auction sales profits to promote environmental goals. Created in 2018 by the same company which developed Cryptokitties (Axiom Zen), Honu is a sea turtle, and profits from its sales are used to raise money for sea turtle preservation in the Caribbean.\textsuperscript{145} Similarly, the Panda Project aims to preserve the endangered Panda. This blockchain-based App creates unique digital copies of real pandas. These tokens can be exchanged with other tokens in the project.\textsuperscript{146}

\textsuperscript{139} See Selen Türkay & Sonam Adinolf,\textit{ Friending to Flame: How Social Features Affect Player Behaviors in an Online Collectible Card Game}, CHI CONFERENCE (May 4–9, 2019).

\textsuperscript{140} CRYPTOKITTIES, https://www.cryptokitties.co/ (last visited Dec. 29, 2022); ul Hassan et. al, supra note 14, at 6; Wilson, supra note 34, at 365; Fairfield, supra note 8, at 1276 (“Kitties’ are pictures of cats tied to a particular ERC-721 token.”).

\textsuperscript{141} ul Hassan et. al, supra note 14, at 6; Fairfield, supra note 8, at 1276 (“When an owner wanted to sell a kitty, the token was traded from one wallet to another, and the ownership interest recorded by the smart contract.”).

\textsuperscript{142} Wilson, supra note 34, at 379. (quoting VICE News,\textit{ This Game Combines the Internet’s Favorite Things: Cats and Cryptocurrency (HBO), YOUTUBE, at 00:26} (Dec. 5, 2017), https://www.youtube.com/watch?v=qCqgRlJKew [https://perma.cc/SPXG-NZU7]).

\textsuperscript{143} Fairfield, supra note 8, at 1276.

\textsuperscript{144} Wilson, supra note 34, at 370.

\textsuperscript{145} Mofokeng & Matima, supra note 20, at 12.

\textsuperscript{146} Id. at 13; Panda Earth,\textsuperscript{ DAPP RADAR} https://dappradar.com/ethereum/collectibles/panda-earth [https://perma.cc/UH9Y-MTWC].
Collectible NFTs have also begun to attract attention in the sports field, where teams can use them as a source of revenue. One such example is NBA Top Shot, a marketplace where basketball fans and collectors can buy, sell, and trade NFTs of the most memorable moments in NBA history.

NFTs have also upgraded online virtual reality worlds, and are expected to gain more popularity with the development of the Metaverse as there is a gordian knot between the two. As such tokens are expected to be an integral part of the metaverse. Undoubtedly, the economy of the metaverse depends to a large extent on the possibility to authenticate visual assets. This authentication can be reached by creating NFTs which are blockchain based. NFTs allow real ownership of collectible assets, rather than a mere license that applies to one specific game only. For example, the virtual sword Dragon Saber can only be owned in the virtual world Legend of Mir and cannot be used in other virtual worlds. Moreover, since NFTs are eternal, they do not disappear when the virtual world shuts down.

In the online world “Decentraland,” for instance, the virtual assets are NFTs, “from its virtual plots of land to the art on the walls in the virtual galleries.” This ownership allows players to interact with the digital asset. The NFTs also allow owners to influence how the virtual world operates and trade the virtual assets in the Decentraland Marketplace.

---

148. See Trautman, supra note 80, at 186–87.
149. Metaverse refers to a 3D virtual shared world where all activities can be carried out with the help of augmented and virtual reality services. For further information see generally Muhammet Damar et. al, Metaverse Shape of Your Life for Future: A Bibliometric Snapshot, 1 J. OF METAVERSE 1 (2021).
151. Trautman, supra note 80, at 400–01 if a person’s avatar earned the Sword of Admiration, the game creator’s database reflected that she owned that sword (and thus the avatar would display it, and the sword appeared in the avatar’s inventory), but no-one else would have such a sword available. But the player could not take the item out of the game, or into any other virtual environment.”).
153. Trautman, supra note 80, at 424.
Lastly, collectible NFTs are used in the digital trading space. “Gods Unchained,”\textsuperscript{157} for example, is a blockchain-based game that incorporates NFTs and gives players complete ownership over in-game items. In this game, players sell and collect playing cards, which can then be resold for potential profit.\textsuperscript{158} The context of the game enhances the value of the card, as ownership is verifiable on the blockchain.\textsuperscript{159}

Thus, it might be said that the advantages of NFTs have opened new economic markets for investment in collectibles and other virtual assets.

3. Markets for Physical Items/ Distribution of Resources and Ticket Brokers

Many NFTs are limited to activity in the digital sphere and function only as digital assets. However, this is not always the case. Some types of NFTs are indeed used for distribution or exchange for real-life items or resources.\textsuperscript{160} For example, Zora is a marketplace to “buy, sell and trade limited-edition goods. All of these goods are launched as tokens.”\textsuperscript{161} These tokens are then tied to a physical item instead of a digital asset. Users can trade items without taking physical possession of them. If an owner wants to physically possess the item, they return the token to Zora, and the tangible item is sent to them.\textsuperscript{162} This saves costs and trouble associated with shipping, as users trade without taking physical possession of the item. Zora and other places like it allow integration of the physical and digital worlds, or example users might buy real baseball cards or other trading cards and play with them against

\textsuperscript{158.} Fairfield, supra note 8, at 1298–99; Shafin Rizvi, How to Buy and Play Gods Unchained Trading Cards, TRUST WALLET (Mar. 29, 2021), https://trustwallet.com/blog/how-to-buy-play-gods-unchained [https://perma.cc/3CL9-XU85].
\textsuperscript{159.} Rizvi, supra note 158
\textsuperscript{160.} See Fairfield, supra note 8, at 1277 (“NFTs have uses beyond digital assets as well. Zora, an NFT-enabled “Everything Exchange,” allows users to sell rare or unique items on their marketplace using tokens. Every item is given an NFT, and users are able to buy, sell, and trade the tokens on the exchange. Rather than being tied to a piece of digital art, the token is tied to a physical item. The token allows users on Zora to trade the item and capture its rise in value without ever having to ship the item. Once someone wants to buy the physical item, they can redeem the token and the physical item will be sent to them.”).
\textsuperscript{161.} @ OurZora, TWITTER (Mar. 2, 2020 8:59 AM), https://twitter.com/ourzora/status/123447845934814976?lang=en; Fairfield, supra note 8, at 1277.
\textsuperscript{162.} Id.
others in online baseball games. Such NFT collectible products are tied to a rich social context in digital spaces.

Tokens can also be tied to real world experiences and activities. One of the most promising areas for NFTs is event ticket sales. The owner of the world’s largest ticket marketplace, Ticketmaster, recently made predictions regarding the potential of NFTs in the concert market. Ticketing companies could use NFT tokens to create transferable tickets. Once an NFT ticket is created, it can be transferred without the involvement or approval of the ticketing company. Thus, ticket owners would be able to exchange the ticket for another date or event, transfer it to a friend, or sell it. NFT systems could even be instrumental in creating an additional social dimension to event attendance. People who attended a play, movie, or rock concert could retain the tokenized ticket, post attendance on social media platforms and connect with others who attended the same event.

4. The Music and Film Industries – Ownership and Rights Management

NFTs are also starting to gain popularity in the music industry, as collectibles and as a tool for gaining royalties. For instance, 30-year-old electronic-musician Justin Blau, known as 3LAU, sold 33

163. Id.
164. Lesavre et al., supra note 46, at 42.
166. Id.; Onkar Singh, What is NFT Ticketing and How Does it Work? Cointelegraph (Feb. 14, 2023), https://cointelegraph.com/news/what-is-nft-ticketing-and-how-does-it-work (“the adoption of blockchain technology enables transparency and traceability, making it simpler to trace the ownership and origin of the ticket. NFT tickets can also be sold or exchanged on online exchanges, with their value depending on how much interest there is in the event.”) [https://perma.cc/K2LK-LTRB].
167. Ledger Insights, supra note 165.
169. See Lawrence Wintermeyer, Lindsay Lohan On Why NFTs Are Destined For Hollywood, Forbes (Mar. 26, 2021) https://www.forbes.com/sites/lawrencewintermeyer/2021/03/26/lindsay-lohan-on-why-nfts-are-destined-for-hollywood [https://perma.cc/U9U2-4iMs] (“Tokenization through NFTs can help content creators and musicians actually own the property rights for what they create, and allow them to profit accordingly...Royalties are likely to be another avenue artists explore. Having been a Disney child star, Lohan is intimately familiar with the revenue model and its long-standing inefficiencies.”).
NFTs, for varying prices, and obtained USD 17,000,000 from NFTs, "helped in part by a tokenized release of his three-year-old album 'Ultraviolet,' which grossed $11.6 million and briefly held the record for the highest price paid for a single NFT, $3.6 million." Recently, Quentin Tarantino tokenized a collection based on his film "Pulp Fiction" which included "...high-resolution scans from his original handwritten screenplay of Pulp Fiction, plus a drawing inspired by some element of the scene..." However, Miramax, the film's production company, filed an action against him in a California court. Miramax claimed that Tarantino violated its copyright and trademark rights, as these NFTs don't fall under Tarantino's reserved rights for the film. Therefore, Miramax demanded that Tarantino halt the upcoming sale of these NFTs. In a recent motion, Tarantino argues that he's not infringing on any of Miramax's copyrights since the NFTs will exploit the screenplay for Pulp Fiction and not the movie itself.

But why would artists want to buy or create an NFT? NFTs allow users to own the assets underlying the NFT, but they can also allow users to receive a license while ownership remains with the original owner or artist. Like most other blockchain applications, NFTs allow artists to circumvent the middlemen, such as record labels, that might dictate unfavorable terms when establishing contracts with most artists. NFTs allows artists to create a license which protects their intellectual property rights while enabling them to monetize their work. For instance, musical works can be tokenized and shared through Tokenized Music License ("TML"). Thanks to blockchain-based platform characteristics such as immutability, transparency, programmability and decentralization, TML can provide innovative models for music distribu-

---

171. Trautman, supra note 80, at 409.
173. Id.
175. Fairfield, supra note 8, at 1295.
177. Evans, supra note 95, at 264; Adjovu & Fabian, supra note 30, at 8 ("The blockchain supply chain is expected to allow musicians to directly publish their music on a blockchain, then have their music reached by consumers on blockchain-based platforms.").
178. See Adjovu & Fabian, supra note 30.
Two blockchain-based music streaming and downloading services, Rocir and Ujo, operate based on a direct fan to artist model, where music is tied to the NFT and cannot be deleted. Under this system, musicians can quickly and efficiently receive royalty payments since “each time a music file is played, a transaction can be recorded on the blockchain with the applicable royalties due to rights holders.” Ujo Music was “one of the first blockchain-based music platforms to issue royalty payments to musicians on the Ethereum blockchain.” This allows artists to collect their royalties and enjoy financial compensation for their work within days or months, instead of what previously took years. Moreover, these platforms allow musicians to obtain information about royalties earnings in real time.

C. The Future: Tokenizing Speech - Sharing NFTs as an Integral Part of the Internet

There is no doubt that more NFT uses will emerge in the future. This part spotlights one NFT use that will undoubtedly expand and become commonplace: tokenizing speech. Speech is already tokenized today, as demonstrated by the examples detailed in the introduction to this Article: Jack Dorsey’s first tweet and Kevin Roose’s New York Times column, both of which were tokenized as NFTs. However, to date, the aim of tokenizing speech has mainly been to sell such NFTs as collectibles for profit, due to their current or prospective historical or artistic value. In other words, the token was created to allow digital scarcity. The expressive value of the token as a unique form of communication was secondary. This Article predicts that NFTs will become a common means for communicating information.

As explained in Part I, blockchain transactions and data cannot be altered or deleted from the blockchain. The blockchain never forgets the information a token carries. Such unchangeable, eternal information

---

179. Id. at 70; see e.g., Rocki, supra note 31; Introducing Streaming Payments for Ujo with Connect Payments Channels and Dai, UJO (Mar 28, 2019), https://blog.ujomusic.com/ [https://perma.cc/TX62-ZZJ].
180. Adjovu & Fabian, supra note 30, at 36.
181. Id. at 10.
183. Peters, supra note 1.
184. Roose, supra note 6.
185. ELROM, supra note 10, at 471; Po-Wei Chen & Bo-Sian Jiang, supra note 50 (“all the transaction records, which have been stored into the blockchain, cannot be modified and deleted”).
can have highly significant implications for communication and freedom of expression, since it removes barriers and impediments to the free flow of information. This potential of NFTs will likely be realized quite soon, due to the anticipated incorporation of NFTs in social media. APIs have already begun to integrate NFTs with the internet, as tokenized tickets can be used as digital souvenirs and “add an attestation of it on social media to connect with other attendees and artist.” NFTs are an integral part of the new addition of social media and in particular in the wake of the metaverse - a network of three-dimensional, virtual or augmented reality that will be Web 3.0 blockchain-based. Such social networks are expected to develop and become a key player in an internet that will become open, without any central intermediary.

The widespread distribution of NFTs has tremendous implications and is a game changer for freedom of expression. NFTs make it possible to create eternal expressions that cannot be censored or forgotten. Because the expressions are recorded on the blockchain, such expressions would perpetually remain accessible to all Web 3.0 blockchain-based network users.

The next part of this Article will focus on the centralized nature of online websites that depend on intermediaries and the practice of gatekeeping. It will explain the main motives intermediaries have for censoring user speech and will explore these types of censorship. It will demonstrate that the law immunizes intermediaries from liability for their editorial decisions and allows them to shape the online discourse.

II. PUBLIC-PRIVATE AND PRIVATE ONLINE CENSORSHIP AND THE NFT ENGINE OF SPEECH

Contrary to what many people may believe, the internet is not a sovereign-free medium controlled by its users from the “bottom-up.”

186. See, e.g., ul Hassan et al., supra note 14, at 11 (explaining that API can be programmed in a way that could allow blockchain infrastructure to “talk” with the internet infrastructure).
187. Lesavre et al., supra note 46 at 42.
189. See Mersch & Muirhead, supra note 48.
190. See Aurélie Bayle, Mirko Kosicina, David Manset & Octavio Perez-Kempner, When Blockchain Meets the Right to be Forgotten: Technology Versus Law in the Healthcare Industry 789 (2018) IEEE /WIC/ACM International Conference on Web Intelligence (WI) Paper, 2018 (“Effectively, the blockchain immutability allows considering that by design, anything cannot be deleted from the ledger”). For more information on Web 3.0 as a decentralized version of the internet, see supra notes 46–47.
Although it may seem as if anyone “can publish freely and instantly online,” many intermediaries “actively curate the content” their users post on their platforms. For this reason, intermediaries have been dubbed the “New Governors of online speech.” Intermediaries use diverse strategies to moderate user-generated content. Unfortunately, there is still not enough transparency regarding these strategies. Yet there is no dispute over the fact that moderation is a fundamental aspect of any platform. This type of intermediary intervention in content is needed for sound operation of the internet. Without it, platforms would overflow with spammers, and this would interfere with the user’s ability to find real and relevant content and to engage with other users. Moreover, without moderation, platforms would be abused by “bad actors,” which as the next part demonstrates, has the potential to flood platforms with negative value content.

One main problem with platform moderation is the fact that there is not enough transparency as to what content is removed, when, and why. This can result in removal of legitimate content and censoring of specific types of agendas without accountability, ultimately impeding the marketplace of ideas. The following subsections will review several motives that could lead to censoring of legitimate content and provide examples.

A. Motivations for Removing Content

1. Removal of Harmful Content to Comply with the Law

The first motivation for intermediaries to remove content is legal rules. In such cases, the reason for removing content is collateral cen-
sorship that occurs “when a (private) intermediary suppresses the speech of others in order to avoid liability.”\textsuperscript{198} For such speech. As this Article will explain in detail,\textsuperscript{199} in the U.S., even if intermediaries refrain from taking action and removing illegal content on their site, they generally remain immune from liability. This is true even where there is knowledge of potentially illegal content.\textsuperscript{200} However, intermediaries might voluntarily take down content in response to reports that unlawful content has offended users. For example, intermediaries may remove a defamatory expression following a report, or a takedown notice, by a person who was defamed. They might do so because of a sense of corporate responsibility, or as an attempt to enhance their social standing and profits, if they think this will make users view them as family friendly.\textsuperscript{201} They might also remove content as a preventive measure to shy away from murky legal areas and diminish the likelihood of claims against them, such as in the case of speech that constitutes a criminal offense. Although intermediaries are shielded from civil liability, such as in instances of defamation, they are not immune from liability for third party expressions that constitute federal criminal offenses.\textsuperscript{202}

Outside the U.S. there is no overall immunity for intermediaries for content published by third parties. Many jurisdictions require intermediaries to remove illicit content. For example, the German government drafted the Network Enforcement Act (“NetzDG”) to target hate speech and fake news.\textsuperscript{203} The Act obligates intermediaries to delete content that

\textsuperscript{199} See Part II. B
\textsuperscript{201} See Lavi, Publish, supra note 65, at 512.
\textsuperscript{203} Netzwerkdurchsetzungsgesetz [NetzDG] [Network Enforcement Act], Oct. 1, 2017, (Ger.), trans. athttps://www.bmjw.de/SharedDocs/Gesetze/3814/2/Dokumente/NetzDG_engl.pdf?__blob=publicationFile&v=2 [https://perma.cc/NT6Z-3BDK]; Heidi J. S. Tworek, Fighting Hate with Speech Law: Media and German Visions of Democracy, 35 J. HOLOCAUST RSH. 106 (2021); id. at 113 (“Despite the law, it remains unclear whether NetzDG has made significant headway in stopping hate speech, which was one of its main stated aims”); id. at 118 (“Singapore, a state that has often used
is “evidently unlawful” within 24 hours of the filing of a complaint.\textsuperscript{204} Similarly, the European Commission has launched a legislative bill for a regulation designed to prevent the dissemination of terrorist content online. This proposed regulation is intended to complement Directive 2017/541 on combating terrorism. The bill requires intermediaries to remove terrorist content within “the first hours after it appears online because of the speed at which it spreads.”\textsuperscript{205} In addition, the proposed regulation would require platforms to “adopt more proactive measures to prevent the spread of terrorist content in the first place.”\textsuperscript{206}

Another motive that may lead intermediaries to remove content is a safe haven provision. For example, Article 14 of the EU’s e-Commerce Directive dictates the framework for intermediary liability.\textsuperscript{207} It states that intermediaries are exposed to a ‘notice-and-takedown’ regime for hosting illegal content. Under this regime they are obligated to remove the illegal content in order to avoid liability.\textsuperscript{208} Although the e-Commerce Directive does not stipulate a general monitoring obligation for online platforms, the European Court of Justice (“ECJ”) has held Facebook and other intermediaries liable for content that was identical to existing content which had previously been declared unlawful.\textsuperscript{209} A similar regime was adopted by the E.U. in the context of online infringing content.\textsuperscript{210} In order to meet this requirement, intermediaries will need to use screening technology to monitor their platforms.

Another aspect of the discussion of content removal relates to the “Right to be Forgotten.” In the EU, citizens benefit from a “Right to be

---

\textsuperscript{204} Lavi, Publish, supra note 65, at 476–77.


\textsuperscript{208} Lavi, Publish, supra note 65, at 476.

\textsuperscript{209} Case C-18/18, Glawischnig-Piesczek v. Facebook Ir. Ltd., ECLI:EU:C:2019:821 (Oct. 3, 2019) (reviewing the decision of the Vienna Commercial Court). For further information on this case see Lavi, supra note 65, at 478.

Forgotten,” that was first established by the Data Protection Directive (“DPD”).

In Google Spain SL, Google Inc. v. Agencia Espaolade Proteccin de Datos, the ECJ held that search engines, such as Google, must remove search results that link to personal information. This includes removal of defamatory content found on third-party websites upon request.

The General Data Protection Regulation (“GDPR”) replaced the DPD in May 2018. This Regulation includes a specific provision titled “right to erasure (‘right to be forgotten’)” that imposes data controller obligations to erase data relating to specific identified individuals. The GDPR imposes more specific obligations regarding information processing.

State regulation, as detailed above, does not focus on the direct wrongdoer but rather on digital infrastructure owners to encourage “them to regulate and surveil end-user speech according to the government’s purposes.” This type of regulation was dubbed the new-school speech regulation and it is directed at online infrastructure. Such regulation is becoming increasingly important in the algorithmic society, in which individuals and organizations disseminate content at enormous speed and scale.

In short, intermediaries can, and many times do, take down lawful expressions because of uncertainty regarding their liability. Liability creates an incentive to remove more content than necessary. Concerns regarding such collateral censorship increase with the use of algorithms for content moderation. Algorithms will likely fail to capture context accu-

216. Id.
217. See Michael L. Rustad & Thomas H. Koenig, Towards a Global Data Privacy Standard, 71 Fla. L. Rev. 365, 418 (2019) (“The GDPR requires companies ‘to take into account the protection of the rights of individuals, both before and during their processing activities, by implementing the appropriate technical and organization measures to ensure that they fulfill their data protection obligations’”).
218. Balkin supra note 69 at 8; id. at 37 (“[G]overnments seek to coopt or coerce infrastructure owners to govern speech in ways that governments like.”).
219. Id. at 8.
220. See id. at 8–9.
221. Id. at 5 (“The speed and scale of digital speech have transformed how speech is governed, regulated, and protected. Social media companies have developed an algorithmic-administrative system for governing speech that does not view speech in terms of rights.”).
222. See Lavi, Platforms, supra note 195, at 537.
rately, resulting in "over-removal" or "false positives": erroneous removal of lawful content.\textsuperscript{223}

2. Public-Private Cooperation in Content Removal

Removing harmful content published by private parties in order to obey the law is one thing, but what about removing content due to a direct request from government officials? In such cases, the motive for removing such content is cooperation with the government. In fact, platforms cooperate with governments on these matters extensively.\textsuperscript{224} Removal of content in accordance with government requests, even when the content at stake is beyond the legal definition of unlawful, is not uncommon around the world. In nations with authoritarian governments that control the internet, the government can directly censor speech.\textsuperscript{225} However, even in democracies, governments cooperate informally with intermediaries to remove content. Such informal cooperation has been dubbed "the invisible handshake."\textsuperscript{226} The problem with this cooperation is that it allows governments to bypass constitutional constraints.\textsuperscript{227} In the U.S, for example, intermediaries are not subject to the First Amendment and can censor speech.\textsuperscript{228}

A prominent example of such an invisible handshake is the cooperation between the Cyber Unit in Israel's Office of the State Attorney and online platform intermediaries. A special voluntary enforcement track was developed for removal of allegedly unlawful content such as incitement or harassment of public officials.\textsuperscript{229} Accordingly, it has been reported that the Cyber Unit conducts an internal procedure for review-

\textsuperscript{223} Id. at 499.
\textsuperscript{225} See Tufecki, supra note 53, at 234 ("The Chinese government's strategy for managing the internet is also centered on a deep understanding of the importance of attention and capacity to movements, rather than merely blocking information").
\textsuperscript{227} See Richards, supra note 61, at 139 (2021) (explaining that government agencies often collaborate with private actors in order to do things they would otherwise be constitutionally forbidden from doing).
\textsuperscript{228} Lavi, Publish, supra note 65, at 488.
\textsuperscript{229} Viki Auslender, Who Gets to Define Incitement?, C1cH By CAlCALIST (May 27, 2021, 2:42 PM) https://www.calcalistech.com/ctech/articles/o.7140.L-3708694.00.html [https://perma.cc/ZN7Q-BUL6] ("When Israeli security sees posts they do not like online, they send them to the cyber unit, which, in turn, selects those it believes violate a platform’s policy and asks, rather than demands, that they will be removed. Since the cyber unit’s inception in 2015, platform compliance rates have risen steadily").
ing the legality of allegedly unlawful content and requests the intermediaries to remove it.230 In practice, about 90 percent of requests made by the Cyber Unit to the platforms are accepted.231 In Adalah v. Cyber Unit,232 two petitioners claimed that the Cyber Unit's activities infringed on their constitutional right to freedom of expression.233 However, Israel's High Court of Justice held that no infringement of constitutional rights had occurred, and thus, de facto legitimized this practice. This type of private-public cooperation is also becoming common in other parts of the world. For example, police agencies in Europe have formed special “Internet Referral Units” to report and flag violations of platform content rules for takedown.234 In the U.S., “[e]ven with Section 230’s liability shield intact, government agencies often engage in efforts to coerce, compel, or convince intermediaries to take down harmful content ....”235 Thus, even though U.S. law does not obligate intermediaries to take down content and even shields them from liability for editorial decisions, they often obey government agencies’ requests and take down user-generated content from their platform.

3. Removal of Content that Does not Comply with the Terms of Service or the Intermediary’s Agenda

As private entities in the U.S., intermediaries are not constrained by the constitutional limitations of the First Amendment.236 They are entitled to set “community guidelines” limiting certain types of speech on their platforms, and they can enforce these restrictions, even if the content concerned is a legitimate expression. Thus, intermediaries are able to remove speech, even speech that is protected by the First Amendment, while they themselves have no legal obligation to remove harmful speech.237 In some instances, intermediaries act out of a sense of corpo-

---

230. Shadmy & Shany, supra note 55 (“The Cyber Unit then conducts an internal process for reviewing the legality of the content in question and the propriety of seeking its removal”).
231. Id.
233. For further information on this case see Shadmy & Shany, supra note 55.
234. See Brian Chang, From Internet Referral Units to International Agreements: Censorship of the Internet by the UK and EU, 49 COLUM. HUM. RTS. L. REV. 114, 120–22 (2018); see also Bloch-Wehba, Content Moderation, supra note 206, at 1300.
235. See Bloch-Wehba, Content Moderation, supra note 206, at 1305.
236. Balkin, supra note 69, at 18 (“In general, U.S. free speech doctrine has a difficult time dealing with the conflicting speech rights of private parties”).
237. Eric Goldman & Jess Miers, Online Account Terminations/Content Removals and the Benefits of Internet Services Enforcing Their House Rules, 1 J. OF FREE SPEECH L. 191, 201 (2021) (“the Internet ser-
rate responsibility: they take down content they believe to be harmful to the public good. In other cases, they may believe that a removal policy can help improve their image and increase profitability. They may also restrict content to distance themselves from specific agendas. Intermediaries outline the terms under which users and platforms interact in their “terms of service.” These address “not just appropriate content and behavior but also liability, intellectual property, arbitration, and other disclaimers.” Intermediaries also outline “community guidelines” that present the platform’s expectations of what kinds of content users should be allowed to share, clarifying the platform’s principles and setting forth a list of prohibited types of content.

One common example of such prohibited content is sexual content, such as nudity or photographs of sexual activity. Another type of prohibited content proscribed by nearly all social media intermediaries is violent or obscene content. A third prohibited category is harassment and threat, such as exposing a person’s private information, address, or location. Fourth, most social media platforms prohibit hate speech such as racism, white supremacy, Holocaust denial and other hateful expressions against groups in society. Although some platforms fall short of naming hate speech explicitly, it can be included within umbrella categories such as “personal attacks.” Intermediaries also prescribe posts that encourage illegal activities such as “instructional bomb making, choking games, hard drug use, or other acts where serious injury may result.” Another category restricted by some social media intermediaries is discussion of self-harm, such as goading users to injure themselves, commit suicide, or develop an eating disorder such as anorexia or bulimia. Some social media platforms, such as Facebook, have a “real name policy” and remove content or accounts of users that
do not use their real names.\textsuperscript{247} Other intermediaries prohibit commercial content, such as self-promotion, in an effort to prevent platform exploitation for commercial traffic.\textsuperscript{248} Lastly, intermediaries encourage users to post high-quality content and refrain from transmitting fake news.\textsuperscript{249}

Intermediaries must moderate content, as explained above. In fact, moderation is a necessary part of any platform, without which platforms could be abused by malicious users who might flood them with negative content, or overload them with cacophony, making it difficult for users to weed out irrelevant content.\textsuperscript{250} Such moderation, however, is conducted without transparency and due process. Often it is even executed through opaque algorithms,\textsuperscript{251} that determines “what speech to take down and leave up, amplify, and demote”.\textsuperscript{252} Social networks may adopt a categorical approach to content removal and fail to differentiate between nuances of context in the content published on their platforms. Moreover, although they may aim to remove negative value content and harmful expressions, due to “false positives” they could remove legitimate speech instead. Finally, intermediaries may remove legitimate content that runs counter to their own agendas, essentially chilling legitimate speech. The following subsections will demonstrate these practices.

\textit{(a) Categorical Classification that Overlooks Nuances of Context}

In some cases, intermediaries limit specific types of content in their terms of service and community guidelines. For example, most intermediaries restrict the publication of nude pictures on their platforms. In many cases, removing nude images has benefits, as such content can be offensive. Yet some intermediaries do not approve any kind of nudity. In cases where the approach is to ban all nudity, intermediaries categorically remove all nude images from their platforms, without consid-

\textsuperscript{248} Gillespie, supra note 195, at 63–64.
\textsuperscript{249} Id. at 64.
\textsuperscript{250} See Lavi, Platforms, supra note 195.
\textsuperscript{252} Balkin, supra note 69, at 33.
Thus, even images that have historical, cultural, or artistic value are prohibited on their platform. The following three examples demonstrate this:

(1) Nudity with Historical Value - “Napalm Girl”

In 1972, several photographers managed to capture photos of children fleeing from napalm bombing during the Vietnam War. The most famous picture was of Kim Phuc. The picture caught her naked, with napalm burns over her back, neck, and arm. They were not alone; other photographers were present. The photo won the Pulitzer Prize award. As this photo had a tremendous effect on the history of warfare, Tom Egeland, a Norwegian writer, included it in a September 2016 article he uploaded to Facebook. Facebook’s moderators then deleted Egeland’s post, as it included nudity and graphic suffering. Egeland refused to accept the deletion of his post and reposted the image, criticizing Facebook’s decision to take it down. Facebook then reacted by suspending him. But the story does not end here: A daily Norwegian newspaper reported Egeland’s suspension and published the photo. In response, Facebook instructed the newspaper to remove the photo, then went ahead and deleted it. The Norwegian newspaper then republished the picture and criticized Facebook on their front page. The incident further escalated when the Prime Minister of Norway posted the story on his page and Facebook removed his post as well. Only after extensive global news coverage criticizing the decision, Facebook finally reinstated the photo. In this case, Facebook preferred to remove the photo, as it depicts images of fully naked children screaming in pain, which would fall under the category of nudity. However, it would appear that Facebook did not grant enough weight to the photo’s context and its great historical and emotional value.

253. See Evelyn Douek, Governing Online Speech: From “Posts-as-Truths” to Proportionality and Probability, 121 COLUM. L. REV. 759, 774–75 (2021) (“The contestation around nudity on Facebook has been about definition of the category. Activists have had some success in getting Facebook to create limited exceptions to the general ban, but they remain narrow. The classification of a post as nudity is generally outcome-determinative: It comes down.”).
254. GILLESPIE, supra note 195, at 1.
255. Id.
256. Id.
257. Id. at 3.
(2) Breastfeeding

Similar to the “Napalm Girl” case, the breastfeeding case also demonstrates categorical removal of content without addressing nuances of context. In this case, breastfeeding mothers posted photos on their personal profiles, or posted them in support groups for motherhood or breastfeeding. These support groups provide mothers with a forum where they can find advice about the challenges involved in breastfeeding and can belong to a community with other mothers. Facebook removed photos of mothers breastfeeding their babies, arguing that such photos violate their general policy restricting nudity. Some of the mothers whose photos were removed by Facebook wrote angry blog posts and turned to the local press in criticism of Facebook. The mothers felt humiliated that the platform removed photos depicting their experiences, judging these photos as obscene. Following these incidents, women set up a group called “Hey Facebook, Breastfeeding Is Not Obscene!” which drew more and more members. In response, Facebook stated that it would only take down photos containing fully exposed breasts. The controversy expanded until finally, in June 2014, Facebook changed its policy and clarified that the company’s intention is to allow images of breasts for medical or health purposes, such as to raise awareness of breast cancer, and allow photos of women actively engaged in breastfeeding.

(3) Artistic Nudity

Another prominent example of content impacted by the bans in platform terms of service and community policies, that neglect nuances of context is artworks such as paintings and sculptures. Here too, the no nudity policy has been implemented without making any distinctions. Even after Facebook changed its policy regarding nudity in June 2014, declaring it would allow artwork depicting nude figures, controversy remains and Facebook still removes artwork containing nudity in some cases. For example, Elisa Barbari posted a picture of a naked statue of the Roman god Neptune, which is located in front of a church
in Bologna’s Piazza del Nettuno and was created in the 1560s by a Flemish sculptor. The statue has an interesting story: it is believed that after the church forbade the artist from making the sculptures’ genitals bigger, as it was deemed not respectable enough to be placed in front of a church, as a retaliation, the artist redesigned the statue so that Neptune’s hand would point in a certain direction. Consequently, visitors who stand at a specific angle might think that the statue has a huge erection. Facebook removed an image of this famous statue on the grounds it was “sexually explicit” and thus violates the platforms’s guidelines. Facebook added in its statement that “the use of images or video of nude bodies or plunging necklines is not allowed, even if the use is for artistic or educational reasons.” In response, Barbary wrote on her Facebook page “yes to Neptune, no to censorship.” Facebook admitted that the picture did not violate its policy and apologized. Although in the end Facebook referred to the removal as a mistake, it is a consequence of categorical removal, neglecting nuances of context.

In another incident, Facebook suspended the account of a French teacher, Durand-Baïssas, after he posted a photo of The Origin of the World, a famous 19th century painting depicting a naked woman’s vulva and abdomen. In response, the teacher filed a legal claim against Facebook for closing his account. About three years ago, a Paris civil court issued a ruling against Facebook over the censorship claims and concluded that the social media company should not have suspended


266. Facebook for closing his account. About three years ago, a Paris civil court issued a ruling against Facebook over the censorship claims and concluded that the social media company should not have suspended


269. Id. (“A Facebook spokesperson later said in a statement that the censorship was a mistake.”)

the account. However, the court dismissed all damages.\textsuperscript{271} According to Durand-Baïssas’s lawyer, Facebook agreed to make an undisclosed donation to the arts.\textsuperscript{272}

Such removals are not a mere mistake, they represent an ongoing negotiation with the platform, or a public reckoning about society’s shared values.\textsuperscript{273}

(b) Removing Content That Is Inconsistent with the Intermediary’s Agenda

Intermediaries can remove content that is inconsistent with the platform’s agenda and in doing so silence free speech. The case of Prager University (“PragerU”) serves as a good example.\textsuperscript{274} In this case PragerU, a non-profit organization, shared conservative ideas by publishing short videos on YouTube. The videos gained popularity and were viewed by millions of people. YouTube categorized the videos as containing potentially mature content. Following this categorization, it limited access to the videos, and PragerU could no longer monetize them. As a result, PragerU lost advertising revenues. PragerU challenged YouTube’s subjective classification in court on a constitutional basis, arguing that this classification separates PragerU’s videos from other videos and violates its First Amendment rights.\textsuperscript{275} The Ninth Circuit rejected PragerU’s claims, explaining that the “First Amendment prohibits the government—not a private party—from abridging speech’ and neither YouTube nor its parent company Google are state actors.”\textsuperscript{276} The court ruled that YouTube’s internal policies are not subject to First Amendment considerations.

\begin{footnotesize}
\begin{enumerate}
\item See Gillespie, supra note 395, at 142.
\item See Prager Univ. v. Google LLC, 951 F.3d 991 (9th Cir. 2020).
\item Id. at 995. For expansion on this case, see Chase Edwards, Developments in Intermediary Liability, 76 BUS. L. 339, 344 (2021) and Lavi, Publish, supra note 65, at 489.
\item Edwards, supra note 275, at 344 (quoting Prager, 951 F.3d at 991); see Isobel Asher Hamilton, YouTube Isn’t Bound by the First Amendment and Is Free to Censor PragerU Videos, a Court Ruled, BUS. INSIDER (Feb. 27, 2020), https://www.businessinsider.com/youtube-google-censor-court-prageru-first-amendment-2020-2 [https://perma.cc/Y5TQ-YSXJ].
\end{enumerate}
\end{footnotesize}
4. False Positives

As explained above, various motives drive intermediaries to intentionally remove content. It should, however, be noted that in many cases intermediaries remove content by mistake. This occurs when they believe it belongs to one of the categories above, when in fact it does not belong to any of these categories. Such removals are different from a policy of categorical removal, which represents negotiation with the platform regarding shared social values. These removals are mere mistakes and did not concern any type of content that the intermediary intends to police.

When discussing mistakes, we must first address different types of moderation. There are two different kinds of moderation: ex ante moderation, which moderates content before it is published on the platform, and ex post moderation, which interferes with content that has already been uploaded to the platform. Moderation can react to user notices of offensive content or seek such content out proactively. It can be executed automatically using learning algorithms and artificial intelligence, or manually, by human moderators. All methods of moderation may lead to removal of false positives.

Humans make mistakes, and human moderators sometimes make moderation mistakes. There are different types of relationships between human moderators and platforms. Many moderators are contract workers who toil long hours for low wages, for example, outsourcing workers in Nairobi Africa, and outsourcing workers in the United States. These workers have to work quickly to complete their task on time and satisfy their employer's platform requirements.

Gillespie noted that “[b]ecause this work is distributed among different labor
forces, because it is unavailable to public and regulatory scrutiny, and because it is performed under high pressure conditions, there is a great deal of room for slippage, distortions, and failure.” Moderators can thus make mistakes and fail to remove content that the intermediary is interested in removing according to their terms of service (“false negatives”), or remove content that is compliant with the platform’s terms of service and community guidelines, and does not conflict with the intermediary’s agenda (“false positives”).

The problem of moderation errors increases with the use of algorithmic tools designed to moderate content. Algorithm sensitivity to the context of speech is still insufficient, thus, automatic decisions by algorithms are often inaccurate. Since algorithms are not great at understanding indications of context, such as tone or intended audience, using algorithms to detect harmful speech is likely to result in far more false positives than false negatives. False positives have extensive implications, especially due to technological advances that allow intermediaries to identify content that has previously been removed and prevent users from reposting it, or take it down immediately. Prevention of dissemination of such content can extend beyond a single platform. Voluntary cooperation between intermediaries “allows them to share unique digital fingerprints, that they automatically assign to videos or photos of offensive content that they have removed from their websites”. This allows peer intermediaries to duplicate their moderation (i.e., to identify the same content on their platforms and prevent re-posting or initiate removal). Such detection trolls, however, are inaccurate in interpreting the context of expressions. This is especially problematic for content which is text-based, as opposed to images, as the algorithms are insufficient at textual interpretation. As these inaccuracies could lead not only to removal of the original expression by mistake, but also to the removal of shares and replications throughout the web, the end result would be extensive censorship of legitimate speech.

283. Gillespie, supra note 95, at 117.
284. See Keller, Filtering Facebook, supra note 69.
287. Lavi, Platforms, supra note 195, at 567.
B. CDA Section 230’s Immunity—No Intermediary Liability for Content Removal, No Liability for Failing to Remove Content

Section 230 of the Communications Decency Act represents “the mindset of internet exceptionalism, differentiating the internet from the media before...” This Section of United States’ legislation generally blocks lawsuits against online intermediaries for any harm caused by content posted by users. Section 230(c)(1), titled “Protection for ‘Good Samaritan’ blocking and screening of offensive material,” dictates that “[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.” In passing this section of the CDA, Congress basically “declared that online service providers could never be treated as publishers of material they did not directly develop.” Section 230(c)(1) absolves online service providers, including website operators, of primary and secondary liability for a host of claims. Section 230(c)(1) encourages online publishers to exercise their editorial discretion, guaranteeing that online actors will favor some types of content over others. The CDA takes immunity from liability for third-party content, stipulated in Section 230(c)(1), one step further in the ensuing subsection, Section 230(c)(2), which directs that online services are not liable for blocking or removing third-party content. Despite this, defendants in court have tended to rely more on the overall immunity in Section 230(c)(1), since Section 230(c)(2) requires that block-

---

289. 47 U.S.C. § 230; Notably that this Section was exported to Canada and Mexico through the United States - Mexico- Canada Trade Agreement which went into force on July 1, 2020. See Article 19/17 of the U.S. Mexico -Canada Agreement; CITRON, THE FIGHT FOR PRIVACY, supra note 67, at 89 (referring to the agreement and stating that “It pledges signatories to refrain from measures that treat interactive computer services as liable for content created by others.”).
290. Lavi, Publish, supra note 65, at 472.
291. CITRON, THE FIGHT FOR PRIVACY, supra note 67, at 86 (“Section 230 (c ), adopted in 1996 has two key provisions. Section 230 (c )(1) addresses the under removal of content...Section 230 (c )(2), conversely, concerns the over-removal of content.”).
JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET 64-65(2019); Lavi, Publish, supra note 65, at 472.
294. Id.
296. 47 U.S.C. § 230(c)(1)-(2); Kosseff, Users’ Guide, supra note 202, at 769 (“Section (c)(2) provides further protection for moderation, as well as for providing tools, such as website blockers, that allow users to control harmful content.”).
ing and removal decisions be made “in good faith.”

However, even when relying on Section 230(c)(2), courts have provided extensive protections to different internet services. Furthermore, a leading scholar on Section 230, Professor Goldman, has stated that no online provider has lost Section 230(c)(2) immunity through failure to make a good faith filtering decision.

The broad discretion Section 230 offers to social media platforms to screen content and block material has come under scrutiny. President Trump’s Executive Order on “Preventing Online Censorship” pertaining to online platforms, aimed to narrow Section 230’s immunity. The Executive order declared that “[i]t is the policy of the United States that large online platforms, such as Twitter and Facebook, as the critical means of promoting the free flow of speech and ideas today, should not restrict protected speech.” However, the Biden Administration recently revoked this Order and it is no longer in force. A 2021 U.S. Supreme Court opinion also voiced criticism of the broad discretion enjoyed by intermediaries and their power to screen content and block material. The court even made an analogy between private platforms and public accommodators, a stance echoed by legal scholars.

---

297. See Kosseff, Users’ Guide, supra note 202, at 778 (“In a 2020 review of more than 500 Section 230 decisions over two decades, the Internet Association found only 19 that involved Section (c)(2).”); Eric Goldman, The Ten Most Important Section 230 Rulings, 20 TUL. J. TECH. & INTELL. PROP. 1, 6–7 (2017); Eric Goldman, Online User Account Termination and 47 U.S.C. Sec. 230(c)(2), U.C. IRVINE L. REV. 659, 666 (2012) (“Several § 230(c)(2) cases have held that good faith is determined subjectively, not objectively. In that circumstance, courts should accept any justification for account termination proffered by the online provider, even if that justification is ultimately pretextual.”) (hereinafter Goldman, Online User Account Termination).

298. E.g., Goldman, The Ten Most Important Section 230 Rulings, supra note 297, at 6 (referring to Zango, Inc. v. Kaspersky Lab, Inc., 568 F.3d 1169, 1176–78 (9th Cir. 2009)); Danielle Keats Citron, How to Fix Section 230, B.U. L. REV. (forthcoming) (manuscript at 15) (proposing to leave Section 230(c)(2) as is).


301. Id. at Sec. 4(a).


303. Biden v. Knight First Amend. Insr. at Columbia Univ., 141 S. Ct. 1220, 1221 (2021) (Thomas, J., concurring) (criticizing the discretion of social media platforms in Section 230 to screen content (“[i]t seems rather odd to say that something is a government forum when a private company has unrestricted authority to do away with it.”). Justice Thomas concluded that the Supreme Court will have to address current positions in which few digital platforms have unprecedented control over large amounts of speech, and determine how to apply legal doctrines such as common carrier status and public accommodation to “highly concentrated, privately owned information infrastructure.” Id. at 1221; see also Michal Lavi, Targeting Exceptions, 32 FORDHAM INT’L. PROP. MEDIA & ENT. L.J. 65, 140–142 (2021).
presents the argument that “international technology companies no longer shirk responsibility for online terrorist content.” This case could lead to narrowing of the immunity granted in Section 230. Yet at present, unless the intermediary is a governmental website or is operated directly by governmental officials, it will continue to enjoy broad immunity for editorial choices without being subjected to constitutional scrutiny.

C. A World Without Censorship – NFTs as The Engine of Speech

Intermediary abilities to govern speech and define the standard of free speech globally can lead to a chill on legal and legitimate expression. As outlined above, intermediaries have various motives for taking down content. Beyond harmful content, private-public cooperation can lead intermediaries to take down political speech or protest posts against the government. Frequently, intermediaries remove content categorically, neglecting nuances of context, or mistakenly remove legitimate content that does not violate any community guideline. As elucidated above, the problem of “false positives” becomes more acute as monitoring by algorithm increases. These dynamics may transform online intermediaries into engines of unaccountable private censorship. Consequently, there is a danger that extremely valuable speech


307. See supra Part II.A.

308. See supra Part II.A (4).

309. See Bloch-Webha, Content Moderation, supra note 206, at 1305.
could be removed, including political speech, as well as speech that has artistic, historical, or cultural value. This is precisely where NFTs step in and change the rules of the game.

As NFTs create eternal, irreversible, and unalterable expressions which cannot be deleted from the blockchain, they could become a new communications tool and an engine of speech. No more secrets, no more censorship. NFTs could be a means to achieve real freedom of expression, enabling anyone to upload protest posts against governments, to act as a whistleblower and uncover corruption, or to post images of provocative art or images that express their cultural values, without worrying about censorship or being silenced. Because speech cannot be removed from the blockchain, it is immune to censorship. Anyone can create an NFT containing a special expression they want to remain eternally accessible to all blockchain users, thus making their expression unforgettable. Moreover, the more NFTs become integrated with social media\textsuperscript{310} the greater the potential for a censorship-free world, with a multitude of important individual expressions. This will become possible as more blockchain-based social networks develop and as Web 3.0 becomes a mainstream, integral part of social media. When this happens, an open, trusting internet without any central intermediary could be created. This could assist in allowing the free flow of information.\textsuperscript{311} Thus, as stated above, NFTs could truly become the engine of speech and perhaps allow a utopia of free expression advocators – a world without censorship. But, as we very well know from other technological advances, every positive advance has its pitfalls.

III. THE FLIP SIDE OF NFTS: PERMANENCY OF ABUSIVE SPEECH

NFTs have many beneficial uses and the innovative use of blockchain promises to be a game changer for promoting freedom of expression, functioning as an engine of speech. As illustrated in the previous section, by creating a permanent expression that cannot be altered or deleted from the blockchain, NFTs could lead to the realization of a utopian world without barriers or limitations on speech. This utopia could however easily turn into a dystopia. Indeed, a system that makes it impossible to delete the information published on the web can promote freedom of expression. Yet the inability to delete information has a flip side. It could

\textsuperscript{310} For example, NFTs can be integrated to social media by API. See ul Hassan et. al, supra note 14, at 11 (explaining that API can be programmed in a way that could allow blockchain infrastructure to “talk” with the internet infrastructure).

\textsuperscript{311} See Mersch & Muirhead, supra note 48.
also enable the dissemination of abusive speech, without any ability to remove it from the blockchain. Thus, it also has the potential to exacerbate existing types of harm resulting from dissemination of harmful expression.

Although NFTs aren’t yet integrated into search engines like Google, the permanency of tokenized abusive speech could become a major problem when blockchain-based systems become a substantial component in the architecture of tomorrow’s web (Web 3.0). This could cause tremendous harm to victims, of an even more severe type than currently exists. Publication of permanent, unalterable, and unerasable speech could even infringe upon the victim’s own free speech—once their good name has been tainted by an NFT, they may self-exclude from online and even offline forums. Consequently, their ability to engage with others as equals would be curtailed and the public debate suppressed. Furthermore, victims might suffer continuous severe harm to reputation, economic harm, and even emotional distress. In some cases, they may even suffer harassment and violent physical attacks, which might have a vast potential to become an ongoing problem, as the expression is perpetuated on the blockchain. Furthermore, when the tokenized abusive expression concerns a public representative or a public matter, the damage would not only effect the individual, but the public interest as well.

As stated above, courts have ruled that social media companies are not state actors and are not bound by the First Amendment. Therefore, they can remove speech that is protected by the First Amendment, and not just unprotected speech, and overcome structural constitu-

tional constraints. Moreover, as explained in Part II.C., Section 230 of the CDA provides online intermediaries with immunity from civil liability for content uploaded by third parties. Therefore, they are immune from liability for failing to remove harmful content and are also immune for removal of legitimate content. In other words, intermediaries are free to remove expressions, even when the expressions are protected by the First Amendment. As demonstrated, intermediaries have various motives for removing user content from their platforms. Indeed, there is a danger that intermediaries could censor speech in cooperation with governments, or to promote a commercial agenda. Yet, they also play an important role in removing offensive expressions and harmful communication, in order to allow civil communication which is not violent or insulting. In fact, one of the main functions of social media is to enforce community standards and by doing so change the public opinion. As explained in Part II, outlining such norms and enforcing them are crucial to the operation of any platform, which would otherwise be filled with negative value content that is offensive to users and third parties. Thus, even though in practice, content is often removed inaccurately and categorically, and there are frequent “false positives”; existing types of moderation are preferable to the alternative of no moderation at all.

NFTs are relatively new, so we do not know all the ways they could be abused to aggravate harmful speech. The following subsections

316. Chachko, supra note 224, at 111 (“Given the constitutional and institutional constraints on government, government is not as well placed as platforms to directly manage platform-enabled threats and geopolitical challenges.”).
318. See supra Part II.B; Citron, Fix Section 230, supra note 298, at 15 (proposing to leave Section 230(c)(2), as is); Kosseff, Users’ Guide, supra note 202, at 12 ("courts would soon interpret Section (c)(1) to mean that platforms are not responsible for the content that their users post, whether or not they moderated user content . . . Section (c)(2) provides further protection for moderation, as well as for providing tools, such as website blockers, that allow users to control harmful content. The provision states that interactive computer service providers cannot be liable due to “any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected” or enabling or providing “the technical means to restrict access” to such material.").
319. See Part II., supra.
321. See supra notes 193–94.
322. Ashutosh Bhagwat, The Law of Facebook, 54 U.C. DAVIS L. REV. 2353, 2354 (2021) (proposing that the right path to address the inaccuracies and fallacies of moderation is to support “sensible, narrow reforms” and leave it to social media companies to correct the flaws).
demonstrate in detail the types of harm that can be caused by different kinds of offensive content, which are projected to intensify should tokens be abused to disseminate offensive content without mitigation. Current content removal methods would be unsuccessful in moderating blockchain content.

A. Tokenized Speech Abused – A Roadmap of Aggravated Harm

1. Harm to Reputation

a. Defamation

Imagine John Doe tokenizes a social media post on Jane Doe accusing her of yelling racist insults, even though she did nothing of the sort. In addition, and in order to make his claims seem real, John also tokenizes a deep-fake video of Jane yelling racist insults, and the video seems believable, even though Jane did not yell racist insults of any kind. Furthermore, imagine that the post goes viral and generates media attention. It is likely that if the post had not been tokenized, social media platforms would remove it upon Jane’s request, despite the lack of any legal platform obligation to do so under U.S. law. For example, Facebook’s community standards forbid posting expressions that degrade others. Moreover, Facebook aims to remove misleading and manipulated media such as deep-fakes. Because social media plat-

323. See La Liberte v. Reid, 966 F.3d 79 (2d Cir. 2020) (plaintiff filing a defamation suit because social media posts of a T.V. personality suggested that she was a racist).

324. Deep-fakes are believable videos, photos, and audios of people doing and saying things they never did. See Robert Chesney & Danielle Citron, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 107 CALIF. L. REV. 1753, 1759–60 (2019) (explaining the emergence of machine learning through neural network methods that increase the capacity to create false images, videos, and audio and that generative adversarial network can lead to the production of increasingly convincing and nearly impossible to debunk deep fakes).


327. Facebook Community Standards: Misinformation, FACEBOOK, https://transparency.fb.com/policies/community-standards/misinformation/ (last visited Mar. 4, 2023) [https://perma.cc/4NCW-23NJ] (“In other cases, the manipulation isn’t apparent and could mislead, particularly in
forms often use embedded link technology for sharing posts, all the replications generated by users that click “share” would also be removed from the platform, and consequently, be obscured in search queries of Jane’s name. Thus, in the long-term, Jane’s reputational harm would be mitigated.

However, in the case described above, John has abused NFT technology to disseminate defamation. Because the post was tokenized it will remain on the blockchain for eternity, even after it has been refuted. The result would be an ongoing attack on Jane’s reputation, potentially curtailling her professional and economic opportunities. To make matters worse, the post would likely gain extensive exposure with the advance of Web 3.0 blockchain-based social media. Indeed, Jane could file a lawsuit against John under defamation law. However, the damage to her reputation would likely persist as the defamatory expression would remain on the blockchain. Moreover, if the identity of the person posting was anonymous, Jane would have to meet requirements for unmasking their identity. Such requirements would make it more difficult to obtain monetary redress.

b. Shaming

In the previous section, we addressed an instance of defamation on the blockchain. However, another possibility is shaming - John Doe might tokenize a true statement in order to shame Jane Doe and cause her tremendous reputational and dignitary harm. Shaming aims to punish individuals who have deviated from social norms or violated the law. It makes such behavior public, criticizes, and denounces it. Yet, shaming is subjective and is based on individual values. It can target

the case of video content. We aim to remove this category of manipulated media when the criteria laid out below have been met.

328. Lavi, Publish, supra note 65 at 511.
329. RESTATEMENT (SECOND) OF TORTS § 558 (AM. L. INST. 1977). To succeed in the defamation suit Jane has to prove that John posted a “false defamatory statement concerning” her; the publication does not benefit from defamation law privilege; the publisher bears at least negligent fault for the publication; “either actionability of the statement irrespective of special harm or the existence of special harm caused by the publication.”
330. See Nathaniel Gleicher, John Doe Subpoenas: Toward a Consistent Legal Standard, 118 YALE L. J. 320, 344 (2008) (explaining the consideration and standards that U.S. courts apply when considering whether to order John Doe subpoenas). In the context of Blockchain unmasking can become more complicated. For a proposal to allow to unmask cryptocurrency users in a related context see Jabotinsky & Lavi, supra note 11.
certain segments of society just because they are different. Furthermore, shaming can cause disproportionate harm to human dignity, particularly when it reaches a wide audience. In addition, the shameful information is disseminated without due process. The shamed person is often not given the opportunity to justify their behavior and clarify that it does not violate norms as if might have first seemed. Finally, shaming can be taken out of context, and defamatory comments can be added to an existing text. As soon as shaming snowballs, it can even lead to mob justice, harassment and violence.

For example, if John had posted the shaming expression on a social media website without tokenizing it, the social media website might have removed it due to a user report or Jane’s request. Shaming generally does not violate the law, and as long as it is not defamatory or does not belong to an unprotected category of speech, shaming is protected under the First Amendment. Even so, social networks such as Facebook perceive shaming as uncivil and consider it to be a violation of their community standards. Thus they can, and indeed do, mitigate shaming’s potential reputational and dignitary harm. Moreover, in the EU there is a legal “Right to be Forgotten.” Accordingly, a person who does not feel comfortable seeing a post published on him appear in search results can claim that the post is irrelevant and request to delete it. As mentioned, the GDPR includes a specific provision titled “Right to Erasure (‘Right to be Forgotten’)” which imposes obligations on the data controller to erase the requested data. The EU’s GDPR expanded this right, imposing more specific obligations regarding information processing, which are interpreted broadly by the European Court of Justice (ECJ). However, when shaming is tokenized, it remains on the blockchain and can never be erased or altered. Without deploy-

332. Lavi, Good, Bad, Ugly, supra note 199, at 2613.
333. Id. at 2621 (“A person may share videos without knowing the context or the full picture. This may lead to misinterpretation and condemnation of a nonreprehensible behavior”); Daniel J. Solove, The Future of Reputation: Gossip, Rumor, and Privacy on the Internet 83 (2007) (referring to shaming a person for stealing a cell phone, even though he was not the thief; instead, he might have received it as a gift without knowing that it was stolen).
334. Lavi, Good, Bad, Ugly, supra note 202, at 2663.
335. See Community Standards: Bullying and Harassment, Facebook, www.facebook.com/CommunityStandards/bullying (last accessed Mar. 3, 2023) [https://perma.cc/5S9-R536] (Facebook will “remove content that’s meant to degrade or shame, including, for example, claims about someone’s sexual activity.”).
337. Lavi, Good, Bad, Ugly, supra note 202, at 2663.
338. See supra notes 209–211.
340. Id.
ing appropriate cryptographic techniques that will help to minimize the risk to data subjects, NFTs could violate the EU’s “Right to be Forgotten” because “the core immutable ledger technology of blockchain prevents data subjects from exercising these rights”. In other words, if we view the Right to be Forgotten in light of blockchain technology, the current regulations fit poorly to the technology of the blockchain. This is a clear case where legal regulation lags behind evolving technology. Thus, shaming cannot be forgotten if it is tokenized as an NFT, and the harm caused to the shamed individual through NFT abuse would be aggravated, in comparison to the harm that might have been caused by shaming in a traditional social media post.

2. Harm to Public Interest – The Case of Fake News

NFTs can be abused to spread fake news stories and increase their harmful effects. Such harm is not necessarily reputational or dignitary harm, as the fake story could also praise a person. For example, during the 2016 U.S. election campaign, fake news was posted on social networks, reporting that the Pope had endorsed Donald Trump. This post was shared a million times. As fake stories spread across global cyberspace, they pollute the flow of information, and make it more difficult to distinguish between true and false information and conduct a truthful dialogue on matters of public importance. Thus, fake news can be detrimental to politics, democracy, and public interest. Beyond the general “truth bias,” that causes people to believe what they hear, when fake news stories circulate, they come to feel so true that people believe them, even after being presented with evidence of their falsi-

342. Anisha Mirchandani, The GDPR-Blockchain Paradox: Exempting Permissioned Blockchains from the GDPR, 29 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1201, 1204, 1224 (2019) (“It may be argued that an immutable blockchain ledger inherently contradicts the pivotal right to be forgotten, or right of “erasure,” under the GDPR, which is too significant of a right granted to data subjects to be compromised for blockchain technology.”).
343. See id. at 1204.
345. See Omri Ben-Shahar, Data Pollution, 11 J. LEGAL ANALYSIS 104, 105, 112–13 (2019) (treating “fake news” as “data pollution” that disrupts social institutions and public interests in a similar manner to environmental pollution).
346. Lavi, Publish, supra note 65, at 444.
Often, efforts to correct fake news by publishing the truth not only fail to counter the falsehood’s impact, they can even increase its credibility. Even when attempts to correct a falsehood reduce its influence, the effectiveness of such corrections is limited because they are often less widely viewed than the original falsehood. Thus, the best remedy for fake news stories is to remove them from the platform entirely.

Even though fake news constitutes protected speech as long as they do not reach the point of defamation, many social media companies prohibit disseminating them on their platform, and encourage users to create high-quality content. Intermediaries usually prefer to decrease the visibility of misleading content by labelling it or fact checking in cooperation with fact-checking organizations, so as to avoid playing the role of arbiters of truth. However, they can and do delete content, or suspend accounts that repeatedly spread such content, thus impairing the access to it. In contrast, when a fake news story is tokenized, it remains on the blockchain for eternity. Social media companies can label it as false or reduce its visibility by algorithmically downgrading it in searches or newsfeeds, but the most efficient tool-taking the content down - is unavailable.

---

348. Lavi, Publish, supra note 65, at 445 (referring to Whitney Phillips, The Toxins We Carry, COLUM. JOURNALISM REV. (Fall 2019)).
349. See id.
350. Id. at 445.
352. See, e.g., Terms of Service, META, https://www.facebook.com/legal/terms https://perma.cc/X6C7-FQSE (last visited Jan. 29, 2023) (prohibiting users from using the company’s products to do or share anything “unlawful, misleading, malicious, discriminatory or fraudulent.”); Gillespie, supra note 195, at 66 (“Then Google and Facebook barred fraudulent news sites from profiting from their advertising networks Facebook also partnered with fact-checking organizations like Snopes, Politifact, and Factcheck.org to mark some stories as ‘disputed,’ and added a way for users to flag stories they believe to be false.”).
353. Gillespie, supra note 195, at 35 (giving the example of YouTube as a platform that “want[s] everyone to post, but it also partners with professional content producers and offers incentives to high-quality and popular amateurs.”).
354. Id. at 66.
355. Id.
3. Privacy Harm

a. The Problem of Doxing - Violence, Economic Harm and Emotional Distress

An additional problem might arise when personal information is tokenized, in a practice known as “doxing.” Take, for example, a case where John Doe tokenizes Jane Doe's personal information, such as home address, phone number, social security number, and even her credit card and bank account numbers. Posting such personal details about an individual could pose a threat to the victim’s safety and expose them to physical danger, burglary, stalking, and on and offline harassment. Such violations of privacy could lead to economic harm because identity thieves could misuse such personal information to conduct fraudulent transactions in the victim’s name. Finally, such privacy violations could have emotional repercussions which impact the victim’s relationships and everyday life.

To date, the U.S. still lacks a comprehensive federal data privacy law. Moreover, constitutional rights are characterized by negative rights vis-a-vis the state, and there are very few constitutional rights that apply against private actors. However, although in the U.S. there is no general legal obligation to remove private identifying information, many platforms derive their privacy rules from their community guidelines document. For example, according to Facebook's community guidelines, the platform will remove content that “shares, offers or solicits personally identifiable information or other private information that could lead to physical or financial harm . . . .” Thus, if John Doe had posted the private identifying information on a social media website without tokenizing it, the social media website might have removed it in response to a user report or Jane's request.

---

358. See CARISSA VELIZ, PRIVACY IS POWER: WHY AND HOW YOU SHOULD TAKE BACK CONTROL OF YOUR DATA 107 (2020); Citron & Solove, supra note 357, at 834–35.
359. Citron & Solove, supra note 357, at 859.
362. GILLESPIE, supra note 195, at 52.
As mentioned above, in contrast to U.S. law, in the EU citizens benefit from the “Right to be Forgotten.” The GDPR somewhat expanded this right in a specific provision titled “Right to Erasure (‘Right to be Forgotten’)” that imposes obligations on data controllers to erase data. Thus, in the EU citizens have a right to request the erasure of private information that is irrelevant, and are less dependent on voluntary social media community standards to remove such content.

However, when private identifying information is tokenized, it remains on the blockchain and cannot be removed. Thus, victims can find no relief or mitigating factors for posts that violate their privacy and they continue to be exposed to doxing, violence, economic harm, and social distress.

b. Sexual Privacy and Intimacy

Yet another problem could arise in the area of sexual privacy and intimacy. John Doe could tokenize a nude photo or video of Jane Doe without her consent and distribute it on the blockchain. The practice of distributing nude pictures without consent has been dubbed “revenge porn.” Beyond the violation of privacy caused by revealing private information, nonconsensual “revenge porn” is also a violation of the victim’s intimate privacy and particularly sexual privacy.

U.S. law currently lacks the tools to address some of today’s sexual privacy violations. If the victim wants to sue the person who posted their nude photos or videos in a civil court, they will need to invest significant resources. Moreover, victims are sometimes reluctant to sue as they fear that the case will cause the perpetrator to upload even more photos or videos and attract more unwanted attention. Many states have

364. See JONES, supra note 213.
365. GDPR art. 17 Regulation 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (EU).
366. Id.
368. See Citron, Sexual Privacy, supra note 367 at 1918; Danielle Keats Citron, Privacy Injunctions, 71 EMORY L.J. 925 (2022) (explaining in details how non-consensual dissemination of nude photos violates the autonomy of victims, leads to distrust in others and makes it difficult for victims to form intimate relationships, deprives the victim’s social and economic opportunities, causes emotional harm and distress and infringes on the victims’ identity.).
passed laws making nonconsensual porn a crime.\textsuperscript{370} Criminal proceedings are possible, but they require law enforcement to invest resources and time, which only happens when the case is very clear-cut and evidence is strong.

If the nude photos or videos are uploaded to social media platforms, these platforms can somewhat mitigate the harm by taking them down and locking the IP address of the publisher or disseminator.\textsuperscript{371} Indeed, as explained previously,\textsuperscript{372} Section 230(c)(1) exempts online service providers, including website operators, from primary and secondary liability for content published by others. However, websites are still exposed to federal intellectual property claims.\textsuperscript{373} Therefore, “copyright law can provide an effective tool in sexual privacy cases involving the distribution of intimate images created by victims.”\textsuperscript{374} As a result, when victims submit a request, social media platforms are likely to promptly remove images or videos, as their only other alternative could be to face pecuniary damages under the Digital Millennium Copyright Act (DMCA).\textsuperscript{375} Moreover, many platforms voluntarily forbid pornography or nudity in their policies.\textsuperscript{376} For example, Facebook’s terms of service agreement has prohibited the upload of nonconsensual sexual content to the platform since 2014.\textsuperscript{377} Platforms that use embedded link technology, such as Facebook, can indeed remove all the shares of such photos when the original post is removed.\textsuperscript{378}

Intermediaries can further mitigate harm by using learning algorithms to remove images that are identical, and even potentially similar to, content which has previously been recognized as offensive, and thus remove all replications of such content. Since revenge porn is based on images rather than text, technology identifies less “false positives” in this context.\textsuperscript{379} Although intermediaries can and do remove revenge porn images, abusers can and sometimes do repost them. Technology however, has a solution for this problem as well — using hash techniques to prevent the cycle of reposting. Thus, following user reports, if

\textsuperscript{370} See Richards, supra note 61, at 74 (2021).
\textsuperscript{371} Citron, The Fight For Privacy, supra note 67, at 88 (“...Grindr [is] in the best position to minimize the harm by fixing its software so that the app could block IP addresses …”).
\textsuperscript{372} See supra Part II.C.
\textsuperscript{373} Citron, Sexual Privacy, supra note 367, at 1955.
\textsuperscript{374} Id.
\textsuperscript{375} Id.; Digital Millennium Copyright Act, 17 U.S.C. § 512.
\textsuperscript{376} Gillespie, supra note 195, at 52–54.
\textsuperscript{377} Citron, Sexual Privacy, supra note 367, at 1955.
\textsuperscript{378} Lavi, Publish, supra note 65, at 511–12.
\textsuperscript{379} Daphne Keller, Stanford Center For Internet And Society, Dolphins in the Net: Internet Content Filters and the Advocate General’s Glawischning-Piesczek v. Facebook Ireland Opinion 18–19 (2019); Lavi, Publish, supra note 65, at 478.
social media determines that the image violates its terms of service, it designates a unique digital fingerprint – a hash. Such technology can block such images from reappearing on any of its platforms, and holds great promise for mitigating harm caused to victims of revenge porn.

However, when a nonconsensual pornographic image or video is tokenized, most of these technological mitigation efforts are rendered useless. The tokenized image remains on the blockchain and cannot be removed. As the original post cannot be removed, embedded links cannot promote removal of replications. Algorithmic techniques might allow removal of replications, but only if the replications themselves are not tokenized and the social media platform is not blockchain-based. Consequently, the original token remains on the platform and in the case of permissionless blockchain-based social media, equivalent NFTs can be created even if hash technology has been implemented. Thus, when NFTs are involved, victims of tokenized revenge porn are left without any meaningful legal recourse or practical tools to combat the wrongdoing, as the content cannot be removed from the blockchain and continues to inflict harm.

4. Extremist Speech that Encourages Physical Violence, arm to Autonomy and Democracy

a. Hate Speech

Now assume Jane Doe is an alt-right white supremacist who wants to spread hate speech. Using NFTs, Jane can tokenize hate speech against African Americans, Hispanics, Arabs, Jews, or other groups, insulting their ethnic backgrounds and calling for them to leave her coun-

---

380. Citron, Sexual Privacy, supra note 367, at 1955 (describing hashing as “a mathematical operation that takes a long stream of data of arbitrary length, like a video clip or string of DNA, and assigns it a specific value of a fixed length, known as a hash. The same files or DNA strings will be given the same hash, allowing computers to quickly and easily spot duplicates.”).

381. Id. at 1956.

382. See Hate Speech, CAMBRIDGE DICTIONARY (“public speech that expresses hate or encourages violence towards a person or group based on something such as race, religion, sex, or sexual orientation (the fact of being gay, etc.),” https://dictionary.cambridge.org/us/dictionary/english-portuguese/hate-speech [perma.cc/XSR4-WN8X] (last visited Jan. 15, 2023). For more attempts to define hate speech by academics and legal frameworks see generally ANDREW F. SELLARS, DEFINING HATE SPEECH 15–20 (2016), https://cyber.harvard.edu/publications/2016/DefiningHateSpeech [https://perma.cc/YCP3-HZ2Y].
Hate speech targeted at assaulted groups can affect all aspects of their lives and infringe on their autonomy and sense of safety. Expressions such as these can promote stereotypes that then become part of the culture. They cause economic harm by promoting discrimination against members of a particular group, depriving them of opportunities for housing, employment, and services. By denying such opportunities, hate speech can help curtail the group’s participation in democracy and in democratic culture. Finally, hate speech can result in violence and even genocide.

For example, in 2018, the military in Myanmar (Burma) launched a violent hate speech attack on Facebook against the Rohingya minority, inciting violence against them. This online attack resulted in a violent reaction leading to physical attacks that took the lives of thousands of people, including 730 children under the age of five, while 700,000 other Rohingyas saved themselves by fleeing to Bangladesh.

Under Section 230(c)(1), online service providers, including website operators, are immune from primary and secondary liability for content published by others. Moreover, U.S. law does not ban hate speech as long as the speech does not specifically incite immediate violence or a true threat of harm. However, in many countries outside the U.S. hate speech is punishable. In Europe, for example, the European Court of Human Rights has often upheld bans on hateful speech.

Although the protection U.S. law provides against hate speech is very narrow, social media companies have voluntarily adopted policies to reduce hate speech on their platforms. “All social media platforms prohibit hate speech; the few that don’t mention it in a specific rule include it under a broader category like ‘personal attacks.’” For example, Facebook’s

---


385. Id.

386. Facebook Admits It Was Used to ‘Incite Offline Violence’ in Myanmar, BBC NEWS (Nov. 6, 2018), https://bbc.in/3AWST23 [perma.cc/6HEA-NZ57].

387. See supra Part II.B.


389. These countries include France, Germany, and Canada. See e.g., Rotem Medzini & Tehilla Schwartz Alshuler, Dealing with Hate Speech on Social Media: Towards an Interoperable Model for Addressing Racism and Strengthening Democratic Legitimacy, in REDUCING ONLINE HATE SPEECH: RECOMMENDATIONS FOR SOCIAL MEDIA COMPANIES AND INTERNET INTERMEDIARIES 25, 44–45 (Yuval Shany ed., 2020).

390. Id.

391. GILLESPIE, supra note 195, at 58.
community standards forbid attacks which are based on “race, ethnicity, national origin, religion, sex, gender, sexual orientation, disability or medical condition.”

Twitter’s hate speech prohibitions “appear under the ‘hateful conduct’ and ‘hateful imagery and display names’ headings of its speech code.” With respect to hateful conduct, users may not promote “violence against, threaten, or harass other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease.” Platforms can and do remove hateful content when reported. Moreover, some platforms use artificial intelligence to detect hate speech and flag it for human review or removal.

Another tool that platforms use to mitigate the harm of hate speech is to make it less visible using algorithmic tools, which hide posts or tweets. Moreover, platforms learn lessons from past events of harmful speech and improve moderation in the future. For example, following the mass atrocities against the Rohingya minority, Facebook created a Strategic Response team to tackle escalation to violence in conflict areas; “[t]he team has been described as the ‘latest evolution in the Silicon Valley’s culture: less ‘move fast and break things,’ and more thinking through the harm they are adding to half a world away.”

The permanency of NFTs, however, is a major problem where hate speech is concerned. As the expressions cannot be removed, they are likely to continue to spread, exacerbate hate and infringe on the autonomy of minorities and entire races without any mitigation. This could give rise to a risk of perpetuating racism, discrimination, or distorted values. Consequently, the autonomy and dignity of such groups could be diminished. Moreover, such groups could suffer from stigma, economic harm and loss of opportunities, and their right to participate in democracy could be curtailed. Even worse, when hate speech is tokenized and cannot be removed from the blockchain, incitement to violence triggered by such expressions could intensify. Ultimately, this could expose groups to ongoing suffering, violence, physical harm, and even death.


394. The Twitter Rules, supra note 393.

395. Gillespie, supra note 195, at 97 (explaining that most types of hate speech are removed, except for holocaust denial, which is deblocked from some countries).

396. Aswad, supra note 388, at 46 (focusing on Twitter).

397. Chachko, supra note 224, at 80.
Yet another potential detrimental use of NFTs is incitement to terrorism. We already know that terrorists sometimes use anonymous cryptocurrencies to fund attacks. Now any aspiring terrorist can incite blockchain users to commit acts of terrorism. Suppose, for example, that Jane Doe is now a member of an alt-right group. Imagine that she has also managed to tokenize a video of a terrorist attack filmed in real time and even massacres. To this token, Jane has also attached a manifesto inciting more terrorist attacks. In recent years, terrorist organizations have exploited social media to spread fear and disseminate propaganda that reaches potential recruits, incites sympathizers and others, and inspires people who are inclined to radicalization to take part in organizing and committing deadly attacks. The terrorist attack in El Paso, Texas serves as an example of the grave consequences of online incitement. The perpetrator of the El Paso attack, Patrick Crusius, announced the start of his rampage on 8chan’s board, a racist alt-right message board, through a post that included a four-page manifesto. The manifesto and posts on 8chan demonstrate Crusius’s radicalization and turn towards white supremacy.

---

398. Jabotinsky & Lavi, supra note 17, at 526.


402. Lavi, Publish, supra note 65, at 483.
sion that Crusius drew profound inspiration from an earlier manifesto by a Christchurch, New Zealand gunman and the video of his attack. As discussed previously, in the U.S. website operators are immune from primary and secondary liability for content published by others. With respect to the person who incites terrorist attacks, certain forms of incitement are not protected by the First Amendment. These First Amendment exceptions include: promotion of “imminent lawless action”, “true intentional threats against individuals or groups, or posts that seek to cooperate, legitimize, recruit, coordinate, or indoctrinate on behalf of groups listed on the State Department’s list of designated terrorist organizations.” Thus, in traditional social media many expressions of incitement can be regulated by the government. Outside of the U.S., in many other states online platforms do not benefit from an overall immunity and are exposed to litigation for failing to remove posts that incite terrorism.

Despite platforms’ broad immunity from material support claims, social media websites prohibit content that incites terrorism and remove content under their own policies out of an abundance of caution. For example, Facebook bans such content in its community standards, under the general subtitle “violence and incitement[,]” and under the subtitle “dangerous individuals and organizations.” Facebook also removes such content when it appears on the platform.

403. See id. at 483 (referring to Evans, supra note 399).
404. See supra Part II–C.
405. Lavi, Publish, supra note 65, at 529.
406. See id. at 506–07 (reviewing laws that impose obligations on intermediaries regarding inciting content); see, e.g., the U.K Terrorism Act, 2006 c.11 (Eng). Under this law, platforms have only two days to comply with a takedown request; otherwise, they are deemed to have “endorsed” the terrorist content. In addition, the Counter-Terrorism and Border Security Act of 2019 updates terrorism offences for the digital age and grants the authorities more power to tackle the threat posed to the U.K. by terrorism. See also Eur. Comm’n, Press Release, State of the Union 2018: Commission Proposes New Rules to Get Terrorist Content Off the Web, European Commission (Sept. 12, 2018), https://ec.europa.eu/commission/presscorner/detail/en/IP_18_5561 [https://perma.cc/C8ET-RZDS] (describing the European Commission proposal that would require hosting service providers to remove terrorist content online upon an order); Gillespie, supra note 195, at 37.
407. Chachko, supra note 224, at 45.
409. Dangerous Individuals and Organizations, META, https://transparency.fb.com/policies/communitystandards/dangerous-individuals-organizations?source=https%3A%2F%2F (last visited Mar. 30, 2023) [https://perma.cc/C8%K-FzP9] (“In an effort to prevent and disrupt real-world harm, we do not allow any organizations or individuals that proclaim a violent mission or are engaged in violence to have a presence on Facebook. This includes organizations or individuals involved in the following: Terrorist activity... We also remove content that expresses support or praise for groups, leaders, or individuals involved in these activities. Learn more about our work to
Beyond removal of such content due to users reports or take down requests, intermediaries also take proactive steps to detect such content and prevent it from being uploaded again. These efforts are voluntary, or preventive due to concerns about potential legal repercussions. For example, technology companies are currently cooperating to establish a database to detect banned violent content. This database would include unique digital fingerprints of banned content, and files could be flagged and removed instantly. Some technology companies have issued guidelines limiting use of such database to only the most extreme terrorist images that violate the content policies of all companies. Regardless of its final characteristics, these are important steps to create an industry database “to help prevent the spread of violent terrorist imagery.” Another developing tool to mitigate terrorist content is the use of AI to detect such content. These measures can help mitigate the dissemination of content that incites terrorism. Thus, extremist content would be reduced, and less people would draw inspiration to commit terrorist attacks from live-streaming or online manifestos.

However, when terrorist content is tokenized, it is a totally different story as the content cannot be deleted or altered. This means that inciting content could potentially reach larger audiences and goad them on to commit acts of violence. Being exposed to such content might even lead individuals who might not otherwise conceive of committing a terrorist attack to adopt extremist views and act on them. In summary, when speech is tokenized, there is no Right to be Forgotten and no escape. As the innovation of NFTs is relatively new, a systematic understanding of how they can be abused to aggravate the implications of harmful speech has not yet been developed. The above section endeavored to propose such a systematic analysis as to why existing online content removal methods are ill-equipped to confront the challenges of the blockchain.

410. See Lavi, Publish, supra note 65, at 508.
411. Id. (referring to Citron, Extremist Speech, supra note 285, at 1043–45).
412. See id. at 1045 (explaining that according to the guidelines the removal of hashed material would not be automatic but rather subjected to a review by the tech company according to its own specific policies).
413. See Lavi, Publish, supra note 65, at 568–69.
IV. TOWARDS A COMPREHENSIVE FRAMEWORK OF SOLUTIONS AND REMEDIES FOR ABUSIVE TOKENIZED SPEECH

Abusive NFTs are created, sometimes even without thinking of the consequences, that would become more severe with the advance of Web 3.0 blockchain-based social media and the development of the Metaverse. The industry is taking first steps to respond. Thus, Opensea, the main market for trading tokens, is making efforts to delist them from their listings. For example, recently, it delisted NFTs of cartoon characters with allegedly right-wing interpretations, to make it difficult to trade them on this specific marketplace, although the creator can seek other markets. The industry is only starting to realize the problem. The law however lags further behind. To date, there is no comprehensive framework of legal and technological solutions for abusive NFT use. This part proposes mechanisms to bridge this gap, and endeavors to provide a comprehensive framework for mitigating the harm caused by such abuse.

This part also addresses free speech concerns raised by the framework, since the First Amendment protects freedom of speech against government censorship, including even raw data and source code.

A. Safety-by-Design - Ex Ante Prevention

In recent years, technological solutions are being developed and deployed in order to prevent harm inflicted by the free flow of information. The way online platforms are designed can help prevent and mitigate harm, as well as shape attitudes towards violations of law and norms ex ante. There are currently ongoing debates regarding the morality and ethics behind coding of online platforms. Decisions made by engineers can unleash new technology, which may affect fundamental rights such as free speech and the right to privacy. Literature on the influence of technology and its potential to protect privacy is already gaining momen-

414. For example, Lithuanian designer Eriks Mališauskas sold a non-fungible token (NFT) of a digital collage made from homophobic online comments for $6000 with the proceeds going to Lithuanian LGBTQ+ charities. Mališauskas had good intentions, but by creating this NFT he in fact perpetuated hate speech. See Ian Smith, Lithuanian Designer Turns Homophobic Hate Into NFT and Sells It for $6000, GCN (May 6, 2021), https://gcn.ie/lithuanian-homophobic-hate-nft/.


416. See DANIEL SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 100 (2004); see also Lavi, Publish, supra note 65, at 493.

417. See Lavi, Publish, supra note 65, at 494.
The concept of privacy by design advocates regulation of technological design ex ante, rather than providing ex post remedies to victims of dissemination of harmful information. Researchers have described how privacy-protective features can be designed to play a central role in functionality and address threats to privacy.

Regulators around the world have discovered the benefits of privacy by design and set forth guidelines that include privacy by design, alongside efforts to incentivize stakeholders to adopt this approach within their business models. A central example is Article 25 of the EU GDPR, which addresses “data protection by design and by default,” building systems meant to protect and advance privacy starting at the beginning of the design process. Accordingly, at the stage of system development, controllers must implement “appropriate technical and organizational measures” in order to protect data subjects’ rights. “Data pro-

---


419. Hoofnagle, supra note 418; Bamberger & Mulligan, supra note 418; CVOUKIAN, supra note 418.

420. See, e.g., Rubinstein, supra note 418.


423. See Lilian Edwards & Michael Veale, Slave to the Algorithm? Why a ‘Right to an Explanation’ Is Probably Not the Remedy You Are Looking For, 16 Duke L. & Tech. Rev. 18, 77 (2017) (explaining that by doing so, it recognizes that “a regulator cannot do everything by top down control, but . . . controllers must themselves be involved in the design” of systems that minimize invasion of privacy).


(“The protection of the rights and freedoms of natural persons with regard to the processing of personal data require that appropriate technical and organisational measures be taken to ensure that the requirements of this Regulation are met. In order to be able to demonstrate compliance with this Regulation, the controller should adopt internal policies and implement measures which meet in particular the principles of data protection by design and data protection by default.”).
tection by default" is required to ensure that data that is unnecessary for processing is not collected.425

The principle of privacy by design can be applied to NFTs, learning lessons from concepts rooted in privacy, providing ex ante protection against the harm of abusive tokenization, promoting privacy and allowing "safety-by-design" which can mitigate this harm.

1. Minimizing What is Stored on the Blockchain

As reiterated above, tokenized information cannot be altered or deleted and remains on the blockchain for all to see. However, the expression itself does not have to be stored on the blockchain. Instead, the actual expression could be stored off-chain while the token would only store hashes, which are considered pseudonymized data.426 In other words, the digital item itself would be stored in a centralized server, while the token would merely keep a record of who owns that item and a pointer that indicates that the token is tied to a specific digital item.427 Focusing on the design of the tokenization process of marketplaces would allow protection against harmful speech. Storing the information off-chain would allow deletion of the information. As the information is off-chain, it would not be eternalized. Deletion of the information would be possible by “deleting the off-chain stored record while keeping the hash of it intact on blockchain.”428 If the data is stored off-chain, then a data subject could request erasure of the data or file a lawsuit. To delete the information, all one would have to do is “erase any off-chain data that could be used to identify the subject, if linked to the on-chain hash.”429

425. Edwards & Veale, supra note 423 (quoting GDPR, art. 25).
426. Hassan et al., supra note 14, at 17; Aurelie Bayle, David Manset, Octavio Perez Kempner & Mirko Koscina, When Blockchain Meets the Right to be Forgotten: Technology Versus Law in the Healthcare Industry, IEEE/WIC/ACM INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE 788, 789 (2018); see, e.g., Fairfield, supra note 8, at 1272–73

("Often, an NFT stands for ownership of something not directly stored on the block-chain—a piece of digital art, for example. The token contains a pointer to find the digital art file, and a hash of the file as proof. So a token representing digital art might contain a URL pointing to the art and a hash of the art file. In this way, an NFT might convey an ownership interest in a piece of digital art, an asset in an online game, a card in a collectible trading card game (think rare baseball cards here), or a plot of land in a virtual world.").

427. Fairfield, supra note 8, at 43.
428. Hassan et al., supra note 14, at 17.
429. Rose, supra note 341, at 41.
This solution has benefits as it is the best possibility to mitigate the
damage of abusive NFTs. However, this solution also has shortcomings
which should be considered. First, storing the data off-chain might
eliminate some of the benefits of using the blockchain. Storing the
data on an off-chain database would separate the hash that is stored on
the blockchain from the original data. When a person exercises the
“Right to be Forgotten” (in the EU), or when a judicial decision orders
removal of the information, the hash and corresponding personal data
in the off-chain database would be destroyed. Such a design would
impede the benefits of immutability and integrity of information and
accuracy that can be achieved through storage on the blockchain, and
would contradict the very rationale of storing information on the block-
chain in the first place. One could, however, argue that deletion
should not be possible in every circumstance, but rather should require
a judicial decision declaring that the abusive NFT violates the law, and a
judicial order to remove it. Requiring a judicial decision could restrict
this technical solution’s potential to lead to censorship.

Another shortcoming is the problem of cybersecurity and data pro-
tection. Using a third-party centralized entity to store the data could
result in loss of the security achieved when information is stored on a
decentralized blockchain. Storing information outside the blockchain
might also interfere with the transparency achieved when information
is stored on the blockchain, since off-chain storage would make it im-
possible to know who is accessing the database off-chain. Moreover,
separating pointers from information would increase the risk of errors,
cyberattacks, and data breaches. There is also no guarantee that the
company that stores the data will survive – should such a company
cease to exist, the result might be a loss of data. Indeed, storing the in-
formation off-chain would compromise some of the benefits of storage
on the blockchain and would raise cybersecurity and data protection
concerns. Such concerns could, however, be partially mitigated by in-

431. Id.
432. Id.
433. Id. at 1230 (“By storing data subjects’ personal data in the off-chain database, the database
still has the same cybersecurity issues as a regular database”).
434. Id.
435. Id.
436. See Fairfield, supra note 8, at 1283 (“When a digital item is stored on-chain, the art itself is
hashed directly into the token. On-chain storage allows for the item to continue existing even if the
original company hosting the item on its servers no longer exists. On-chain storage thus provides
greater security for a purchaser because the value of the NFT is no longer tied to the continued ex-
istence of any one particular server, or company”).
vesting in securing the systems, and obligating marketplaces to use only storage companies that comply with adequate cybersecurity standards and guarantees regarding the company’s stability.

Alternatively, policy makers could create incentives to use only trusted data storage companies. For instance, fiduciary intermediaries could grant a stamp of approval (such as a trust mark) to marketplaces that use data storage companies that comply with adequate security standards and guarantees regarding company stability. Potential owners that consider tokenizing speech or other items would consider whether the marketplaces have a trust mark, and would prefer to tokenize expressions with a reputable company, to make sure that the information the token points to is less vulnerable to breaches. Such marketplaces would be superior to marketplaces that do not comply with adequate security standards, thus allowing them to gain a competitive advantage.

Finally, storing off-chain draws criticism with respect to notions of property rights and ownership of NFTs that do not store information directly on the blockchain. In such cases, owners have less control over the context in which their own property is used and less control over the property itself. Such a method of storage conflicts with traditional notions of ownership and could lower the value of such tokens. Yet in any case, NFTs have been criticized for upending notions of ownership and uniqueness, even when they are stored on-chain. Thus, one might say that off-chain storage does not change much. Moreover, it is likely that over time, new social norms and conceptions will come into being to address new types of digital property, even if these new norms might challenge conventional notions of property. Moreover, one might argue that off-chain data storage is necessary anyway in cases where the transaction data exceeds “the current storage limitations of the Ethereum blockchain.”

This design would enable deletion of the information on the central server, leaving the NFT with a token that leads nowhere. Deletion would only occur following a court order, when there is prima facie evidence that the speech the token points to is unprotected within the First Amendment, or upon request by the token owner themselves. Since deletion of information would be subject to a court order, it is likely that

---

437. See Lavi, Publish, supra note 65, at 503 (information on trust marks on a related context).
438. Fairfield, supra note 8, at 1283–84.
439. Frye, supra note 84, at 4 (criticizing the concept of NFT).
440. New norms regarding digital property could focus on gaining social recognition rather than control. For expansion, see Frye, After Copyright, supra note 49.
only the worst types of speech that clearly violate laws would be deleted. Thus, such a solution would enable mitigation of the harm of abusive tokens by allowing the technical possibility to delete the information that is stored outside the blockchain. However, as explained the deletion would be limited only to rare circumstances and truly abusive NFTs, thus avoiding the extensive continuing moderation and even public-private censorship, that traditional intermediaries engage.44

One might argue that this proposal infringes on the marketplace’s right to free speech, and its right to conduct business, because marketplaces would be forced to change their technological code, and code is considered speech. However, the proposed framework does not specifically forbid storing information directly on the token. Marketplaces can choose between minimization of what is stored, prescreening, or any other method that would prevent the perpetuation of abusive speech on the blockchain. We believe that many marketplaces might prefer to adopt this proposal even without specific regulation, out of a sense of corporate responsibility,44 and sensitivity to the potential harm that could be caused by permanent expressions in the absence of a Right to be Forgotten. Another reason marketplaces might elect to adopt this proposal might be a desire to distance themselves from murky legal areas and minimize the potential for legal claims such as fines for noncompliance with the provisions of the GDPR, or liability for their role as marketplaces, as outlined above in Part IV. A.1.

One might, however, argue that the very possibility of imposing liability for negligent design would violate the marketplace’s free speech right to shape its systems as it sees fit, even if they may still have a choice between several options. Yet, even if technological tools are treated as speech,444 the value of such speech is not absolute. Programming a tool that shapes NFT marketplaces is not direct participation in the marketplace, rather it is a form of market behavior that utilizes “speech.”445 Due to the commercial nature of code, as part of a product, or a tool in the marketplace, possible liability for negligent design should not be subject to strict
scrutiny. Rather, it should only be subject to intermediate scrutiny. Due to the importance of regulation to reduce unsafe and abusive design that could result in irreparable harm, there are substantial interests in allowing liability. This liability would be content neutral, independent of the nature of speech tokenized and would not discriminate between viewpoints. Storing NFT information off the chain would also not interfere with the overall operation of the system. As such, it is narrowly tailored to pass the intermediary test.

Although the solution of storing information off-chain has its shortcomings, we believe it is more efficient than other possible solutions that will be presented in this Article, and that the benefits outweigh the risks and costs. Many marketplaces may choose to adopt this solution voluntarily, preferring to store information off the blockchain and keep only a pointer to the NFT on-chain. This would be due to a sense of corporate responsibility and sensitivity to the possible dangers of permanent expressions without the Right to be Forgotten. Another reason why some marketplaces might choose to adopt such a solution is a desire to shy away from murky legal areas and minimize legal claims. First, as marketplaces conduct business and tokenize expressions of EU citizens, inter alia, they are often subject to the EU General Data Protection Regulation (GDPR) which regulates controlling and processing personal data. The GDPR applies “…anytime a business, or data controller, collects personal data,” namely data that relates to an identified or identifiable natural person. As NFTs store personal data on the blockchain, the GDPR would appear to apply. Data subjects therefore have the right to request rectification of privately identified information, and have a “right to erasure,” namely they can request that their “personal data be erased from a business’ storage without undue delay.” Furthermore, under the GDPR, data subjects have rights to data portability and they can request that “their data be transmitted from one data controller to another, without any hindrance from the original data controller.” Theoretically, the entire blockchain ecosystem is subject to the GDPR. However, the blockchain’s structure of im-

446. See Citron & Norton, supra note 443.
447. See 2016 O.J. (L 119) 1, art. 17. The GDPR applies "wherever the use of personal data by a business relates to the offering of goods or services to individuals in the EU, irrespective of whether a payment is required, or monitoring of those in the EU." Mirandani, supra, note 337, at 1219.
448. 2016 O.J. (L 119) 1, art. 4.
449. Id. art. 16.
450. Id. art. 17.
452. 2016 O.J. (L 119) 1, art. 20.
453. Mirandani, supra note 342, at 1222.
mutability makes GDPR compliance impossible. Moreover, due to the decentralized nature of the blockchain, enforcement raises difficulties. Marketplaces that strive to reduce the risk of fines due to GDPR violations will therefore opt to store NFT content off the blockchain.

Indeed, it might be argued that hashed data can still be considered personal data, because if an entity possesses such a hash, “it might be able to reconstruct the information and decipher what was stored on blockchain in the first place”. However, it is possible to work around this problem by using “hash peppering whereby a random and secretly kept nonce is appended to the blob of data before taking its hash and storing it on blockchain.” As explained, this method has its shortcomings, yet it increases the likelihood that marketplaces will voluntarily comply with the GDPR.

The second murky area that might incentivize marketplaces to store information off the blockchain is the possibility of lawsuits. If courts were to treat abusive tokens which contain unprotected speech as defective products, the marketplace could be held responsible for the damage. It’s true that Section 230 of the CDA provides immunity for content created by other content providers. However, as scholars argue, the internet has developed beyond a mere communication medium for speech. Therefore, Section 230 should be redefined to meet the reality of marketplaces, which can be the cheapest cost avoiders and should be held liable for their own conduct, or for their contributory liability.

In the context of marketplace liability, different courts currently disagree both about the scope and substance of immunity. Many courts have applied immunity and dismissed a variety of claims against online marketplaces. Yet, recently some courts have allowed such claims to pro-

454. Rose, supra, note 341, at 38 (explaining that if every node on the blockchain was classified as controller, enforcement of GDPR would be virtually impossible, due to the sheer volume of nodes in a permissionless blockchain).
455. See Hassan et al., supra note 14, at 17.
456. Id.
457. See discussion supra Part IV.A.
460. See Agnieszka McPeak, Platform Immunity Redefined, 62 Wm. & Mary L. Rev. 1557 (2021); Gregory M. Dickinson, Rebooting Internet Immunity, 89 Geo. Wash. L. Rev. 347 (2021) (referring to online marketplaces and arguing that “Where a claim is preventable other than by content moderation—for example, by redesigning an app or website—a plaintiff could freely seek relief, just as in the physical world. This approach empowers courts to identify culpable actors in the virtual world and treat like conduct alike wherever it occurs.”).
461. Dickinson, supra note 460, at 377. For example, claims of negligence in Dyroff v. Ultimate Software Grp., Inc., 934 F.3d 1093, 1097–101 (9th Cir. 2019), unfair competition in Marshall’s Locksmith
gress beyond preliminary stages. For example, in *Bolger v. Amazon.com, Inc.*, a defective laptop battery that the plaintiff purchased from a third-party seller through Amazon marketplace exploded, causing the plaintiff physical harm. The court held that Amazon could be held strictly liable because of its role in the transaction. The court did not apply Section 230, because the claim was for product liability, not publication of third-party speech. Similarly, in *Loomis v. Amazon.com LLC*, the California Courts of Appeal recently held Amazon liable for marketplace items. In this case, a hoverboard was sold from a third-party retailer via Amazon’s website. After plugging it in to charge the battery, a fire broke out in the plaintiff’s bedroom, causing the plaintiff burns. The court held that since Amazon is in the “vertical chain of distribution” between the seller and the buyer, it can be held strictly liable. In the same way that marketplaces bear liability for the products they list on their websites, even when they are not the sellers, NFT marketplaces could bear liability for abusive tokens.

The third murky area is liability for unsafe design that encourages violations of law and harmful behavior. In *Lemmon v. Snap Inc.*, three boys lost control of their vehicle and were killed in a fatal accident. They had been speeding at 123 MPH, after using Snapchat speed filter, an app designed to calculate the user’s speed and show it in a photograph. The boys’ parents argued that the speed filter was an unsafe, negligently designed product, and brought a claim against Snapchat. A civil suit was allowed to proceed in the U.S. Court of Appeals for the Ninth Circuit as an independent negligent design claim is independent of Snapchat user messages. Thus, the court differentiated it from claims that focus on content published by other content providers. On remand from the Ninth Circuit’s Section 230 denial, the *Lemmon v. Snap, Inc.*
mon district court denied Snapchat motion to dismiss and decided that the Plaintiffs have adequately alleged that the design of the Speed Filter itself encouraged Plaintiffs to engage in reckless driving. This case demonstrates that Section 230 cannot protect online marketplaces from all types of claims. This exposure may lead them to seek ways to reduce litigation and possible liability. In the context of NFT marketplaces, ex ante storage of information off the blockchain might help provide them the protection they seek.

2. Prescreening Tokens and Preventing NFTs Containing Unprotected Speech from Entering the Blockchain

A second solution by design is prescreening abusive tokens that contain speech that is unprotected within the First Amendment before they enter the blockchain. Such a solution would make it difficult for abusive speech to enter the blockchain in the first place and would reduce the likelihood of perpetuating abusive speech. Unlike the previous design-based solution, this solution would not curtail the benefits of blockchain immutability and integrity. As it would enable full storage of NFTs that have not been prescreened out, it does not raise cyber security concerns and it conflicts less with traditional notions of property. The blockchain itself is unable to pre-learn works before adding them. The blockchain can safeguard against data tampering, but it cannot guard against inaccurate or unverified data. Marketplaces that allow the creation of NFTs, however, are able to do so and have the ability to prevent the creation of abusive tokens.

Despite this, prescreening has more shortcomings than the previous solution. First, it does not provide an overall solution for the problem of perpetuating speech. Once a pre-screened NFT enters the blockchain it still cannot be deleted at a later stage. Thus, this solution still does not provide a Right to be Forgotten and would therefore be incompatible with the provisions of the GDPR and unsuitable where EU
citizens are involved in the transaction. Moreover, as the context changes the meaning of an expression can change too. A statement that was true at the time of tokenizing could constitute defamation at a later stage, when time passes, or when new facts are revealed. Thus, such a solution could still potentially allow the perpetuation of defamatory, and other unprotected speech, as the decision to allow the token to enter the blockchain is based on the context and facts known at a specific point of time.

Prescreening also curtails the benefits of the blockchain as a decentralized system: by assigning the roles of mediation and moderation to the marketplaces, it in fact creates new gatekeepers. Even though marketplaces are external to the blockchain system, one might argue that gatekeeping by marketplaces, making them the arbiters of what does and does not enter the chain, would conflict with the idea of crypto-democracy, which is based on the absence of intermediaries. Such a solution would replace traditional intermediaries and shift power to marketplaces, which would come to control the type of information included on the blockchain.

Beyond the conceptual conflict with the idea of crypto-democracy, assigning marketplaces the role of gatekeeper could de facto result in chilling legitimate expression and private censorship. Much as described in Part II with respect to online intermediaries, marketplaces, as private entities, are likely to pre-screen protected speech, to elect to avoid murky legal areas and comply with laws both in and outside the U.S. It is reasonable to assume that they would be likely to do so in cooperation with the government and to bar speech that is not compatible with their agendas or their community standards. Indeed, like intermediaries, different marketplaces would adopt different policies regarding barring expressions from being tokenized, and a degree of diversity would exist. One cannot, however, ignore the possibility that such a practice could chill legitimate protected speech, make it difficult to perpetuate it, and undermine the benefits of NFTs as the engine of speech. Moreover, even if marketplaces aim to weed out only unprotected speech, just like online intermediaries, marketplaces would not have the capacity to prescreen the high volume of content tokenized manually and

475. See Lavi, Good, Bad, Ugly, supra note 202, at 2640.
477. See supra Part II.A.1.
478. See supra Part II.A.2.
479. See supra Part II.A.3.
would need to use algorithms. As explained, this method of automatic prescreening results in inaccuracies in identifying the context of messages, leads to many “false positives”, and can chill protected speech. In short, both design-based approaches proposed above have a certain degree of potential to reduce harm caused by perpetual, abusive NFTs ex ante. Storing all information off the blockchain and using pointers, or prescreening only NFTs that contain unprotected, abusive expressions, are potential solutions to the problem of perpetual tokenized abusive expressions. Each of these possibilities has its benefits and shortcomings. We, however, believe that the solution of storing all information off of the blockchain is preferable, as it would allow deletion of information at any given point of time and not only at the date of tokenization. And yet, it remains compatible with the decentralized crypto-democracy idea, preserving the function of NFTs as the engine of speech.

It should be noted that similarly to the solution of storing information off-chain, the solution of prescreening tokens also give rise to free speech concerns. One might argue that legislation imposing liability for a failure to prescreen abusive NFTs is not content neutral and should be subject to strict scrutiny. However, this proposal is directed only at unprotected low-value speech that does not benefit from First Amendment protections. Yet, it could still be argued that even so, mandating prescreening of only unprotected speech raises constitutional problems, as such algorithmic prescreening could erroneously weed out NFTs that include legitimate protected speech as well. Algorithms are inaccurate in identifying the context of messages, resulting in “false positives,” which could infringe on the freedom of expression of NFT owners. All that may be true, but if this indeed occurs, courts would be able to instruct the marketplaces to reinstate the content as an NFT where needed. Indeed, we are well aware of the fact that the solution of prescreening could raise more First Amendment concerns than our first proposal if it is imposed on marketplaces through mandates. However, if marketplaces elect to prescreen tokens, this is their choice.

480. See supra Part II.A.4.
483. Id.
and they would be responsible for conducting risk management. This is not very different from moderation of intellectual property violations, that can also be seen as a type of speech, but are legally subject to a type of notice and takedown regime. Nor is it different from moderation in states that do not grant immunity to intermediaries.

Marketplaces can choose not to prescreen tokens and adopt the first proposal instead. Marketplaces that store the information elsewhere, separate from the NFT, and include only a pointer on the NFT, would be exempt from the proposal to prescreen. Potential NFT owners could also choose between marketplaces that use different mitigation methods: prescreening or storage off the blockchain.

**B. Ex Post Mitigation- Obscuring Harmful NFTs**

1. “Superseding” Old, Inaccurate Data with New Data on the Blockchain

This Article focuses on technological mitigation directed towards NFT marketplaces that can store all content off the blockchain, or at least prevent the tokenization of abusive speech. Such ex ante solutions are preferable to ex post solutions. However, although we believe that ex post solutions are less efficient, they can still provide a level of mitigation when ex ante steps have not been taken. The following subsections will focus on ex post mitigation through obscuration.

Although tokenized information on the blockchain cannot be deleted or altered, the victim can still obscure it. In other words, they can make the information “relatively difficult to find, or understand.”

This concept of obscuration was first defined by Hartzog & Stutzman,

---

484. See Digital Millennium Copyright Act, 17 U.S.C. § 512 (imposing a notice and takedown regime with respect to intermediaries liability regarding intellectual property infringements).
485. See e.g., Network Enforcement Act, Netzwerkdurchsetzungsgesetz [NetzDG] [Network Enforcement Act], Oct. 1, 2017, at § 3(2)(4) (Ger.).
486. Marketplaces might prefer this option over storing the information off the blockchain and using a pointer because storing on the blockchain does not require them to secure a separated database, and it allows a structure that is more in line with traditional conceptions of ownership. See discussion supra Part IV (explaining the security problems of storing information outside the blockchain and on the benefits of doing so from a property law doctrine perspective); see also Mirchandani, supra note 342; see also Fairfield, supra note 8.
487. Hartzog & Stutzman, supra note 418, at 188 (“[T]he concept of online ‘obscuring,’ defined here as a context in which information is relatively difficult to find or understand, is a much more defined and attainable goal for social technology designers”).
The scholars understood that the risks posed by harmful information cannot always be completely eliminated: they can, however, be mitigated if the information becomes less accessible. Obscured information lacks essential factors which are necessary for discovery or comprehension. Literature has identified four such factors which can be obscured: (1) search visibility, (2) unprotected access (free, universal access to the information), (3) identification (the ability to connect the information to a specific person), and (4) clarity (complete information that is not vague). Limiting the visibility of information, restricting access, obscuring the identity of the data’s subjects, or publishing partial information can all reduce the potential harm of exposure to the information.

Obscuration can help mitigate the harm caused by information that has entered the blockchain. As data cannot be removed from the blockchain, victims of harmful data cannot request removal of their personal data, even when the blockchain is a permissioned or private blockchain, meaning one would need to register and gain permission to become part of the network. Such private blockchains are distinct from permissionless networks that are open to the public and allow anyone to participate. On both permissioned and permissionless blockchains, the information cannot be deleted or altered.

On permissioned blockchains however, NFTs can be obscured and made less visible or accessible. This can be done by “superseding”: a process that replaces old, inaccurate data with new data. This process works as follows: Permissioned blockchains record transactions with a

489. See Hartzog & Stutzman, supra note 418.
490. Id. at 32–39.
491. See Melody Moh, David Nguyen, Teng Sheng Moh & Brian Khieu, Blockchain for Efficient Public Key Infrastructure and Fault-Tolerant Distributed Consensus, 79 BLOCKCHAIN CYBERSECURITY, TR. AND PRIV. 69, 83 (Kwang et al. ed., 2022); see also Horst Treiblmaier, Toward More Rigorous Blockchain Research: Recommendations for Writing Blockchain Case Studies, in FRONTIERS IN BLOCKCHAIN 2, 3 (May 2019) (explaining that permissioned blockchain can limit who can publish on the blockchain and controls the access to information as configuration is performed by a trusted user who also controls runtime network reconfigurations).
492. See Dirk A. Zetsache, Ross P. Buckley, & Douglas W. Arner, The Distributed Liability of Distributed Ledgers: Legal Risks of Blockchain, 2018 U. Ill. L. Rev. 1361, 1372 (2018) (“[P]ermissioned systems require an organization and governance structure regulating at least who is permitted to participate and usually the basis upon which they may participate...In contrast, permissionless blockchains such as Bitcoin operate on public domain software and allow anyone who downloads and runs the software to participate.”).
493. Mirchandani, supra, note 342, at 1223 (“The most that a permissioned blockchain node can do is supersede old, inaccurate data, with new data. It cannot alter or delete previously entered data. Even though old data is not physically deleted from the chain, it becomes a part of an older block and is not added upon directly and monitored by nodes.”).
Every blockchain has a special algorithm that scores different versions of the history. When one of the participants on the blockchain receives a higher scoring version “they extend their own database and transmit the improvements to their peers.” As blockchains are built to add the score of new blocks onto old blocks, older entries are likely to be superseded. In such cases, the old data physically remains on the blockchain. However, it becomes part of an older block and is “not added upon directly and monitored by nodes.” When the blockchain is permissioned, superseding abusive NFTs by adding new information to the blockchain upon the victim’s request could be an option to reduce the abusive token’s potential for harm. Although this would not eliminate the damage of harmful speech, it could however substantially reduce it, as the expression would remain local. Superseding the information would obscure the information, reduce its visibility and accessibility, and decrease the likelihood that the information would spread to a wider audience.

This solution can be likened to the practice of reputation management systems and search engine optimization, which obscure harmful information by bombarding the internet with information that downgrades the page-rank of the harmful information. Yet, this method of mitigation will only work if regulation is put in place, instructing search engines to collect data only from the most recent blocks on the blockchain. Such obscurcation would not raise First Amendment concerns, as superseding is the victim’s own self-help initiative and it is not in any way limited by state or federal legislation.

2. Harnessing Online Intermediaries to Obscure Abusive Tokens on the Level of The World Wide Web

Right now, tokenized speech posted on the blockchain might result only in limited harm, because it is exposed only to a limited audience – blockchain users. Indeed, blockchain-based search engines are developing, yet they are directed towards blockchain users and the results
do not yet appear on prominent search engines. When NFTs integrate into the internet via API, that allows computers and applications to communicate with one another,⁴ ninety NFTs start to appear on social media, and will also begin to appear in the search results of general search engines. This means that everyone will be able to access and see them. Abusive NFTs will become easier to search, access and view if Web 3.0 blockchain-based social media becomes more popular. Such a situation would make it easier to share NFTs with a wide audience and gain vast circulation and credibility.⁵

The practice of obscuring, introduced in the previous section, can be useful not only for obscuring information on the blockchain, but for making it less visible, accessible, or searchable on the level of the World Wide Web. Such a solution would not eliminate the potential damage of harmful speech. However, it could substantially mitigate it, as the expression would remain local. Superseding the information with new information would reduce the likelihood that the older information would spread to a wider audience. Making information “relatively difficult to find or understand” in fact obscures it.⁶

Spreading information has many benefits, as explained previously. However,⁷ dissemination can also be devastating when harmful information is spread and cannot be deleted. Yet, deleting content is not the only available remedy.⁸ Much like abusive speech that has been tokenized, online intermediaries can obscure abusive speech, reduce its visibility, and make it harder to search. There are various ways to do this. First, intermediaries can “place a ‘no-index’ tag on a page so that the page does not appear in external search indexes like Google, even though it remains fully accessible on the service.”⁹ This remedy can be likened to the Right to be Forgotten, which obscures access to the information.¹⁰ Another option is to downgrade the visibility of the integrated token on the internal search results page of the website with which

---

499. On API, see generally Gazarov, supra note 44.
501. Hartzog & Stutzman, supra note 418.
502. See supra Part III.
503. See Eric Goldman, Content Moderation Remedies, 28 MICH. TECH. L. REV. 1 (2021) (introducing the toolkit of moderation that include much more than taking down context).
504. Id. at 32–33.
it was integrated, block auto suggestions and completion of search queries with the token’s name in this search engine. Intermediaries can also use AI algorithms to reduce the visibility of abusive NFTs on newsfeeds, or reduce virality, thus reducing the ability to disseminate them widely.

Obscuring is not optimal, as harmful expressions remain on the blockchain. However, this practice can reduce their exposure and provide some relief.

One might argue that this obscuration could chill free speech and infringe on the token owner’s right to free expression and the public’s right to receive information. Indeed, mandating intermediaries to obscure abusive tokens on the internet raises more challenges as Section 230 immunizes intermediaries from liability. Although obscuring is different from removing content altogether, and would lead to a milder chilling effect, such proposed mandates could conflict with Section 230, since they impose liability on editorial discretion exercised by intermediaries regarding third-party content. However, we believe that the interpretation of the law should adapt to a changing reality which includes emerging technologies and their implications. In light of the unprecedented damages abusive tokens could potentially cause, a narrow interpretation of Section 230 should be adopted, in order to allow the imposition of a nuanced obligation to obscure. This is a relatively moderate practice, since it does not involve deleting content altogether, but merely reducing its visibility.

It should be noted that outside the U.S. this solution could be made mandatory. Obscuration obligations already exist under current regulation in related contexts throughout the world, as in most states freedom of expression does not take precedence over the right to privacy. In fact, a similar solution has already been applied to search engines regarding information on EU citizens. In the EU, citizens benefit from the “Right to be Forgotten”. Upon request, search engines such as

---

506. Brunton & Nissenbaum, supra note 505, at 44.
507. Id. at 45.
509. See generally Part II.B (on obscuring in a related context of the right to be forgotten).
510. Neil Richards & Woodrow Hartzog, Privacy’s Constitutional Moment and the Limits of Data Protection, 61 B.C. L. Rev. 1687, 1729–30 (2020) (“In Europe, free expression is safeguarded by Article 10 of the European Convention and Article 11 of the EU Charter. Like other European fundamental rights, these provisions are subject to proportionality analysis – where they conflict with another fundamental right such as the right to privacy or to data protection, courts must balance the rights on an equal footing. By contrast, in the United States, the fundamental right of free expression protected by the First Amendment is not subject to proportionality analysis. . . .”).
511. See supra Part II(1).
Google are required to delist search results that link to information that is personal, inaccurate, or irrelevant, as well as excessive data found on third-party websites. 512 Such model is a model of obscuration, since the data remains on the website of origin and is only removed from search results. Despite differences between the EU and the U.S. with respect to the status of freedom of expression, in the context of the “Right to be Forgotten,” scholarly work advocates the adoption of nuanced obscuration obligations even in the United States. 513 Such obligations become particularly critical in the context of NFTs.

Indeed, delisting search results that link to information that is personal, inaccurate, or irrelevant, can curb free speech, even though it only obscures the expression without removing it. A limited right to delist does, however, provide a proportional balance between values and can even promote speech. First, an abusive NFT could cause the victim to seclude themselves from conversations and delisting it could in fact achieve a balance between the free speech of the owner and the victim. Second, obscuring information does not delete it. The information would remain on the blockchain, and even on the internet, and would not be removed completely. As a result, there would be less of a chill on the token owner's speech. Finally, although NFTs do indeed contain speech, there could be some debate about whether the value of this speech should be treated as absolute. “With time, data may express fewer elements of free speech” and can be removed from the original context. 514 Therefore, the law should legislate nuanced protection of speech which accounts for the extended period of time that may have lapsed since the token's creation, as well as the differences between deletion and obscuration.

As explained above, since it is impossible to remove expressions once they have entered the blockchain, there is justification for a mandatory obligation to obscure directed at intermediaries. However, even if such a mandate were to be adopted, it should not be automatic, and should have limitations. Obligating intermediaries to obscure based on victim allegations of violations would result in extensive collateral censorship, as anyone might argue that an NFT violates their rights. Therefore, in order to avoid such consequences, a duty to obscure should be imposed only after a judicial decision has determined that the token does indeed violate a legal right.

513. See JONES, supra note 213, at 164 (“Even in the US there are ways to make the digital age a forgiving era. But the ways must be within the boundaries of what makes Americans most free.”).
514. Lavi, Good, Bad, Ugly, supra note 202, at 2644.
We are well aware of the challenges that could be posed by mandating obscuration of NFTs that contain unprotected speech. Accordingly, we propose that if binding regulations are not promoted, intermediaries could voluntarily adopt the solution of obscuring content on their platforms. Intermediaries already apply similar voluntary measures in other contexts and adopt obscuring practices in their community standards and moderation policies. In these contexts, intermediaries use algorithms to downgrade some types of content and make them less visible. Intermediaries could use the same practices to voluntarily obscure abusive NFTs.

C. Prosecution or Lawsuits Against Abusive Token Owners – Remedies

The third proposed method of mitigation is direct lawsuits or prosecution against the token owner. Unlike online intermediaries and individuals that share or transfer the original token, which are immune to liability, the victim can file a legal claim against the person who created the token. This option does not come to replace the technological solutions proposed above in Part IV. A. It cannot prevent the harm ex ante, as the NFT would remain on the blockchain, available for blockchain users to see. Unlike obscuring, direct legal action against the owner also would not mitigate harm. It can, however, provide some recourse against the damage caused by abusive NFTs, or deprive the owner of profit from the abusive token. Thus, it has the potential to partially rectify the injustice and allow relief. Moreover, the threat of possible damage claims or disgorgement penalties, that could strip the owners of abusive NFTs of every penny from ill-gotten gain, could constitute a substantial negative incentive to avoid creating and transferring abusive tokens in the first place. However, as the following subsection will explain, substantial and procedural difficulties stand in the way of prosecution or direct legal claims.

515. See supra Part III; see generally Goldman, supra note 503.
516. 47 U.S.C. § 230; see Barzel v. Smith, 333 F.3d 1018 (9th Cir. 2003). (Section 230 also exempts internet users who share content published by others; see also Barrett v. Rosenthal, 146 P.3d 510, 525, 528–29 (Cal. 2006) (“[C]ongressional purpose of fostering free speech on the Internet supports the extension of section 230 immunity to active individual ‘users.’”).
1. Difficulties and Obstacles to Direct Lawsuits

   a. Harmful Expressions Do Not Necessary Violate the Law

   The first difficulty is a substantial one: in order to prosecute or file an action against the NFT's creator, the NFT's content has to violate the law. Thus, even if an NFT has caused tremendous harm, and has violated the community standards of all mainstream platforms, the police would not interfere, and the victim would not have a direct cause of legal action, unless the NFT violates the law. Not all categories of behavior and damages presented in Part III violate laws, especially in the U.S., where there is a presumption against restrictions of speech. This is because freedom of speech enjoys stronger protection in the U.S. than in other democracies. A legal action can therefore be successful mainly regarding categories that constitute wrongdoing or criminal offenses.

   Abusing NFTs to commit defamation is unlawful, since defamation is considered unprotected speech. Therefore, victims could file an action against the person who first posted an NFT containing defamation against them. Shaming however, is different: if the NFT contains shaming but does not reach the level of defamation, it can be posted without legal liability. This is because shaming is not considered wrongdoing, even if intermediaries do often voluntarily remove shaming that is not tokenized from their social networks. Fake news that does not reach the level of defamation is also considered protected speech. Therefore, even if NGOs might have legal standing regarding an infringement of public interest, the lawsuit is likely to fail. Victims can file actions regarding violations of privacy that meet the requirements of Prosser's identified privacy invasions. These violations fall under the


520. See KILLION, supra note 315.

521. See supra Part III.A.1.b.

522. United States v. Alvarez, 567 U.S. 794, 718–19 (2012); Louis W. Tompros, Richard A. Crudo, Alexis Pfeiffer & Rahel Boghossian, The Constitutionality of Criminalizing False Speech Made on Social Networking Sites in a Post-Alvarez, Social Media-Obsessed World, 31 HASTINGS J. L. & TECH. 65 (2017). For criticism, see CASS R. SUNSTEIN, LIARS: FALSEHOODS AND FREE SPEECH IN AN AGE OF DECEPTION 48 (2021) (“the plurality in Alvarez was myopic in focusing largely on established categories of cases, such as defamation, in which false statements of fact can sometimes be regulated or sanctioned. In the modern era, false statements falling short of libel are causing serious problems for individuals and society; if they cause such problems, there is a legitimate argument that they should be regulable.”).
Restatement (Second) of Torts. Action can therefore be filed for intrusion upon seclusion, or public disclosure of embarrassing private facts. Victims can also file action for infliction of emotional distress. When nude photos are tokenized and disseminated without consent, victims can file actions based on privacy laws, infliction of emotional distress, or specific laws that address revenge porn. Surprisingly, speech that encourages physical violence is not always a violation of the law. For example, U.S. law does not ban hate speech unless it specifically incites to imminent lawless actions, such as incitement to terrorism, or includes a true threat of harm.

Even before the creation of NFTs, victims suffered harm from hate speech, and lacked legal remedies. Yet tokenizing speech can aggravate this harm substantially, and victims remain empty handed. The problem will get worse with the advance of Web 3.0 blockchain-based social media. In fact, once the harmful expression has entered the blockchain, the only way to mitigate harm is to try to obscure the information by adding more information in the case of permissioned blockchain, or by harnessing online intermediaries to make it less visible.

b. The Problem of Recourse, Reluctance to File Legal Action and Collection of Monetary Remedies

Even where a direct legal cause of action regarding the abusive token is present, filing a legal action requires significant resources. Criminal procedures require law enforcement entities to invest resources, and they do so only in very clear-cut and extreme cases. Private civil suits require victims to invest significant resources and, in many cases, victims cannot afford to hire a lawyer. Victims may be reluctant to take legal action in some cases, such as invasions of privacy and violations of a sexual nature. In these instances, victims of abusive NFTs may fear that a civil suit could attract more unwanted attention to the

525. See Nonconsensual Pornography Laws, Cyber C.R. Initiative (last visited Jan. 29, 2023), https://cybercivilrights.org/nonconsensual-pornography-laws/ (demonstrating that as of today, at least 48 states have revenge porn laws that make it a crime to disseminate such photos); Franks, supra note 367, at 1256; Citron, Sexual Privacy, supra note 367, at 1932.
527. See Citron, supra note 367, at 1930.
528. Id. at 1929–30.
harmful expression and exacerbate harm. Therefore, they might elect to refrain from taking action.

Furthermore, legal proceedings can be complicated, cumbersome, and lengthy. Many requirements must be met for a court to reach a decision, and the contexts in which remedies are available can be limited, as some causes of action require plaintiffs to prove damage, or to demonstrate evidence that the defendant has profited from the abusive speech.

c. The Problem of Anonymity and the Procedural Difficulties of Unmasking the Violator

Even if wrongdoing or criminal offenses have been committed, and even if authorities or the victim of the abusive NFT, are willing to invest the necessary resources to prosecute or file a lawsuit and embark upon a cumbersome legal process, in many cases the victim faces yet another obstacle – token user anonymity.

NFT owners do not have to identify themselves on the blockchain. The token’s anonymity makes it possible to hide and protect the identity of the chain of owners. The original creator’s identity might be known to the marketplace (usually due to regulation that requires this); however, if the owner does not identify themselves voluntarily, their identity will remain unknown to blockchain users, as will the identities of subsequent buyers.

In previous work, we expanded on the right to anonymity as a constitutional right. We criticized such anonymity in the context of fungible cryptocurrencies. We proposed that corporations that issue cryptocurrencies be required to verify the identity of blockchain users. In order to balance between legitimate privacy, free speech interests to maintain anonymity and crime prevention, we proposed that unmasking blockchain user identity should be subject to a court warrant.

Similarly, we believe that NFT owner identity should be verified by marketplaces, as well as on the blockchain, and that the law should allow unmasking of NFT owners. In order to balance privacy, free speech interests, and the interest of protecting against violations, unmasking should

529. Lavi, Good, Bad, Ugly, supra note 202, at 2622.
532. Id. at 569.
not be a routine matter. Instead, it should be subject to a judicial decision and balancing tests.\footnote{See id. at 577–79 (expanding on unmasking cryptocurrency users and the right to anonymity).} Otherwise, individuals would avoid creating and transferring NFTs for the beneficial purposes described in Part I, such as artwork and collectibles. Moreover, automatic unmasking of NFT users would impair their function as “the engine of speech.” If users were to know ex ante that they could be exposed to extensive government and private intervention and could be unmasked for tokenizing speech that is protected under the First Amendment, they would be disincentivized from using NFTs to perpetuate their speech. Consequently, they might avoid using NFTs to tokenize political speech, words of protest, criticism of governmental institutions, or whistleblower leaks, without being censored by governments or private intermediaries. Without limitations on unmasking, such as subjecting it to a judicial decision, the innovation achieved by NFTs would grind to a halt, and society would be deprived of its benefits.\footnote{Id. at 582 (referring to a related context of unmasking cryptocurrency users and explaining that without limitations unmasking obligations could hinder innovation: “such obligations might impair user trust in the system and hinder innovation.”).}

In the related context of using fungible cryptocurrencies to fund illicit activities, we proposed that unmasking fungible tokens should be subject to a warrant.\footnote{The conclusion that a warrant is needed is based on the U.S. Supreme Court decision in Carpenter v. United States, 138 S. Ct. 2206 (2018).} Where there is probable cause to suspect that a cryptocurrency user’s activities support crime, or infringe on national security, courts should order cryptocurrency issuers to disclose the identity of users on the blockchain.\footnote{Jabotinsky & Lavi supra note 17, at 564–65.} Similar standards should apply where there is suspicion that a person is using an NFT to commit or support crime.

Furthermore, we propose that procedures should also be developed to unmask NFT owners in the context of civil law. Similar procedures have been developed regarding unmasking of cybersmear on the internet.\footnote{Kosseff, supra note 530, at 93–111.} Obligations to store information on users, such as IP addresses, and general unmasking procedures, are not uncommon in other civil law contexts, and U.S. scholars proposed to condition Section 230’s immunity to content providers on maintaining IP logs.\footnote{Id. at 137–39.} People posting defamatory comments on websites might thus be exposed to unmasking of their Internet Service Provider (ISP).\footnote{Jabotinsky & Lavi, supra note 17, at 583.} However, careful thought is needed, as
overly broad unmasking of NFT owners could chill protected expressions under the First Amendment. Similar to the widely employed standard for unmasking in traditional defamation actions,540 unmasking should be subject to a judicial decision, and should be exercised only after identifying precisely which statements constitute a tort. Additionally, plaintiffs should be required to produce prima facie evidence supporting every element of the claim. The court would be tasked with balancing between the risks connected with unmasking the defendant and potential harm to the plaintiff, in light of the specific circumstances of each case.

Courts should weigh “the risk of unmasking the defendant against the harm to the plaintiff on a case-by-case basis.”541 There should be a need to weigh similar considerations when police or government suspicions arise that an NFT supports crime. Courts would only order unmasking where there is probable cause that the NFT supports crime.

To illustrate the above, imagine that John Doe subpoenas unmasking the identity of anonymous speakers from their ISP, or from a website on which they have posted defamatory comments that are not new. The idea of unmasking anonymous users should be all the more applicable to NFT owners who not only publish harmful expressions, but also perpetuate them on the blockchain. As unmasking is possible only under specific circumstances, and the anonymous speaker’s right to free speech is balanced against the plaintiff’s claim, this unmasking would not violate the First Amendment.

It should be emphasized that prosecution or direct legal action against token owners are not intended to replace the other types of mitigation outlined above, but rather to supplement them.542 Prosecution can punish violators who abuse tokens. Direct legal action can provide remedies such as compensation or disgorgement. Neither paths prevent the entrance of abusive tokens to the blockchain system ex ante or

540. The widely employed standard was established in the case of Dendrite Int’l, Inc. v. Doe, 775 A.2d 756, 760 (N.J. Super. Ct. App. Div. 2001); Koseff, supra note 530, at 119–33 (2022) (addressing the need to set the rules and balancing tests to online anonymity); see also Madeline Lamo & Ryan Calo, Regulating Bot Speech, 66 UCLA L. REV. 988, 1022 (2019) (“First, plaintiffs must notify anonymous speakers in order to provide them with a reasonable opportunity to contest a potential unmasking; second, they must identify precisely which statements are allegedly defamatory; third, they must produce prima facie evidence supporting every element of their claim; and finally, the court must weigh the risk of unmasking the defendant against the harm to the plaintiff on a case-by-case basis.”). On the different standards for unmasking anonymous speakers see Ronen Perry & Tal Zarsky, Liability for Online Anonymous Speech: Comparative and Economic Analyses, 5 J. EUR. TORT L. 205, 215 (2014).
541. Lamo & Calo, supra note 540, at 1022.
542. See supra part IV A-B (preventing the entrance of expressions to the Blockchain by using a pointer to a cloud outside the blockchain, pre-screening abusive expressions ex ante, or obscuring abusive expressions ex post facto).
obscures them ex post. Thus, in these cases, the tokenized expressions would remain on the blockchain. Moreover, as explained, there are difficulties and obstacles to direct prosecution. These obstacles are significant. However, they are not fatal. Plaintiffs who are able to surmount these obstacles would be entitled to damages for harm caused by tokens that include unprotected harmful expressions, or to the remedy of disgorgement, which would deprive the owners of such NFTs of profit.

New technologies emerge, markets change, and the harm that can result from the use of such technologies is greater than ever. We therefore believe that policy makers must rethink the path forward. This Part outlined a comprehensive framework of solutions for abusive tokenized speech. An understanding that the harm of abusive tokens can be perpetual, and can exceed any type of harm known before, should lead to the conclusion that a framework for regulating tokenized speech cannot be left for market self-regulation. This is where the law should step in. We believe that voluntary measures should be taken in the interim until the regulator steps in, and/or as a choice by intermediaries and marketplaces, to self-regulate tokenized speech beyond the threshold of unprotected speech. However, due to the severity of the harm that could potentially be caused by abusive tokens, we believe that legal regulation is preferable to self-regulation in the long run.

We are well aware of possible conflicts that could arise between mandating a framework that includes marketplace liability and Section 230. Yet we are confident that most of these conflicts can be reconciled by adopting a narrow interpretation of Section 230, to adapt it to new emerging technologies and NFT markets. Thus far, courts have used two primary arguments as the basis for denying Section 230 immunity: “(1) the platform at least partly developed or created the content; or (2) the claim did not treat the platform as the publisher or speaker of third-party content.”

544. As explained, prosecution can occur only if the legal authorities invest efforts in prosecuting token owners of expressions that are criminal offences. Civil legal actions can succeed only if publishing the speech on the NFT is considered a wrong, the plaintiff is willing to invest time and money in filing a legal action and can prove every element of the case. Moreover, an additional procedural process of unmasking might be required before the plaintiff could file a direct action, and the plaintiff would have to meet a standard of evidence to support his claim in order to balance free speech and the plaintiff’s constitutional rights such as reputation and dignity.
A possible interpretation that might allow narrowing the immunity granted in Section 230 would be to treat the act of tokenizing as an act of content development. Thus, the marketplace would be seen as a content developer, excluded from Section 230 immunity. Another interpretation that could enable a narrowing of Section 230 immunity is that Section 230 does not apply to platform design. As mentioned above, some courts have already ruled that Section 230 does not shield online intermediaries in all circumstances and have distinguished negligent design claims from claims that focus on content published by other content providers. This narrow interpretation could be adopted with respect to platform design and marketplaces, thereby allowing most lawsuits regarding platform design and marketplaces to advance beyond the preliminary stages.

It might further be argued that, due to the irreparable harm that could be caused by NFTs containing unprotected speech, and the fact that the proposed obligations target the marketplace of commerce and not traditional platforms, there is justification for treating NFT marketplaces as an exception to Section 230. It is important to remember that the same expressions could still be shared on traditional platforms without prescreening obligations.

CONCLUSION

NFTs emerged about three years ago and ascended rapidly in 2021, with a growing sales volume reaching billions of USD. These unique

---

547. 47 U.S.C. § 230(f)(3) “Information content provider” is defined as “any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the internet or any other interactive computer service.” Congress declared that online intermediaries could never be treated as “publishers” of material they did not develop. 47 U.S.C. § 230(c)(1) (emphasis added).
550. See Alexander Tsesis, Marketplace of Ideas, Privacy, and Digital Audiences, 94 NOTRE DAME L. REV. 1585, 1588 (2019) (differentiating between marketplace behavior and freedom of expression). A broad interpretation of marketplace behavior might be a possible path to justify prescreening by marketplace, as the act of tokenizing is marketplace behavior that allows the owner of the token to transact with it.
551. Elizabeth Howcroft, NFT Sales Volume Surges to $2.5 Bln in 2021 First Half, REUTERS (July 6, 2021) https://reut.rs/3jWPxYW [https://perma.cc/3Y7T-TE9S] (“[T]he market for non-fungible tokens (NFTs) surged to new highs in the second quarter, with $2.5 billion in sales so far this year, up from just $13.7 million in the first half of 2020, marketplace data showed.”).
digital assets are a game changer in the digital world, revolutionizing conventions of ownership, artwork, and music, as well as developing new markets around the world. As expressions can be tokenized and stored on the blockchain and these expressions cannot be changed or amended at a later stage, they could become a common means for communicating information and an engine of speech. As NFTs gain popularity, it becomes more likely that the world could finally attain genuine freedom of information and expression, freeing the globe from the shackles of censorship and the control of intermediaries which often remove content erroneously, or disregard context.

This Article has identified the promise of such a world, yet it has also recognized that this utopia could easily deteriorate into a dystopia. NFTs have the potential to tokenize harmful expressions and allow the dissemination of abusive expression, without any ability to alter or remove this abuse from the blockchain. Abuse would thus be perpetuated, and the E.U. Right to be Forgotten could no longer be upheld. Moreover, because of the structure of the blockchain, voluntary practices of content removal are inapplicable in many cases. As a result, NFTs could substantially aggravate existing types of harm caused by dissemination of harmful expressions. Thus, NFTs are not just an engine of speech – they could also become an engine for defamation, shaming, incitement to terrorism and fake news. They have the potential to curtail privacy and intimacy and spread hate and incitement that could even lead to physical violence. Tokenizing speech allows people to publish harmful expression without accountability, leaving victims empty-handed without recourse to any remedy or mitigation by third-parties (such as online platforms) who otherwise remove expressions that do not comply with their community standards.

NFT use is growing exponentially, and NFTs are soon expected to be integrated in social media. Thus, there is real potential that token abuse aimed at perpetuating harmful expressions, and disseminating them to a wider audience, could intensify. Yet, a systematic understanding of the potential of abusive NFTs to cause harm, and the suitable legal mechanisms to mitigate such harm, are currently absent from academic scholarship. This Article endeavors to bridge this gap and offers a comprehensive framework designed to meet the challenges presented by abusive NFTs, allowing different avenues for mitigation and remedy. We also address possible free speech objections to the framework we have proposed and demonstrate that it is in line with the First Amendment.

The solutions we have identified and proposed include an ex ante safety-by-design approach to prevent storage of all information on the
blockchain in the first place. Alternatively, the Article has advocated for the regulation of marketplaces, making them serve as gatekeepers by prescreening abusive expressions so that abusive expressions would not be tokenized to begin with. A second avenue we have identified and proposed in order to mitigate harm is ex post obscuration of abusive information, either by adding more information to the blockchain (a solution that could have an impact especially in permissioned blockchains), or by harnessing online intermediaries to reduce the visibility of such information. Lastly, we have proposed that complementary mitigation of compensation for damage, or disgorgement, should be made possible through legal action against NFT owners.