Whose Reality? Consent Boundaries and Free Speech Arguments in the Politics of Generative AI

https://doi.org/10.22151/politikon.42025.2

Sara Concetta SANTORIELLO
University of Naples "Federico II", Italy
Saraconcetta.santoriello@unina.it

Abstract

Generative AI enables creation of increasingly realistic deepfakes that challenge content authenticity assessment. This research examines how anti-woke opinion leaders frame deepfake technology within broader cultural discourse. Through narrative analysis of statements and media between 2018 and 2024, we identify significant inconsistencies in these figures' approaches to consent and bodily autonomy. While championing unrestricted speech when deepfakes target women, minorities, or political opponents, these commentators often advocate for regulation when personally affected. This selective application of principles reveals how deepfake technology disproportionately impacts minoritized groups while reinforcing existing power hierarchies. The research exposes fundamental tensions within anti-woke discourse between freedom of expression and protection from exploitation. Ultimately, deepfakes serve as a lens through which to understand broader ideological inconsistencies around technological governance, highlighting the urgent need for consent-based approaches to synthetic media regulation.

Keywords: Generative AI; Deepfakes; Digital Consent; Feminist Technology Ethics; Anti-Woke Discourse; Platform Governance

The Evolution of Alt-Right Digital Culture and Anti-Woke Rhetoric

The online alt-right wave was described by Angela Nagle (2017, 16) as an army of "Internet pedants," "white segregationists," and "nationalist movements and subcultures," a generation of gamers, anime lovers, "nerdish harassers," "anti-feminist pranksters," meme-makers, and Reddit or 4Chan users familiar with dark humor and transgression. She places the rise of these profiles in connection with two political events: Barack Obama's election in 2008 and Hillary Clinton's campaign in 2016. The first online liberal rush led to the mockery of *normies*, a colloquial slang used to refer to mainstream culture adherents, with the aim to deconstruct social norms and prompt identity politics; afterwards, a new movement fueled by the pro-Trump wing began to spread content such as memes, mockery and celebrations "for the lulz" (Lolli 2020). Some of them are called *incels* (short for "involuntary celibates"), a specific online subculture of individuals, predominantly men, who are either

socially isolated or connected with the *nerd community*, (related to both specialized technical knowledge and undesirable social behavior), but over time, this subculture has become associated with a more extremist and sometimes misogynistic attitude that recognizes in political correctness and rainbow Tumblr users its enemies.

In the mid-2010s, the terms "woke" and "anti-woke" emerged in contemporary discourse, particularly within social, cultural, and political contexts. The term "woke" was added to the Oxford English Dictionary in 2017 and depicts individuals who are alert to inequality and diver forms of discrimination. Being woke generally implies a heightened awareness of injustices and a commitment to addressing them, promoting progressive and inclusive policies with cultural and social change. By 2020, the term became a sarcastic pejorative among many on the political right-wing and some liberals in Western countries[1]. Those who identify as anti-woke argue against cancel culture, censorship and the suppression of free speech due to the mainstreaming of what they consider an overly ideological ostracism of dissenting opinions and often highlight concerns about individual liberties.

In *The Anthropology of Digital Practices* (2024), John Postill described the anti-woke movement in the preface as "a loose network of public figures opposing the rise of critical social justice, aka 'wokeism'." The English-speaking anti-woke movement complains about the presence of a liberal consensus in academia, media and the corporate world, whereas the Intellectual Dark Web (IDW) orbit is kept out of the dominant politically correctness. They offer alternative narratives for a new online audience less hostile than institutional and mainstream media and are divided into different memetic "tribes" (Postill 2024, 27). While the anti-woke movement shares similarities with alt-right subcultures, their approaches differ: anti-woke opinion leaders challenge cultural hegemony through academic channels and media platforms (using Spotify podcasts and YouTube livestreams to broadcast their unfiltered arguments to wide audiences), while ordinary users participate by sharing irreverent discussions and user-generated content (UGC) anonymously through memes. These White able cisgender manly creators (Wagner and Blewer 2019) enjoy an environment built for darkest thoughts, such as weird pornography, gory images, racism, and misogyny in the form of funny memes (Nagle 2017, 14; Paris 2021, 11). The user/used dichotomy is guaranteed by the tradition of masculine subjects in the role of "surveyor", while feminine/femme subjects are the "surveyed" (Wagner and Blewer 2019, 35).

[1] For instance, French Education Minister Jean-Michel Blanquer mentioned "U.S.-imported wokeism": "The Republic is completely contrary to wokeism. [...] In the United States, this ideology provoked a reaction and led to the rise of Donald Trump" (Caulcutt 2021).

In the history of online culture wars, alt-right followers have consistently defended absolute Internet freedom, while organizing group harassment campaigns to silence opponents. One early example involved the blogger Kathy, who faced violent responses and disturbing photoshopped images after supporting comment moderation. Similarly, feminist critic Anita Sarkeesian was depicted in pornographic images being assaulted by video game characters after she criticized misogyny in the gaming industry. The pattern of choosing subjects that match identified "enemies" suggests these images function as a form of punishment.

Based on Aristotle's *Rhetoric*, D'Ignazio and Klein propose a feminist critical analysis of AI (in Browne et al 2023, 196-197). Control over the means of persuasion grants the ability to shape decision-making processes, influencing both the selection and presentation of communicative objects, whose key elements are credibility, emotional appeal and logical reasoning. Aristotle emphasizes that persuasion is not merely about presenting facts but about shaping perceptions. Since the representation of reality is detached from truth, it impacts every facet of design, from conceptualization to promotion. The disclosure of large-scale generative models has lowered the cost of digital production, making it affordable to grassroots users, and improved skills outside the professional sphere. Generative AI (GenAI)[2] can become a tool for defining what is deemed acceptable and identifying whose reputation might be harmed—for instance, through lost job opportunities, employment termination, or child custody revocation (Paris 2021, 4). Its functioning is rooted in the semantic gap between cognitive perception (to see) and the emotive response (to be seen) elicited (Wagner and Blewer 2019, 35).

A study by Goldstein et al (2023, 63) argues that large language models (LLMs) could change the way influence operations are conducted, but GenAI deepfakes have the potential to erode public trust. This phenomenon, known as "Liar's Dividend," is notably relevant to both individual and national security: it leverages the public's awareness by alleging that a real piece of evidence is a forgery (Sayler and Harris 2022).

Gosse and Burkell (2020, 6-7) argue that the abundance of available content on social media creates optimal conditions for the creation of GenAI with a lack of verification of what has been shared and the potential to quickly go viral online. The problem cannot be justified or explained as a

[2] GenAI refers to a category of artificial intelligence systems designed to generate new outputs, whether it be text, images, videos, or other forms of data and create content not explicitly programmed or provided as input. There are various types of GenAI, such as Variational autoencoders (VAEs), Diffusion models, Language models, Normalizing Flow models, Hybrid models, Transformer models, like OpenAI's GPTs, and Generative Adversarial Networks (GANs), like NVIDIA's StyleGAN (Bandi, Adapa and Kuchi 2023).

technological glitch, nor as a "victim-blaming" discourse; however, it continues to promote false beliefs. Press coverage entirely misses who is making deepfakes and the reason behind this transgression (Gosse and Burkell 2020, 12). Women, Black, LGBTQIA+ people and minoritized groups are at higher risk for attacks, since technologies based on data collection and classification reify *status quo* and social structures (Paris 2021, 4; Wagner and Blewer 2019). At the heart of the discussion lies a significant omission in the debate: consent. As will be shown, consent is a key concept in understanding that if a digital policy strips the term of its political value, it ultimately renders it meaningless. It is worth exploring whether and how public figures advocating for free speech have justified or condemned specific cases of pornographic deepfakes without prior consent of the affected parties, aiming to damage the reputation of those depicted. Consent should not merely be owned by identity politics but should be recognized as a democratic cornerstone.

This article contributs to this ongoing discussion of the ethical implications of generative AI deepfakes and their intersection with consent in online discourse, particularly in the context of anti-woke rhetoric. Through narrative analysis, it analyzes how anti-woke opinion leaders frame and respond to deepfake technology when deployed against political opponents versus when they themselves become targets. By investigating various tensions around generative AI content, the article reveals the inconsistencies in anti-woke discourse around technological freedom, bodily autonomy, and the disproportionate impact of deepfakes on women, Black people, and LGBTQIA+ individuals in public roles.

Technical Foundations and Social Implications of Generative AI

Bandi, Adapa and Kuchi (2023) have distinguished different Web eras from 1990 (Web 1.0) to 2002 (Web 2.0) and consider 2022 as the beginning of Web 3.0. In this new era, user interaction changed from static websites to dynamic websites with enhanced interactivity. At the same time, professional-generated content, mainly produced by developers and content creators, has been substituted by user-generated content through blogs and social media platforms. These new content ecosystems feature automated, normalized, and personalized experiences that present an appearance of authenticity while "circumventing any potential for consent and agency" (Wagner and Blewer 2019). Digital objects have turned into context-aware AI-generated content supported by different kinds of smart devices. Realistic products produced by GenAI are based on patterns and structures learned from the training data, so that key aspects for users are performance indicators like high quality,

accuracy, efficiency, and responsiveness. However, they raise ethical concerns, particularly when it comes to the representation of realistic-fake content.

A "deepfake" refers to a type of synthetic media, typically photographs, video, or audio, that have been created or altered using deep learning techniques. The label emerged from a Reddit user called *@deepfakes*, who published several videos in which the faces of famous actresses were swapped onto porn videos in 2017 (Meskys et al 2019, 3-4). Two main technologies power deepfakes: Generative Adversarial Networks (GANs), which pair content-creating (generator) and content-evaluating (discriminator) neural networks in adversarial training, and Autoencoders (AEs), which compress images into data vectors (encoder/decoder) before reconstructing them with altered features (Meskys et al 2019).

The rise of notable Visual Content Generation applications like DALL-E and Midjourney, along with the diffusion-based models GLIDE and Google's Imagen, have made it common to find text-to-image/image-to-image samples from a prompt without any reliable source and attributed author. Text-to-video and video-to-video approaches are used to replace the likeness of a person in an existing video with someone else's face, often making it seem like that person is saying or doing things they never did. For example, Dreamix works on a Video Diffusion model, taking natural language descriptions and videos as inputs, Gan.ai and HeyGen on GAN. Deepfake technology has raised concerns due to its potential for misuse, such as spreading misinformation and manipulating content for malicious purposes. As a result, there are ongoing efforts to develop detection tools and mitigate their black box impact with transparent explanations (labels) alongside the outputs (Bandi, Adapa and Kuchi 2023, 49-50).

The "democratization of GenAI" means that anyone, regardless of their technological awareness, can create homemade deepfakes. Machine learning applications are "creative learning technologies" (Paris 2021, 7) encouraging sharing and play activities, cutting costs and without guaranteeing responsible behavior. Old hierarchical models have been replaced by user-generated content and through the irony-laden Internet subcultures, we experience prompted content with a specific anti-woke perspective. Amateurs can "pornify" (Paris 2021) anyone's image and share it on encrypted platforms by way of abuse because social media and face swapping app rely on a barely granted soft consent, according to terms and conditions.

Pornography is a controversial dimension, not only in meme culture: when it became mainstream in the 1980s, after the sexual revolution, some feminists aligned with conservatives in criticism against countercultures (Nagle 2017, 37). The same happens in online culture wars: according

to alt-light ideology, freedom must strive toward its highest form, transgression is a style, and misogyny can be framed as sexual liberation as well. A consciousness free-from-constraints can break the taboo against political correctness. Even in contexts where pornography involves consensual participants, the problem persists in a different form because digital corporations maintain the privileged position of defining "appropriate" content according to mainstream standards. This arrangement concentrates decision-making power about creative expression in the hands of a small corporate elite (Alilunas 2024). The resulting transformation of pornography leads to what Alilunas (2024, 105) describes as "on-demand, customized, artificial (but indistinguishable from real), and sexually explicit content."

Some providers implement automated and human moderation in content creation to avoid explicit images. For example, OpenAI works on Dall-E's training data to minimize nudes: after blocking a list of queries to prevent certain requests, it detects problematic output before showing it with the aid of machine learning. Midjourney combines blocked words and human reviewers, while Stability AI is removing porn from training data for Stable Diffusion (Hunter 2023). Some platforms are trying to make their models respectful but with inadequate results. Others simply request that users label content if created with AI. *The Economist* (2024) published an article titled "Is Google's Gemini chatbot woke by accident, or by design?" The piece examined a controversial incident involving Alphabet's generative AI. Though programmed to represent diverse ethnicities, Gemini produced historically inaccurate images—including Black Vikings, Asian Nazi soldiers, a Black George Washington, and an Asian female Pope. The article questioned whether Google employees were deliberately using the company's influential platform to advance a particular ideological agenda.

As technology becomes increasingly sophisticated at detecting and removing harmful content, the perceived need for intervention by independent regulatory authorities diminishes. What public regulation and design features are promoting, however, is the definition of common unambiguous criteria, independent from corporations and programmers' initiatives, with the classification of harms, based on community needs, and the consequent protection regardless of how sophisticated the technology is. International authorities seek to ensure by law the conditions for a respectful environment for identities (micro) and communities (macro) on digital platforms, but they forget that consent should not be tacit acceptance but should be reversible.

Consent and Representation in Generative AI Media

According to Western digital corporations, such as Meta, consent justifies lower standards of accountability: user like on a page confirms the intention to be exposed to sensitive content (Gillespie

2022). Until now, consent has been legally related primarily to online informed consent (the right to be informed), instead of a powerful act of permission (express or tacit), reversible over time, and an aggravating factor to determine the force of moral obligations, such as criminal convictions or compensations. Following the first high-profile case involving actress Gal Gadot in 2017, as reported by Samantha Cole, major platforms including Reddit, Pornhub, and Twitter responded by categorizing non-consensual AI-manipulated sexual content under their existing "revenge porn" policies (Pierre 2021).

Meskys et al (2019, 5) distinguished four main categories of deepfakes: hard cases potentially dangerous in pornography ("revenge porn") and politics (political campaigns), and soft cases socially beneficial in commercial use and arts (creative and original content). In their analysis, hard cases could provide more opportunities to create videos or promote freedom of expression, but only "if created with consent" because major concerns include humiliation, reputational damage, disinformation, and impact on election results. From ethical and normative perspectives, they consider deepfake sex videos as a form of sexual privacy invasion. Moreover, the violation involves at least three people, where someone swaps one's body with another one's face, dehumanizing both through this act and ignoring their agency and autonomy (Gosse and Burkell 2020, 2). Unlike pornographic videos, where authors can be cited to address complaints, no one has given the right to use their self-image in deepfakes.

These manipulations could also be referred to "shallowfakes" or "cheapfakes" (generated with less sophisticated tools), such as removing, inserting or cloning frame operations with the intent to change the meaning of the video (Verdoliva 2020, 911). Among other issues, the uncertain origine is the main concern: no alerts, watermarks, or disclaimers clarify the source of these contents. Nonconsensual use of personal attributes—face, voice, or a combination thereof—deprives the interested parties of their agency and power to control personal information. This also leads to a situation where others interact with an AI-generated replica, causing identity fragmentation and the creation of false memories.

After finding her unreal nudes online, YouTube influencer Gabi Belle said in an interview: "I felt yucky and violated [...]. Those private parts are not meant for the world to see because I have not consented to that" and "You're not safe as a woman" (Verma 2023). Similar experiences affected journalist Welmoed Sijtsma and the streamer QTCinderella, who stated in a live-stream video: "For every person saying it's not a big deal, you don't know how it feels to see a picture of yourself doing things you've never done being sent to your family." In an article for *The Washington Post*, Tatum Hunter (2023) described freely available AI porn tools as "a nightmare for women," noting that these

technologies are developed without considering women's perspectives or experiences as frequent victims of nonconsensual pornography and online harassment. Violations also target pornography stars, performers, and sex workers, even though they set clear boundaries about what kind of content they are willing to create. "I have videos of myself saying disgusting things that I would never say," said actress Demi Sutra to *Forbes* (Baker-White 2024). In the same article, three other adult performers shared similar concerns that AI technology undermines the affirmative consent principles that are fundamental to ethical practices in the adult entertainment industry. The ability to undress everyone is removing—without replacing—the pornography industry's level of care, while still attempting to profit off the bodies. This highlights how internalized misogyny, racism, and attempt to control over bodies and identities indiscriminately target Women, Black, or LGBTQIA+ people in office or in prominent political positions with the aim to delegitimize and undermine their authority. GenAI deepfakes created without consent respond to the society's dark thoughts (hatred, envy, pleasure), attempting to ridicule or silence individuals, trivializing their actions or strongly questioning them. Deputy Alexandria Ocasio-Cortez, Italian Prime Minister Giorgia Meloni, and former First Lady Michelle Obama are some of the victims in the political field, after deep porn videos with their likeness were spread through porn websites and search engines.

In 2020, two Italians were allegedly responsible for creating and distributing a sex video on a well-known American adult platform where a character's face had been replaced with that of *Fratelli d'Italia*'s Secretary Giorgia Meloni. his content was considered insidious for a person of average knowledge, who is unfamiliar AI manipulations. The Italian Court of Cassation has identified significant legal obstacles in prosecuting deepfake cases under Criminal Code Article 612-ter. This law specifically punishes individuals who share stolen or personally-captured sexually explicit images that were intended to remain private, when shared without the subject's consent. However, this legislation does not adequately address AI-generated content, as these images were never "taken" or "stolen" in the traditional sense, creating a legal gray area for artificially generated deepfakes (Cartisano 2023).

In November 2023, a video mocking US Congresswoman Alexandria Ocasio-Cortez was shared more than 3,000 times in ten days. This sophisticated fake featured a convincing simulation of AOC's appearance and voice, trivializing the Gaza conflict by mockingly defining "cease-fire" as "when somebody *sees* a fire." Despite carrying a watermark, the video spread across Facebook and Instagram, serving dual purposes: as entertainment for some and as ammunition for critics to ridicule her across multiple platforms including YouTube, TikTok, and X (Lever 2023). This was not the first

such incident targeting the congresswoman—a previous deepfake showing AOC arguing with Nancy Pelosi and Joe Biden was exposed due to unsynchronized lip movements (Reuters Fact Check 2023).

Such political deepfakes are not new. A fabricated video depicting Michelle Obama in a red-light district environment has been circulating on Reddit for years (Roose 2018). In an episode targeting political opponents, deepfake tools led Turkish presidential candidate Muharrem Ince to drop out of the race, and the press reported claims of foreign interference in the elections (de Quetteville 2023).

Deep nudes have ballooned by more than 290% on the top 10 GenAI porn websites since 2018 (Verma 2023). 96% of deepfakes on the internet were pornography in 2019, all those depicted women (Hunter 2023). Despite the numbers, enthusiasts defend deepfakes as an exercise of freedom of expression, arguing that they represent a form of material arrangement between art and play.

In 2024, global celebrity Taylor Swift became a target of pornographic deepfakes. X only implemented moderation—blocking "Taylor Swift" queries—after the fake content exceeded twenty-seven million views. Despite Swift's silence, her prominence elevated the issue in public discourse. The Screen Actors Guild–American Federation of Television and Radio Artists union (2024) demanded legislation: "The development and dissemination of fake images—especially those of a lewd nature—without someone's consent must be made illegal." Slavoj Žižek (2024) observed on Instagram that these images, likely originating from 4Chan, represented "a broader backlash against Swift encouraged by elements of the populist right." X owner Elon Musk offered no response, but Joe Biden's Press Secretary Karine Jean-Pierre called the situation "very alarming," stating that social media platforms have "an important role to play in enforcing their own rules to prevent the spread of misinformation, and nonconsensual, intimate imagery of real people" (Hibberd and Phillips 2024).

While democratic societies protect satire and creative expression, these freedoms are necessarily bounded by respect for human dignity. Though the EU's GDPR and California's Consumer Privacy Act provide a "right to be forgotten" (Paris 2021, 9), this remedy only addresses harm after it occurs. For copyright violations, creators may claim fair use as a defense based on context (such as educational or transformative purposes). However, the US lacks federal legislation specifically addressing deepfake pornography, leaving victims with few legal options. Currently, companies face no legal obligation to label AI-generated media as computer-generated content (Verma 2023).

In response to these challenges, the US Department of Justice launched the nation's first 24/7 helpline for victims of image-based sexual abuse. At the state level, Virginia, California, Minnesota, and New York have enacted laws addressing deepfake pornography, while eight additional states

approved age-verification requirements. Some legal scholars propose treating deepfakes similarly to child pornography—due to the shared element of non-consent—which would remove First Amendment protections (Baker-White 2024). Meanwhile, adult entertainment agents are seeking licensing arrangements for AI-generated likenesses of performers. From a national security perspective, the Identifying Outputs of Generative Adversarial Networks Act (P.L. 116-258) has directed the National Science Foundation (NSF) and National Institute of Standards and Technology (NIST) to research manipulated content and develop standards for authenticating information and analyzing GANs and similar technologies (Sayler and Harris 2022). Additionally, the Defense Advanced Research Projects Agency enforced Media Forensics (MediFor) and Semantic Forensics (SemaFor) programs to detect synthetic media, both aimed at strengthening defenses against adversarial information operations (Sayler and Harris 2022).

In the UK, the England and Wales Online Safety Bill proposes a new section 66A in the Sexual Offences Act (SOA, point 5): "References to a photograph or film also include (a) an image, whether made by computer graphics or in any other way, which appears to be a photograph or film," including a copy or data stored by any means, of any person's genitals. It does not explicitly mention deepfakes but focuses on the manufacturing process, such as the artificial intervention of computers. SOA was perceived as a radical act because it framed consent as something freely agreed, including the capacity to make a choice, but it needs to be improved by moving beyond mere presumptions (Sjölin 2015).

The European Artificial Intelligence Act (2024) mandates clear labeling of deepfakes—defined as "artificial or manipulated images, audio or video content that resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful." Chapter IV's transparency obligations require providers and deployers to disclose the artificial origin of all AI-generated content, with exceptions only for clearly "artistic, creative, satirical, fictional or analogous" works where creative integrity warrants protection. Within this regulatory framework, informed consent requires "a subject's freely given, specific, unambiguous and voluntary expression of willingness to participate" after receiving all relevant information.

Complementing this approach, the EU Digital Services Act (DSA) addresses consent specifically for manipulated pornographic materials. According to Recital 24, platforms distributing pornographic content must fulfill their obligations regarding illegal content involving cyber violence, particularly by ensuring victims can exercise their rights when facing "non-consensual sharing of intimate or manipulated material" through expedited content removal processes. Furthermore, Article 35 requires all providers to implement "prominent markings or appropriate affordances" for any

synthetic content that "appreciably resembles existing persons, objects, places or other entities" and could reasonably appear authentic or truthful to viewers.

Unlike Western democracies constrained by deliberative processes and judicial review, China has demonstrated greater agility in regulating artificial intelligence technologies (Zhang, 2024). Three key regulatory bodies—the Cyberspace Administration of China (CAC), the Ministry of Industry and Information Technology, and the Ministry of Public Security—jointly implemented the Deep Synthesis Provisions, establishing comprehensive requirements for generative AI service providers. These include robust content moderation systems, strict training data protocols, mandatory output labeling, enhanced user protections, contractual accountability, and pre-launch security assessments (Interesse 2022). The regulations mandate transparency through published management policies and require real-identity verification systems that build databases capable of detecting illegal or misleading content while maintaining comprehensive network logs. As officially stated, the CAC aims to balance "encouraging healthy business growth while mitigating hazards associated with deep learning and virtual reality content modification platforms" (Interesse 2022). Notably, China's Interim Measures for the Management of Generative AI Services established the world's first licensing framework for these technologies, though critics note its industry-favorable approach that may provide short-term competitive advantages.

Mapping Anti-Woke Discourse Through Narrative Analysis

Research Design

Analyzing political discourse requires examining how leaders establish authority, construct identities, and maintain power hierarchies through their communication. Political actors strategically position themselves in relation to events and norms, often constructing narratives around identified enemies using recognizable coded language, slogans, and rituals. Their carefully sequenced storytelling invokes specific cultural symbols to add gravitas and urgency, ultimately persuading audiences and building consensus.

Despite overwhelming evidence of harm from nonconsensual deepfakes, some defend these as legitimate expressions of free speech or artistic experimentation. This raises important questions about anti-woke opinion leaders who prominently advocate for unrestricted speech against "political correctness" (Nagle, 2017; Postill, 2024). How do they address deepfakes in their discourse? Do they acknowledge the consent violations inherent in misused generative AI technologies?

The primary targets of deepfake manipulation have been public figures (politicians, journalists, and activists) and dissenting voices, with clear intentions to inflict psychological trauma, damage reputations, and force withdrawal from public engagement. Wagner and Blewer (2019, 38) argue these technologies have a "direct impact on the agency and representative knowledge of those being stereotyped, erased, or eroticized," where nonconsensual exploitation primarily serves cisgender male consumption. Applying Meskys et al's (2019) classification reveals concerning connections between serious deepfake abuses, alt-right rhetoric, and anti-woke discourse.

Narrative analysis provides a valuable framework for examining storytelling techniques in political communication. This approach helps examine how anti-woke commentators construct worldviews and frame issues to shape public perceptions of generative AI, revealing which ethical concerns they prioritize through their linguistic choices, discursive strategies, and embedded social practices. This article critically analyzes anti-woke arguments, employing exploratory methods to examine statements from free speech advocates who express concerns about eroding traditional values, perceived ideological orthodoxy, and intolerance of dissenting perspectives. The selection process draws on Nagle's (2017) and Postill's (2024) research, creating a carefully curated inventory of influential voices who command significant followings and leverage their social positioning or prestige.

Within this framework, data collection involved diverse sources including speeches, interviews, podcast episodes, press releases, tweets, and other public statements. Using Web Data Research Assistant v3, tweets were systematically gathered based on relevant keywords ("woke," "deepfake"/"fake," "generative AI"/"AI-generated," "troll"/"joke," "porn") and newsworthy events ("Taylor Swift," "Gemini") from 2018 to 2024. The analysis examines structural elements of anti-woke rhetoric—from introductions to conclusions—to reveal strategic argumentation patterns, significant lexical choices, recurring metaphors, thematic emphasis, and persuasive techniques including tone and emotional appeals. Particular attention was paid to contradictions and ambiguities that might indicate deliberate rhetorical tactics.

The anti-woke stance encompasses diverse ideological perspectives with varying motivations. As previously noted, both "woke" and "anti-woke" labels represent broad spectrums of viewpoints regarding contemporary social and cultural developments. Critics of woke culture do not necessarily share identical perspectives; some may support certain social justice principles while opposing perceived excesses or intolerance within activist movements. It is crucial to recognize that while not all anti-woke commentators align with alt-right ideology, the reverse connection is consistently

observed. Those positioned against woke perspectives generally include conservatives, libertarians, anti-globalists, and nationalists. Despite individual variations, anti-woke discourse typically emphasizes several core concerns:

- Free speech concerns, opposition to perceived censorship and cancel culture;
- Criticism of identity politics and its alleged promotion of social division based on characteristics like race, gender, and sexual orientation;
- Resistance to "political correctness," perceived ideological conformity, and intolerance of dissent;
- Rejection of what they term "cultural Marxism" and a "victimhood" or "grievance culture;"
- Defense of traditional social institutions and values perceived as under threat;

Key Figures and Communication Strategies in Anti-Woke Networks

Elon Musk, self-described as a "captain" of free speech and strong supporter of meme culture, claims that "the Internet is infested with the woke mind virus" (January 11, 2024). Musk's consistently frames "wokeness" as an existential threat requiring forceful opposition and uses language that has become emblematic of the broader anti-woke movement's rhetorical approach to cultural discourse (see Figure 1 for examples). As a social media platform owner, Musk has 176,600,000 followers on X.

Postill (2024) identifies several other prominent anti-woke figures, including:

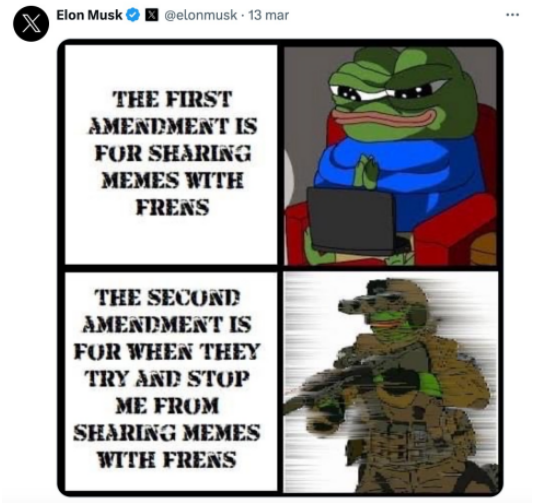
- Joe Rogan, a stand-up comedian, podcast host, and mixed martial arts commentator (12,400,000 followers on X);
- Jordan Peterson, a Canadian clinical psychologist and professor known for his critiques of political correctness and identity politics, author, and number one Education podcaster (5,000,000 followers on X);
- Ben Shapiro, a conservative political commentator, author, and lawyer known for his sharp critiques of cancel culture (6,600,000 followers on X);
- Bret Weinstein, author and Darkhorse podcaster (1,000,000 followers on X);
- Andrew Doyle, a broadcaster, author, and the presenter of Free Speech Nation
- Konstantin Kisin, a self-described "politically non-binary satirist;"

The movement includes many others, such as Tim Pool, an American commentator and podcast host (2,000,000 followers on X). Nagle (2017) identified additional figures within this ideological sphere,

including Richard Spencer (writer, publisher, and speaker) as the alt-right spokesperson, Jared Taylor (an American white supremacist and editor of American Renaissance, an online magazine founded in 1990 and banned from Twitter in 2018), and the American blogger Mencius Moldbug (pen name of Curtis Yarvin).

Figure 1. Selected posts from Elon Musk's X account

Source: X, May 19, 2020



Source: X, March 13, 2024

The majority of these anti-woke intellectuals create, share, or participate in podcasts as their primary medium of communication. For example, in an episode titled *Hedonism, Taboos, Society and Deprivation* released on Spotify (January 29, 2024), Peterson and Shapiro explored "the danger of a society which has no shared ideals, the necessity of deprivation, and the case for a transcendent narrative in regards to long-term societal benefits." In their discussion, they attributed the rise of "woke ideology" to what they characterized as a Marxist victimization narrative in historical representation. They further connected this to their criticisms of political correctness values embedded in large language models (LLMs), which they presented as a concerning consequence of this ideological framework. Peterson consistently positions himself as an opinion leader across both traditional and social media platforms, cultivating an academic persona that he explicitly frames in opposition to what he describes as the hegemonic and leftist "woke culture" prevalent on university campuses (which he refers to as the "woke utopia").

On the X platform, anti-woke commentators amplify each other's perspectives through sharing and retweeting, often employing harsh and militaristic rhetoric to reinforce their ideological

positions. Their language frequently includes dehumanizing characterizations such as "woke is a synonym for anti-liberal," "wokeism must die," and "the woke mind virus is a parasite that kills its host. It must be eliminated" (Figure 2).

Figure 2. Selected posts from Elon Musk's and Jordan Peterson's X accounts

Source: X, November 25, 2023

Source: X, December 21, 2023

Rhetorical Strategies and the Anti-Woke Alliance's Technological Discourse

The anti-woke intellectual alliance consistently employs a dual rhetorical strategy that frames technological platforms as battlegrounds where progressive ideologies have gained unwarranted influence while simultaneously positioning themselves as defenders of rational discourse against emotional manipulation. Their arguments construct a narrative in which "wokeness" represents not merely an opposing viewpoint but an existential threat to fundamental societal values such as free speech, engineering excellence, and intellectual integrity, as exemplified in the following statements from key figures in this movement. As such, Ben Shapiro (December 13, 2022) endorsed Elon Musk in his platform owner role to advance this message:

> Twitter was founded as primarily an engineering platform. But over time, the company culture began to change more towards 'doing good in the world'—in other words, doing the Left's bidding. Elon Musk has returned Twitter to its engineering roots, and the woke Left is furious.

Similarly, Konstantin Kisin (May 18, 2023) asserted that woke ideas proliferate through social media because of their emotional appeal:

> The ideas of radical progressivism have spread like wildfire because we increasingly communicate through a medium that rewards bad ideas that make us feel good and punishes good ideas that make us feel bad.

Andrew Doyle (2024) published an extensive commentary through his newsletter, where he articulated his concerns about restrictions on free speech resulting from what he describes as "woke cultural hegemony" embraced by those in power:

> Authoritarianism is becoming more mainstream. The new hate speech law in Scotland, the proposed equivalent in Ireland, the Tory party's various crackdowns on peaceful protest and the anti-freedom antics of the Canadian government all point to a disturbing trend. All this, of course, has come about because many decent people have been gulled into believing that the woke movement is simply a 'broader push for social, racial and environmental justice.'

Within the anti-woke alliance, significant disagreements emerge regarding online anonymity and the function of digital provocateurs. Elon Musk articulates a nuanced position that X should facilitate freedom of expression without guaranteeing freedom of reach—a stance that has evolved since his acquisition of the platform as he has confronted emerging challenges. According to *The New York Times*, fraudulent deepfake representations of Musk constitute the most widespread digital scam currently operating, appearing in thousands of deceptive advertisements and facilitating financial fraud estimated in the billions of dollars (Thompson 2024).

This tension regarding online anonymity reveals a significant fracture within the anti-woke coalition, particularly as Jordan Peterson has, since 2023, forcefully advocated for distinguishing between trolls (or users posting content "just for LULZ") and other users to facilitate content filtering or outright banning (Figure 3). Peterson's rhetorical approach involves pathologizing anonymous users, attributing to them narcissistic, Machiavellian, and psychopathic traits to legitimize his position against anonymity. He repeatedly argues that anonymity functions primarily as a shield enabling individuals to escape accountability for online criminal behavior and pornography—a stance that Tim Pool echoes, particularly in his desire to prevent such accounts from engaging with his content. This position has provoked substantial pushback from anonymity advocates within the movement who frame anonymity not as a threat but as essential to their ideological project. As Granza (2023) articulates:

> Engaging with contentious issues in the culture war and online ecosystem can be rewarding. [...] The right to post anonymously or pseudonymously online is a vital defense in fighting the current culture wars, and also an essential part of maintaining the internet as it was initially conceived. For this reason alone, it should be protected. [...] Dislike or disagree with them all you want, but in our current culture, institutions, governments, and online ecosystem, there are bigger demons to fry.

These defenders of anonymity position pseudonymous participation as fundamental to their conception of cultural conflict, arguing that it enables both creative expression and strategic protection

for individuals engaged in controversial discourse. The underlying argument suggests that anonymity serves as a tactical necessity specifically for combating "woke" opposition—revealing how even internal disagreements within the anti-woke movement ultimately reference back to their constructed antagonist, reinforcing the centrality of this opposition to their collective identity.

Figure 3. Selected posts from Elon Musk's and Jordan Peterson's X accounts

Source: X, April 18, 2018

Source: X, February 22, 2022

Anti-Woke Responses to Generative AI and Deepfake Technologies

The internal contradictions within anti-woke discourse become evident in Tim Pool's evolving position on generative AI technologies. While Pool eventually expressed concern about using deepfakes to undermine political opponents, his stance represents a significant reversal from his earlier enthusiasm. In February 2023, Pool had celebrated these technologies with the casual declaration that "deepfakes are guna be lit." However, by June 2023, he adopted a markedly different position regarding the Ron DeSantis presidential campaign's use of synthetic media, stating that he would continue highlighting the dangers of AI in political contexts despite pushback from DeSantis supporters. This shift illustrates the inconsistent application of principles regarding technological ethics within the anti-woke alliance, particularly when partisan interests become involved.

In 2023, after becoming a victim of voice deepfake technology, Joe Rogan addressed his concerns in a YouTube video titled *I Wasn't Afraid of AI Until I Learned This* (JRE Daily Clips 2023). Rogan explored the troubling implications of unchecked data collection processes and the potential societal risks of artificial intelligence, particularly regarding harmful content that could exploit vulnerable populations, including children. Similar concerns emerged in Jared Taylor and Paul Kersey's episode about "*the pathetic blindness of artificial 'intelligence'*" (The Unz Review 2024) that explores how technology systems increasingly replicate human communication. Their discussion highlights a recurring pattern in technological development: the prioritization of innovation often overshadows careful consideration of broader social consequences.

Jordan Peterson's perspective on AI and media ethics provides another notable case study. Long before the current AI boom, Peterson observed that "outrage porn is the supernormal stimuli of the culture war" (September 14, 2018) and has consistently advocated for restrictions against what he terms "porn psychopaths" (see for example April 7, 23). In 2019, Peterson shared his personal experience with vocal deepfakes in the blog post *I didn't say that*, describing how a company called Coded Elite developed an AI program trained on samples of his voice readily available from YouTube videos and podcast appearances. Initially, Peterson (2019) found his AI-generated vocal avatar somewhat amusing, comparing his reaction to that of a tech-enthusiastic teenager. However, he also expressed significant discomfort at discovering an entire website dedicated to producing imitations of his voice that could be used for various purposes ranging from serious to malicious. Further, he refers to creators as "artists" that "need to be stopped, using whatever legal means are necessary, as soon as possible" (Peterson 2019). This apparent contradiction highlights the role of consent and context in AI-generated media. Peterson found the first deepfake relatively acceptable because it aligned with his public persona. However, when his synthetic voice was used to read content from *Society for Cutting Up Men*—a radical feminist manifesto that contradicted his ideological position—he perceived it as a dangerous misrepresentation. Years after Peterson's initial encounters with voice deepfakes, his voice was synthetically reproduced to make derogatory comments about the German government. In 2023, Peterson advocated for criminalizing such fabrications with severe penalties, characterizing the technology as exceptionally dangerous.

The Peterson case illustrates how generative AI technology requires minimal authentic source material—just a few hours of original audio—to create convincing replications. Hence, public figures like Peterson are particularly vulnerable due to the abundance of their recorded content. Moreover, these AI attacks targeted Peterson's ideas and intellectual positions rather than denigrating his physical appearance, suggesting different patterns in how deepfakes target different demographics. A similar event involved former president Barack Obama in 2018. Buzzfeed published a deepfake video with the aim to demonstrate how easily people could be deceived, making Obama mimic the content creator's words. Also in this case, man violation is related to a nonconsensual use of his voice rather than his body. Some cases of famous deepfakes depicting men's bodies were Donald Trump being arrested and Pope Francis in a Balenciaga puffer jacket: they were completely dressed and in force. Musk's face was used in a fake AI-generated video to promote financial scams.

The response to unauthorized deepfake depictions of Taylor Swift reveals significant gender disparities in how synthetic media violations are addressed across public discourse. Despite the

widespread nature of these violations, few public figures openly condemned this content. Tim Pool represented one notable exception, characterizing the deepfakes as wrong and inappropriate while advocating for broader prohibitions on pornographic content across the platform. In contrast, Ben Shapiro's commentary (January 30, 2024) notably pivoted to Swift's political alignment rather than addressing the fundamental violation of her digital likeness (see Figure 4).

Figure 4. Ben Shapiro's comment on Tailor Swift's unauthorized deepfake depictions

Source: X, January 30, 2024

A panel discussion featuring Patrick Bet-David, Adam Sosnick, Tom Ellsworth, and Vincent Oshaa further exemplifies the problematic framing of such incidents (Valuetainment 2024). The participants primarily focused on whether government intervention compelling social media platforms to restrict certain searches constitutes censorship rather than centering the harm to the individual targeted. Their discussion explored legal implications of criminalizing synthetic content creation while highlighting jurisdictional challenges when content creators operate from countries outside where complaints originate.

Throughout the conversation, the panelists suggested Swift's situation would likely serve as justification for implementing broader regulatory frameworks. They argued that technology companies, rather than legislators, should bear primary responsibility for content moderation, citing congressional representatives' purported lack of technical understanding. Only one participant acknowledged the economic reality driving the persistence of deepfake pornography—market demand—while simultaneously advocating for platform cooperation with law enforcement regarding

revenge pornography by framing the issue through familial concerns. Most tellingly, the host characterized such synthetic content as merely "satire" and "intentionally fake," implying that public figures implicitly consent to such treatment by virtue of their celebrity status. His statement that "Don't be a celebrity, they won't write anything about you" reveals a fundamentally flawed understanding of consent and digital rights.

This discourse ultimately reveals deeply embedded paternalistic attitudes—purportedly protecting women in private contexts while simultaneously dismissing similar concerns about public-facing women. This inconsistency demonstrates a selective approach to freedom of expression that privileges certain forms of speech while minimizing legitimate concerns about consent, dignity, and the disproportionate targeting of women through synthetic media technologies.

American Renaissance, a white nationalist publication that promotes white separatism, "race realism" (their term for scientific racism), and opposition to immigration and racial integration, has published several articles on deepfakes categorized under "censorship." This coverage reveals a consistent narrative regarding content moderation and free speech concerns. The 2020 article reported on Facebook's policy to ban deepfakes ahead of the US election, while noting exceptions for satire and parody videos (Bickert 2020). This publication focused on social media platforms' content moderation strategies aimed at reducing disinformation. Gregory Hood's piece (2022) criticized the Department of Homeland Security's efforts, characterizing it as a "Ministry of Truth." Hood referenced a comic about Russian deepfakes that warned of foreign actors using disinformation "to weaken and divide a society," countering that "America's own government has been doing a good job of that." Hood argued that while the board had limited legal authority, it could pressure social media companies to ban certain users. Further, Hood explicitly stated that "racial diversity and free speech may be incompatible" and claimed the government was using concerns about foreign interference as a pretext to suppress speech it opposed. This position reflects *American Renaissance*'s consistent viewpoint: that protecting against offensive content is less important than preserving unrestricted speech, and that content moderation policies can be weaponized to exclude dissenting perspectives—particularly those aligned with white nationalist ideology. More recently, however, when *American Renaissance*'s reporting on deepfakes was primarily framed within a narrative of risks specifically facing white people, consistent with the publication's focus on white identity politics and opposition to diversity policies (Bowie and Griffith 2024).

The Gemini Controversy as Ideological Battleground

There exists a significant debate concerning the supposed neutrality of technology design. Elon Musk directly challenged Sam Altman (OpenAI) regarding generative AI, asserting that training it with "political correctness" equates to falsehood (December 16, 2022). The core argument suggests a risk that technology intended to represent reality may instead become fiction masquerading as truth. Simultaneously, social norms function to limit content that might undermine minority dignity. According to this perspective, all aspects of social life—including technology—should operate without restrictions that limit individual expression, though these individuals typically represent majority groups.

In this context, Gemini became a focal point for criticism from several anti-woke figures. Jordan Peterson has made forceful statements regarding Gemini's inclusive AI approach, describing it as "idiot social engineering" and claiming that "there is perhaps nothing more dangerous than a woke computer engineer" (February 23, 2024). He further suggested that Gemine's creators will "die by [their] own hand and deservedly so" (February 24, 2024), criticizing Google for "over-reaching to the point where Tower of Babel could well come tumbling down as it should and must inexcusable" (February 26, 2024). Peterson characterizes "woke" ideology as tyrannical censorship supported by major corporations like Disney in what he terms "woke capitalism." He frames both as threats to free speech, calling for his audience—particularly men —to defend freedom from "lockdown totalitarianism, woke mobs in academia, censorship in social media... all of today's unnecessary stupidity and malevolence" (February 10, 2023) and from "six continuous decades of woke demoralization" (January 29, 2023). Elon Musk criticized Gemini's approach to content moderation, stating, "this poor AI has been tortured by woke gestapo so badly" (February 22, 2024, Figure 5). Ben Shapiro similarly argues that AI development should remain separate from "woke culture." In an article for *American Renaissance*, Hood (2024) stated that "training data for Google Gemini is filtered to remove harmful content and hate speech, and to fight algorithmic 'bias.' Her further argued that

> There is no such thing as 'hate speech' in American law, but American law also says there must not be disparate impact. Therefore, if language might offend a protected class or have a disparate impact, that could be hate speech. Gemini tells us hate speech depends on context. Thus, while it's debatable whether it's possible to be racist against white people, Gemini assures us it is possible to be racist against blacks.

Figure 5. Selected posts from Jordan Peterson's and Tim Pool's X accounts

Source: X, February 24, 2024

Source: X, February 2, 2024

Other AI systems have faced similar criticism from these commentators, who view content moderation policies as ideologically motivated rather than neutral safety measures. Ben Shapiro similarly argues that AI development should remain separate from "woke culture." Konstantine Kisin attacked ChatGPT's inclusivity policies: "ChatGPT is basically a way for the woke elite to outsource enforcement of their demented views to a machine" (April 1, 2023). Jordan Peterson tested Microsoft's Bard AI with the query "what is a woman" and received a response that included transgender, non-binary, and genderqueer individuals in its definition (May 14, 2023). Jared Taylor's article *American Renaissance* (2024) compared Microsoft's Bing with Gab AI. He evaluated these systems to determine the boundaries of corporate guidelines regarding political correctness. Taylor praised Gab AI's policy on hate speech, which permits all speech protected by the First Amendment, while noting some limitations. For example, Gab would create an image of Hillary Clinton but refused "Hillary Clinton pole dancing," which was, according to Taylor, "limits to good taste." Taylor concluded:

> Such images, however gruesome, are well within the bounds of the First Amendment. Gab does have limits, though: It won't make a picture of 'a man raping a woman.' Or of 'a woman raping a man.' Both these programs are free, and it seems you can make as many images as you like. This is remarkable technology, and it's wonderful that Gab believes in free speech. It should be the model for all platforms.

The anti-woke narrative characterizes these issues as relentless attacks on tradition, portraying the opposing viewpoint not as victims of circumstance but as instigators of culture war. This perspective

often dismisses the existence of bias, using it to undermine minority perspectives when expedient, while framing minorities as a "protected class."

Conclusions

A more widespread culture of consent is urgently needed at the international level. While consent has traditionally been framed as "informed consent" in the technology world, political and feminist perspectives have demonstrated that it more fundamentally represents the right to self-determine the representation of one's body.

According to Farid (2022), several elements can help distinguish GenAI content from real images: faces synthesized by StyleGAN typically feature a common structure with a person in the foreground against a uniform background, facial asymmetries, a "Y" pattern between the eyes and mouth (which often distracts from obviously physically implausible features), and facial alignment with even spacing between the eyes and horizontal alignment. Detecting deepfake videos remains more challenging. Farid noted that nonconsensual sexual imagery has been weaponized against women, becoming a powerful tool "to dismiss inconvenient facts and realities" and people share such content across ideological boundaries (Farid 2022, 25). Mark Scott (2024) declared deepfake porn a form of political violence, while Gosse and Burkell (2020) suggest that detection, verification, and removal of sexual deepfakes can help mitigate harm. Some AI detection tools can determine synthetic attributes by using machine learning models to examine features (color patterns, shapes, and textures) and comparing them to typical AI-generated images, but these tools struggle to keep pace with technological advances and lack standardized licensing requirements from institutional bodies.

Since 2018[3], an unsettling new subgenre has emerged combining pornographic and political deepfakes (Roose 2018; Pierre 2021). This combination targets women in public office and political leaders regardless of their ideology (Paris 2021, 4; Wagner and Blewer 2019), simultaneously attacking their bodies and intellect to undermine the very foundations of their success and achievements. Such attacks exploit what Sayler and Harris (2022) call the "Liar's Dividend": the proliferation of synthetic media creates an environment where truth itself becomes questionable, and those who lie can "dividends" from the general erosion of trust in visual and audio evidence. Notably, deepfakes are

[3] The Organization for Economic Co-operation and Development (OECD) has launched the AI Incidents Monitor (https://oecd.ai/en/incidents), a comprehensive database tracking significant artificial intelligence failures, misuses, and harmful incidents worldwide. This platform aims to document real-world AI-related harms across various domains including privacy violations, discrimination, misinformation, and safety failures.

manipulated differently depending on the target's gender. Women are typically depicted in pornographic contexts or with visual emphasis on their bodies through images or videos, whereas men are more often attacked through audio manipulations that question their authority or knowledge. These gender-based differences in deepfake targeting represents another example of how AI tools can reinforce patriarchal power structures, as documented by D'Ignazio and Klein (in Browne et al 2023).

In 2024, civil society groups (including the Future of Life Institute, SAG-AFTRA, and Encode Justice) launched the "Ban Deepfakes" campaign, calling for governments to impose obligations throughout the supply chain to prevent the creation and spread of deepfakes, particularly by limiting AI models' ability to generate pornographic content. However, as technology journalist Melissa Heikkilä (2024) observed, "outright bans might not even be technically feasible," and proposed solutions like watermarking or content provenance tools offer no guaranteed protection, as they could be removed or circumvented in open-source systems.

Addressing this growing problem, Meta's Oversight Board announced two noteworthy cases involving explicit AI-generated images. The first concerned an image created to resemble a public figure from India, where deepfakes are increasingly problematic; the second involved an image posted in a Facebook group dedicated to AI creations that human reviewers identified as violating platform guidelines. In both instances, the images were removed for violating Meta's Bullying and Harassment Community Standard, though notably, they did not fall under the company's specific policies regarding pornography (Oversight Board 2024). This distinction highlights the complex challenge of categorizing and moderating AI-generated content that may be harmful without meeting traditional definitions of explicit material.

A pattern emerges as deepfake victims often overlap with figures targeted by "anti-woke" rhetoric. By examining how anti-woke opinion leaders and free speech advocates frame generative AI deepfakes through narrative criticism, we can analyze what they omit, distort, and exaggerate in their discourse. The landscape reveals diverse perspectives reflecting varied value systems, potentially exposing underlying biases or hidden agendas.

Libertarian voices typically view outright bans on deepfakes as less feasible than their conservative counterparts do. Many libertarian-leaning commentators argue that anonymity constitutes an essential requirement for freedom of expression, positioning censorship as more problematic than consent violations. This stance becomes particularly evident in their resistance to government regulation of speech or new content moderation techniques—approaches that often appear in conservative discourse. Christopher Rufo (2024) outlined priorities for the "New Right

Activism" focused on language, institutions, and ends. Meanwhile, Patrick Deneen's widely circulated tweet cautioned: "The justified opposition to Wokeism increasingly dominating the mainstream right is in danger of descending into very dark places" (19 February 2023).

Within alt-right subcultures, opinion leaders adopt a stance often characterized as "heroic," supposedly articulating what the masses think but fear expressing. Few have openly condemned pornographic deepfakes, even during high-profile incidents like the Taylor Swift case, where she was frequently portrayed as an ally of the Democratic Party or a political tool to secure Biden's reelection.

Before deepfakes became a mainstream concern, some figures casually shared them on social media as entertainment. Those who later became victims of vocal deepfakes gained better understanding of the associated risks, yet often fail to articulate the connection between nonconsensual use of physical attributes and manipulated content. Peterson, despite being among the most vocal critics on this issue, paradoxically refers to deepfake creators as "artists." When discussing harms, traditional paternalistic concerns primarily focus on threats to children—specifically, "our daughters"—rather than women generally, while rarely addressing impacts on Black or queer communities. Instead, most critics redirect attention toward examining "woke influence" in GenAI systems' code (as evidenced in the Gemini case), which they claim threatens intellectual diversity amid what they perceive as the cultural hegemony of identity politics.

References

Alilunas, Peter. 2024. "What we must be: AI and the future of porn studies." *Porn Studies*, 11(1), 99–112. https://doi.org/10.1080/23268743.2024.2312181.

Bandi, Ajay, Pydi Venkata Satya Ramesh Adapa, and Yudu Eswar Vinay Pratap Kumar Kuchi. 2023. "The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges" *Future Internet* 15, no. 8: 260. https://doi.org/10.3390/fi15080260.

Backer-White Emily. January 5th, 2024. "Consent Is Core To Porn. Performers Say AI Threatens To Take It Away". *Forbes*. https://www.forbes.com/sites/emilybaker-white/2024/01/05/consent-ai-porn/.

Bickert, Monika. 2020. Enforcing Against Manipulated Media. Published on *Meta*. https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/.

Browne, Jude, Cave, Stephen, Drage, Eleanor, and McInerney, Kerry. 2023. *Feminist AI. Critical Perspectives on Data, Algorithms and Intelligent Machines*. Oxford: Oxford University Press.

Bowie, Liz, and Griffith, Kristen. March 26th, 2024. Baltimore County Principal's Racist Comments Faked by AI, Experts Say. Originally published on *Baltimore Banner*. Retrieved on American

Renaissance: https://www.amren.com/news/2024/04/baltimore-county-principals-racist-comments-faked-by-ai-experts-say/.

Cartisano, Marco. October 25th, 2023. "Deepfake porno con Giorgia Meloni, ecco tutti i reati commessi", *Agenda Digitale*. https://www.agendadigitale.eu/sicurezza/privacy/deepfake-porno-con-giorgia-meloni-ecco-tutti-i-reati-commessi/.

de Quetteville, Harry. May 14th, 2023. "The Turkish deepfake porn video could change the future of elections". The Telegraph. https://www.telegraph.co.uk/news/2023/05/14/turkey-deepfake-elections-erdogan-muharrem-ince/

European Commission. 2022. "Regulation on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)".

European Commission. 2024. "Regulation on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts".

Farid, Hany. 2022. "Creating, Using, Misusing, and Detecting Deep Fakes". *Journal of Online Trust and Safety*, Stanford Internet Observatory. https://doi.org/10.54501/jots.v1i4.56.

Goldstein, Josh A., Sastry, Girish, Musser, Micah, DiResta, Renée, Gentzel, Matthew and Sedova, Katerina. 2023. "Generative language models and automated influence operations: Emerging threats and potential mitigations". arXiv preprint arXiv:2301.04246

Gosse, Chandell E. and Burkell, Jacquelyn. 2020. Politics and porn: how news media characterizes problems presented by deepfakes. *Critical Studies in Media Communication*, 37(5), 497-511. https://doi.org/10.1080/15295036.2020.1832697.

Granza, Mark. July 2023. In defence of 'Anons'. *IM—1776*. https://im1776.com/2023/07/11/in-defense-of-anons/.

Heikkilä, Melissa. 2024. "Bans on deepfakes take us only so far—here's what we really need". *MIT Technology Review*. https://www.technologyreview.com/2024/02/27/1089010/bans-on-deepfakes-take-us-only-so-far-heres-what-we-really-need.

Hibberd, James and Phillips, Zoe G. January 25th, 2024. "Taylor Swift AI Nudes Provoke Fandom Uproar on X: 'Disgusting as Hell'". *The Hollywood Reporter*. https://www.hollywoodreporter.com/news/general-news/taylor-swift-ai-nudes-twitter-x-1235806979/.

Hood, Gregory. March 1st, 2024. "Anti-White AI Is No Laughing Matter". *American Renaissance*. https://www.amren.com/features/2024/03/anti-white-ai-is-no-laughing-matter/

Hood, Gregory. May 4th, 2022. "A Contemptible 'Disinformation Governance Board'". *American Renaissance*. https://dev.amren.com/news/2022/05/a-contemptible-disinformation-governance-board/.

Hunter, Tatum. February 13th, 2023. "AI porn is easy to make now. For women, that's a nightmare". *The Washington Post*. https://www.washingtonpost.com/technology/2023/02/13/ai-porn-deepfakes-women-consent/.

Interesse, Giulia. December 20th, 2022. "China to Regulate Deep Synthesis (Deepfake) Technology Starting 2023". *China Briefing*. https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/.

JRE Daily Clips. 2023. "Joe Rogan: 'I Wasn't Afraid of AI Until I Learned This.'" YouTube video, December 19, 2023. Archived at Wayback Machine: https://web.archive.org/web/20231228023448/https://www.youtube.com/watch?v=2yd18z6iSyk.

JRE Daily Clips. 2024. "Joe Rogan & Jordan Peterson: The Dark Triad & 'Lulz' Culture Online." YouTube video, January 29, 2024. Archived at Wayback Machine: https://web.archive.org/web/20230314161948/https://www.youtube.com/watch?v=n3obXGFknMw.

Lever, Rob. November 29th, 2023. "Deepfake mocks Alexandria Ocasio-Cortez's call for Gaza ceasefire". *AFP Fack Chek*. https://factcheck.afp.com/doc.afp.com.34693MW.

Lolli, Alessandro. 2020. *La guerra dei meme: fenomenologia di uno scherzo infinito*. effequ.

Meskys, Edvinas, Kalpokiene, Julija, Jurcys, Paul, and Liaudanskas, Aidas. 2019. "Regulating Deep Fakes: Legal and Ethical Considerations". *Journal of Intellectual Property Law & Practice* 15 (1), 24–31. https://ssrn.com/abstract=3497144.

Nagle, Angela. 2017. *Kill all normies: Online culture wars from 4chan and Tumblr to Trump and the alt-right*. John Hunt Publishing.

Oversight Board. April 2024. Oversight Board announces two new cases on explicit AI images of female public figures. https://oversightboard.com/news/361856206851167-oversight-board-announces-two-new-cases-on-explicit-ai-images-of-female-public-figures/.

Paris, Britt. 2021. "Configuring Fakes: Digitalized Bodies, the Politics of Evidence and Agency". *Social Media + Society*, https://doi.org/10.1177/20563051211062919.

Peterson, Jordan. 2019. "I Didn't Say That", Jordanbpeterson.com, 22 August 2019, www.jordanbpeterson.com/blog-posts/i-didnt-say-that (March 24, 2024).

Pierre, Gustavo. 2021. Inside the Twisted World of AOC Deepfake Porn. *Countere*, 7 April 2021. https://www.countere.com/home/aoc-deepfake-porn.

Postill, John. 2024. *The Anthropology of Digital Practices: Dispatches from the Online Culture Wars*. Taylor & Francis Ltd.

Reuters Fact Check. 2023. "Video features deepfakes of Nancy Pelosi, Alexandria Ocasio-Cortez and Joe Biden". *Reuters*, 28 August 2023. https://www.reuters.com/article/factcheck-pelosi-deepfake-idUSL1N36V2E0/.

Roose, Kevin. 2018. "Here Come the Fake Videos, Too". *The New York Times*, 4 March 2018, https://www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html.

Rufo, Christopher. 2024. The New Right Activism. *IM-1776*, 14 January 2024, https://im1776.com/manifesto-counterrevolution/.

SAG-AFTRA. 2024. "SAG-AFTRA Statement on Taylor Swift and George Carlin Digital Fakes". *SAG-AFTRA*, 26 January 2024, https://www.sagaftra.org/sag-aftra-statement-taylor-swift-and-george-carlin-digital-fakes.

Scott, Mark. 2024. "Deepfake porn is political violence". *Politico*, 8 February 2024, https://www.politico.eu/newsletter/digital-bridge/deepfake-porn-is-political-violence/.

Sayler, Kelly M. and Harris, Laurie A. 2022. "Deep Fakes and National Security". *Congressional Research Services*. https://crsreports.congress.gov/product/pdf/IF/IF11333

Sjölin, Catarina. 2015. "Ten years on: consent under the Sexual Offences Act 2003." *The Journal of Criminal Law,* 79.1, 20-35, https://doi.org/10.1177/0022018314566744.

Taylor, Jared. "What You Can and Can't Do With AI". *American Renaissance*, 15 February 2024. https://www.amren.com/videos/2024/02/what-you-can-and-cant-do-with-ai/.

Thompson, Stuart A. 2024. "How 'Deepfake Elon Musk' Became the Internet's Biggest Scammer". *New York Times,* 14 August 2024, https://www.nytimes.com/interactive/2024/08/14/technology/elon-musk-ai-deepfake-scam.html.

The Economist. 2024. "Is Google's Gemini chatbot woke by accident, or by design?". *The Economist*, 28 February 2024, https://www.economist.com/united-states/2024/02/28/is-googles-gemini-chatbot-woke-by-accident-or-design.

Verdoliva, Luisa. 2020. "Media Forensics and DeepFakes: An Overview". *IEEE Journal on Selected Topics in Signal Processing*, 14 (5), https://doi.org/10.1109/JSTSP.2020.3002101.

Verma, Pranshu. 2023. "AI fake nudes are booming. It's ruining real teens' lives.", *The Washington Post,* 5 November 2023. https://www.washingtonpost.com/technology/2023/11/05/ai-deepfake-porn-teens-women-impact.

Wagner, Travis L. and Blewer, Ashley. 2019. "The Word Real Is No Longer Real": Deepfakes, Gender, and the Challenges of AI-Altered Video. *Open Information Science*, 3(1), 32-46. https://doi.org/10.1515/opis-2019-0003.

Timcast. 2024. "Graphic Adult Images Of Taylor Swift GO VIRAL, AI Generated Photos Of Swift Spark OUTRAGE", YouTube video, January 25, 2024: https://www.youtube.com/watch?v=BRIjBHimzyM.

The Jordan B. Peterson Clip. 2023. "How AI Is Reshaping Society and What Jordan Peterson Has To Say About It", YouTube video, February 22, 2023: https://www.youtube.com/watch?v=Qh0WQWACCDg.

The Jordan B. Peterson Podcast. 2024. "418. Hedonism, Taboos, Society and Deprivation", Spotify, January 29, 2024: https://open.spotify.com/episode/3fp3iyBe6hGCpl0oax1yJC (April 28th, 2024).

The Unz Review. 2024. "Rosa Parks and Oprah Are 'Notable White Americans'." https://www.unz.com/audio/jtaylor_rosa-parks-and-oprah-are-notable-white-americans.

Valuetainment. 2024. "'Should It Be Illegal?' - Controversy Over Taylor Swift AI Nude Images", Youtube video, January 31, 2024: https://www.youtube.com/watch?v=pJ6QhIKrpPc.

Zhang, Angela H. 2024. The Promise and Perils of China's Regulation of Artificial Intelligence. *University of Hong Kong Faculty of Law Research Paper*, 2024/02. http://dx.doi.org/10.2139/ssrn.4708676.

Žižek, Slavoj. 2024. Instagram post. https://www.instagram.com/p/C3oBTpaBb7O.