**A Multidisciplinary Look at History and Future of Deepfake with Gartner Hype Cycle**

Can Yavuz

Ghent University

can.yavuz@ugent.be

***Abstract*—Deepfake has been rapidly developed, misused, and democratized in a handful of years, which created an overhype around this technology. To separate this overhype from reality, this study presents a multidisciplinary (technical, legal, and societal) history of deepfake and speculates on the future of deepfake using the Gartner Hype Cycle. The Hype Cycle explains the typical trajectory that emerging technology traverses during its development and technology adoption in reference to its visibility and maturity and has proved to be a practical instrument to set realistic expectations for nascent technologies.**

**Keywords**

Deepfake, Generative Artificial Intelligence, Gartner Hype Cycle, Technology Adoption, Generative Adversarial Networks, Autoencoder, Diffusion Models, Artificial Intelligence Spring, Artificial Intelligence Winter

**Introduction**

Generative adversarial networks (GANs), autoencoder, and diffusion models can be considered three of the most ingenious inventions of the ongoing AI spring. Without diving into technical details, an aspect of these AI architectures should be emphasized: they can create synthetic but hyper-realistic audio, photo, and video. This hyper-realistic synthetic audiovisual media, in a wide sense, is referred to as deepfake. Noticing a variety of definitions of deepfake have been suggested [1], [2], this study defines deepfake as realistic synthetic audiovisual media (audio, photo, and video) that is entirely created or partially edited by an AI system. Taking an etymological look at deepfake can also help comprehend it. Deepfake is a portmanteau of the terms deep and fake. The term deep refers to deep learning-based creation and editing of deepfake, and fake denotes the synthetic nature of deepfake.

The term deepfake was coined after a Redditor (also named Deepfakes), who created a subreddit on AI-generated nonconsensual pornography and publicly shared its open-source code [1]. Nonconsensual pornography was the catalyst of deepfake technology; a 2019 study [3] demonstrated that 96% of deepfakes were pornographic. Additionally, deepfake emerged in the age of post-truth politics, which fueled concerns over the misuse of deepfake for dis- and misinformation. Face presentation attacks [4], deepfake-powered child pornography [5] and online grooming [6], fabricated evidence [7], financial fraud, and identity theft are other perils of deepfake. On top of the malicious uses and growing concerns, the media covered deepfake with eye-catching headlines, which will be detailed in Section 3 (see **Table 1**). These factors resulted in a great deal of (negative) overhype around the capabilities and societal impacts of deepfake. Before proceeding further, let’s take a step back and discuss the overhype around nascent technologies and its potential impacts.

According to Amara’s law, ‘we tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run’ [8], which is visible in the history of AI. The AI springs created short-term overhype that brought sizeable investment and immense enthusiasm in the AI domain. Subsequently, the gap between short-term overhype and reality (the challenges of developing and adopting new technologies) brought disillusionment and skepticism about the capabilities and future of AI, as exemplified in the Lighthill report [9], [10], [11]. A combination of many factors beyond the scope of this study paved the way for AI winters, one of which was media overhype [12].

It can be argued that a similar overhype is emerging around deepfake due to its swift development and democratization, enormous potential, and overhyped media coverage (as will be detailed in Sections 2 and 3). As AI history reminds us, if deepfake technology fails to meet the inflated expectations attributed to it, the likelihood of a ‘deepfake winter’ escalates. Consequently, setting realistic expectations for deepfake (without forgetting its great long-term potential) is of utmost importance for anyone interested in synthetic media, especially for technology and innovation management domains.

This raises the question of how to set feasible expectations for exciting nascent technology. One of the instruments that has proved to be useful in this endeavor is the Gartner Hype Cycle. It helps separate overhype from reality while characterizing the potential progression of emerging technologies. Developed by technological research and consulting firm Gartner, the Hype Cycle ‘explains a general path a technology takes over time, in terms of expectations or visibility of the value of the technology’ [13]. Many sectors employ the Gartner Hype Cycle as a working management tool to track innovations because it can shed light on uncertainties surrounding the development and adoption of emerging technologies [13], [14].

Gartner Hype Cycle speculates how an emerging technology can mature and be adopted by society and posits that it passes five stages on its path to productivity. These phases are 1) the innovation trigger, 2) the peak of inflated expectations, 3) the trough of disillusionment, 4) the slope of enlightenment, and 5) the plateau of productivity. Now, let’s elucidate these stages.

The innovation trigger phase immediately follows the innovation. Generally, a technological advancement that has been under development for quite some time comes to light and starts to generate excitement about its potential. In this phase, companies seeking a first-mover advantage consider investment in technology [13], [14], [15].

The innovation trigger is followed by the peak of inflated expectations, a phase marked by over-enthusiasm about the capabilities of emerging technology. Media overhype fuels excitement about the capabilities of technology. Along with the overhype, companies that aspire to gain a competitive edge enthusiastically embrace the nascent technology with high hopes. Companies that do not want to be left behind follow suit (often without a sound business case) instead of carefully assessing capabilities, development, and adoption of technology, which creates a bandwagon effect [13], [14].

The third stage of the Hype Cycle—the trough of disillusionment—is a phase of facing reality and replacing original excitement by impatience as technology fails to progress swiftly and meets overhyped expectations [13]. ‘Problems with performance, or slower-than-expected adoption, or a failure to deliver financial returns in the time anticipated all lead to missed expectations’ [14]. The media reshares the initial success stories of the technology until they become outdated. As the need for fresh stories arises, the media shifts its focus to highlighting challenges and presenting less favorable aspects of the technology [14].

In the fourth stage—the slope of enlightenment—the early adopters and organizations overcome the initial issues and gain a more profound and contextual understanding of the technology and its practical applications [15]. The light at the end of the tunnel renews interest in technology [14]; thus, this stage is also referred to as ‘the time-to-value gap’ [13]. The final stage of the Hype Cycle is the plateau of productivity, where technology’s real-world benefits and productive value are widely demonstrated and accepted [15]. Due to declined risk factors, companies exhibit a greater propensity to invest in and deploy technology [14].

**Figure 1** summarizes the Gartner Hype Cycle and its stages.

A diagram of a line graph

Description automatically generated with medium confidence

**Figure 1. Gartner Hype Cycle** [13]**.**

Having concluded that deepfake is an emerging technology surrounded by overhype, and the Gartner Hype Cycle is a useful instrument to separate the overhype from reality and speculate on the trajectory of development and adoption of technology, we can move on to the aim and outline of the study.

The major objectives of this study are to 1) provide a technical, societal, and legal history and future of deepfake, 2) separate overhype surrounding deepfake from reality, and 3) speculate on the development and technology adoption of deepfake with Gartner Hype Cycle. The study can be considered descriptive and speculative, akin to the Gartner Hype Cycle.

The study has been divided into five sections. The initial section examines the requisite duration for the development and technology adoption of deepfake, and Sections 2, 3, 4, and 5 trace deepfake’s journey on the Gartner Hype Cycle.

**Methodology**

The study conducts secondary research and interviews with three leading experts from academia, industry, and civil society (see **Appendix A**) to fulfill its aims.

Moving on to the limitations and delimitations of the study, firstly, the small sample size of interviews may generate bias and makes it difficult to generalize the research results. Secondly, the study concentrates on deepfake generation and pays no attention to deepfake detection. As of writing, deepfake detection is one of the leading methods to mitigate the misuse of deepfake, which is relevant to the impacts of deepfake. Thirdly, there are uncertainties about the development, technology adoption, regulation, and societal impacts of deepfake. Because they depend on complex economic, technological, educational, and political factors that are subject to change at any time, the forecast on the future and technology adoption of deepfake should be taken with reservation. Lastly, the Gartner Hype Cycle is open to criticism regarding its theoretical underpinnings, industry-centric approach, lack of consideration for different stakeholders, and oversimplifying the trajectory of emerging technologies [13]. A study based on the Hype Cycle naturally inherits its limitations.

**1. Time-to Maturity Assessment: Deepfake as a Long Fuse Technology**

Technologies may follow divergent trajectories to travel through the Hype Cycle. Gartner suggests three adoption speeds for emerging technologies: 1) fast track, 2) normal, and 3) long-fuse, which typically take two to four, five to eight, and ten to fifteen years to traverse the Hype Cycle, respectively [15].

Maturing in a decade or over, the overhype around the capabilities of the technology, regulation issues, and dependence on professional skills in short supply are some of the typical features of long-fuse technologies, according to Gartner [15]. Deepfake possesses each of long-fuse technology characteristics. Firstly, the technology behind deepfake is expected to mature (reach the plateau of productivity) between eleven to seventeen years [16], [17]. Secondly, the media created an overhype surrounding deepfake, as will be detailed in Section 3. Thirdly, hyperrealistic deepfake generation (particularly deepfake video), at least for now, requires skills that are short in supply. Lastly, there are many regulation issues around deepfake, e.g., how to tackle deepfake generated dis- and misinformation, safeguard post-mortem privacy from deepfake, enforce the labeling requirement for deepfake, and prevent nonconsensual deepfake pornography. Having concluded that deepfake is a long-fuse technology, the following sections trace deepfake’s journey on the Gartner Hype Cycle.

**2. Innovation Trigger**

The origins of deepfake and the technology behind it can be traced back to the late 1990s. Video Rewrite [18] (1997) was the first automated system that could create lip-syncing videos, ancient examples of deepfake. Another technological breakthrough was active appearance models [19] (2001), a computer vision algorithm that matched statistical models of appearance to images. Furthermore, the advancement of deep learning, increasing computing power, and the ease of access to a large volume of data prepared the ground for the technology trigger of deepfake. Ian Goodfellow invented the technology behind deepfake (GANs) [20] in 2014.

GANs consist of two neural networks—generator and discriminator—that compete against each other in a zero-sum game. The generator manufactures fake data, and the discriminator tries to distinguish authentic samples from fake ones. Analogously, the generator takes the role of counterfeiter and prints counterfeit money; the discriminator plays the role of detective and evaluates the banknotes’ authenticity. Two neural networks sharpen each other thanks to the feedback loop between them [20].

‘Not all synthetic images are born equally and by same manufacture methods’ [21]. Another neural network architecture used to generate deepfake is variational autoencoder (VAE), a modernized version of traditional autoencoder. Generic VAE is skillful for deepfake manufacture because it ‘generates latent vectors that follow a Gaussian unit distribution. By doing this, it allows us to generate new images by sampling a latent vector from the Gaussian distribution, which could then be passed to the decoder network’ [22].

The third model behind deepfake is diffusion models, a variety of probabilistic generative models based on the forward diffusion stage (which progressively injects noise and destroys data) and forward diffusion stage (which reverses the process for sample generation). A famous example of diffusion models is OpenAI’s text-to-image model, DALL·E 3. Diffusion models are considered the new driving force of generative models and show great potential in many tasks, e.g., image generation, image super-resolution, image inpainting, and mage-to-image translation [23], [24]. The above-stated three models have advantages and drawbacks against each other [21], [22], [23], and each has a wide range of variations [23].

After a concise account of the technology behind deepfake, we can turn to the Hype Cycle analysis. The first wave of venture capital investment is a stage indicator of technology trigger, as shown in **Figure 1**. This investment in deepfake arrived three years after the invention of GANs [25].

**3. Peak of Inflated Expectations**

As stated in the Introduction, the term deepfake was coined after a Redditor who made the deep learning code behind nonconsensual deepfake pornography publicly available (November 2017), which sparked the media interest in deepfake [3], subsequent to Cole’s article [26]. Social media platforms were the first responders to emerging misuse of deepfake: Discord, Gfycat, Twitter, Pornhub, and Reddit swiftly banned nonconsensual deepfake pornography (February 2018) [27]. Nonconsensual deepfake pornography served as a tool to discredit and silence investigative journalist Rana Ayyub (April 2018) [2].

The researchers voiced concerns about the misuse of deepfake beyond nonconsensual pornography and labeled deepfake as ‘a looming challenge for privacy, democracy, and national security’ (July 2018) [28]. The incidents in Gabon, Malaysia, and Ukraine echoed the concerns. Gabon’s president released a strange New Year’s speech video after months of absence from political life due to his deteriorating health. The New Year’s speech video was deemed as a deepfake and provoked a *coup d’état* attempt (January 2019). Another video allegedly showing the Malaysian finance minister engaging in sexual conduct with a man sparked debates in a country that criminalizes same-sex sexual conduct. The finance minister Azmin denied the video’s authenticity, stating it was a deepfake (June 2019) [2]. Lastly, deepfake was weaponized in the Russia-Ukraine war. A deepfake video of the Ukrainian president announcing surrender to Russia circulated on social media and quickly debunked (March 2022) [29].

Deepfake Tom Cruise [30] was born during the above-stated debates (February 2021). The marriage of the acting skills of the actor behind deepfake, his resemblance to Tom Cruise, and the expertise of the deepfake producer made deepfake Tom Cruise look more realistic compared to previous deepfakes and took public interest in deepfake to new heights. According to Google Trends [31], the first peak of the search query ‘deepfake’ occurred in March 2021, coinciding with the first deepfake Tom Cruise video.

In addition to Google Trends data, an assessment of mainstream media coverage of deepfake can provide valuable insights into the visibility and level of hype surrounding this technology. It may be an important indicator because mainstream media is a powerful influencer that can shape hype dynamics. **Table 1** demonstrates mainstream media’s headlines covering deepfake.

**Table 1. Mainstream Media’s (Overhyped) Coverage of Deepfake**

|  |  |  |
| --- | --- | --- |
| Media Outlet | Headline | Date |
| The New York Times [32] | Will Deep-Fake Technology Destroy Democracy? | 17 October 2017 |
| MIT Technology Review[33] | AI Could Set Us Back 100 Years When It Comes to How We Consume News | 7 November 2017 |
| The Vice[26] | AI-Assisted Fake Porn Is Here and We’re All Fucked | 11 December 2017 |
| The Guardian[34] | You thought fake news was bad? Deep fakes are where truth goes to die | 12 November 2018 |
| Politico[35] | Welcome to the age of uncertainty | 17 December 2019 |
| Forbes[36] | Deepfakes Are Going to Wreak Havoc On Society. We Are Not Prepared | 25 May 2020 |
| The Washington Post[37] | Deepfakes are coming for American democracy. Here’s how we can prepare | 10 September 2020 |
| Venture Beat[38] | Why deepfake phishing is a disaster waiting to happen | 9 December 2022 |
| Newsweek[39] | Deepfakes Could Destroy the 2024 Election | 24 March 2023 |
| Thomson Reuters[40] | Seeing is no longer believing — the rise of deepfakes | 18 July 2023 |
| inews[41] | How deepfake videos could derail democracy at the next general election | 27 October 2023 |

This study acknowledges the misuse of deepfake technology and its increasing potential as a powerful tool for malicious purposes. There is a pressing need for amendments, educational initiatives, policy interventions, and technical measures to tackle this issue. Nonetheless, deepfake destroying democracy, truth, trust, and elections and setting news consumption back to a century ago could be considered threat inflation. This term is defined as ‘the attempt by elites to create concern for a threat that goes beyond the scope and urgency that a disinterested analysis would justify [42].’ [43] also concluded that panic is the default tone of most media coverage of deepfake.

In addition to the eye-catching headlines, credible media outlets reported the stories of how a deepfake-powered impersonation attack in a video call deceived European parliamentarians [44] and a fifty-year-old woman—lacking professional knowledge—effortlessly manufactured a convincing deepfake video to harass her daughter’s cheerleader rival [45]. These stories can also be regarded as threat inflation because deepfake was never utilized in the incidents. Also, the democratization of deepfake arguably fueled the overhype [21]. Consequently, the ‘fear of deepfakes seems to have outpaced the technology itself’ in recent years [44], and WITNESS—a prominent non-governmental organization focused on synthetic media—suggested to ‘de-escalate rhetoric’ on deepfakes [46].

The study concludes that overhyped media coverage—combined with the rapid advancement, democratization, and exciting potential of deepfake—created (negative) overhype around the capabilities and societal impacts of deepfake during this phase. Such media overhype is the defining characteristic of the peak of inflated expectations stage[13].

It is challenging to determine deepfake’s current position on the Hype Cycle. In the first stages of the Hype Cycle, the position of a technology ‘is guided more by its hype levels than its perceived maturity’ [15]. Recalling the ongoing media overhype around deepfake, one may speculate that deepfake is currently traversing the peak of inflated expectations stage.

The growing concerns and democratization of deepfake encouraged novel initiatives, policies, and laws to fight back against the misuse of deepfake. The Content Authenticity Initiative [47] and the Coalition for Content Provenance and Authenticity [48] were formed to curb disinformation and address content authenticity. Another noteworthy development was the ban of deepfake training on Google Colaboratory—a free and popular cloud-based service—due to potential exploitation of the platform to generate malicious deepfakes (May 2022) [49].

Turning to the legal responses against misuse of deepfake, 2019 was the year the law and policymakers arrived on the scene. The U.S. and Chinese lawmakers were particularly proactive in this stage. Starting with the USA, the United States Congress held a hearing on the national security concerns over malicious deepfakes in 2019 [50]. Subsequently, Virginia criminalized nonconsensual deepfake pornography [51]. To tackle the misuse of AI with AI, the U.S. Department of Defense [52], [53] and Facebook [54] launched deepfake detection programs (August and September 2019). As the concerns about the misuse of deepfake extended beyond nonconsensual pornography, Texas [55] and California [56] criminalized the use of malicious deepfakes that target elections (September and October 2019). New York approached deepfake from a different perspective and established the postmortem right of publicity [57] (October 2020). One month after the turbulent 2020 United States presidential election that exemplified the dire consequences of dis- and misinformation, the United States Congress enacted an act to support research on deepfake generation and detection (December 2020) [58].

Turning to the other side of the Atlantic, the European policymakers lagged behind their U.S. counterparts and noticed deepfake with a delay. The European Parliamentary Research Service published a comprehensive report on deepfake (July 2021) [1]. The Strengthened Code of Practice on Disinformation 2022 enacted transparency and content moderation obligations to major online platforms (June 2022), and the EU AI Act (Article 52) imposes transparency obligations for deepfake.

While the United States and the European Union were trying to gain a deeper understanding of the societal impacts of deepfake, China became the pioneer of deepfake regulations. China’s comprehensive deepfake laws (the Provisions on the Administration of Deep Synthesis Internet Information Services [the supplementary of 2019 regulations]) came into force in January 2023 [59].

**4. Trough of Disillusionment**

Deepfake manufacturing models have significantly advanced in recent years thanks to numerous GANs, autoencoder, and diffusion models variants [21], [22] [23], [60], [61]. A retrospective look at deepfake supports this argument. **Figure 2** visualizes GANs’ progress in face generation from 2014 to 2022. Deepfake videos of Tom Cruise [62] and music band AllttA’s duet with deepfake Jay-Z [63] demonstrate that hyperrealistic deepfake production is now within the realms of possibility. Lastly, OpenAI’s text-to-video model Sora indicates the capabilities of text-to-video on the horizon [64].



**Figure 2. Face Synthesis Deepfakes Between 2014-2022** [29]**.**

While automated systems can effortlessly produce fairly realistic deepfakes, hyperrealistic deepfake (without any inconsistencies) production typically necessitates the finishing touch experts to a certain extent, as of writing. This is particularly true for deepfake videos. The creator of deepfake videos of Tom Cruise reminded that hyperrealistic deepfake videos cannot be produced by ‘just pressing a button’ [65], which holds true at least for now. Turning to audio deepfake, a leading audio deepfake company collaborates with both machine learning experts and sound engineers to achieve hyperrealism [66].

Regarding deepfake’s future from a technical viewpoint, the technology behind deepfake faces many limitations, albeit the timeline in **Figure 2** is tremendously encouraging. Some of the current generation of deepfake manufacturing methods slowly showed signs of reaching their potential. Working around their shortcomings may not be sufficient to produce more realistic deepfakes than the existing ones [67]. Research also underlines serious challenges and open issues of deepfake image, audio, and video generation [21], [23], [61], [68]. In other words, further innovations are necessary to advance deepfake generation. This may be the start of the trough of disillusionment phase. Because technological progress often follows the S-curve and once technological limits are reached, the progress (the advancement to the next generation) may become increasingly difficult, time-consuming, and expensive [69]. Therefore, the progression of deepfake technology may not proceed at the rapid pace observed during its initial stages. Resolving the challenges, addressing the open issues outlined in the existing literature, and meeting overhyped expectations may take some time.

**5. Slope of Enlightenment and Plateau of Productivity**

Deepfake is an emerging technology that has yet to reach its full potential. Foreseeing an emerging technology’s technical development and especially societal impacts is challenging. This challenge stems from the fact that the technical side of technology swiftly advances and outpaces our understanding of the emerging technology’s societal impacts [70]. Despite this, the following paragraphs will speculate on the future of deepfake from technical, legal, and societal perspectives, respectively.

Commencing with the possible technical developments, two forthcoming trends are democratization and enhanced quality of deepfake. As stated above, present deepfake generation methods seem to be pushed to their limits and face many limitations. Innovations are expected to help overcome limitations at the cost of increasing the entry barrier for the hyperrealistic deepfake generation and slowing down the democratization of the hyperrealistic deepfake generation [67]. That said, this delay may be short-lived. Some experts argue that deepfake generation may become easier and more accessible before 2028 [71]. Also, Gartner suggests that businesses seeking a competitive edge should include diffusion models in their investment plans between 2025 and 2027 [17]. The consulting firm also forecasts that GANs may reach the plateau of productivity between 2026 and 2031 [16]. This means that the real-world benefits of GANs may become visible and accepted, and one-third of the target audience of GANs can adopt this technology [15]. The projected advancements can empower the democratization of deepfake, resulting in the proliferation of both deepfake creators and deepfake content.

The forecasted success of deepfake should not come as a surprise because deepfake has a wide range of use cases and can create enormous societal and economic value. [72] argues that thanks to deepfakes, ‘we are about to see a major paradigm shift in media creation and consumption that will likely change the equation for entire industries.’

Turning to the technical side of deepfake, it is anticipated that the next generation deepfakes will look more natural and the realism gap between authentic and synthetic media will be shorten [21]. More realistic deepfake audio, full-body puppetry, real-time deepfake and more capable text-to-image and text-to-video models seem in the not-too-distant future [61], [68].

Regarding the potential legal developments related to deepfake, the prospective trends in the short term are the criminalization of nonconsensual deepfake pornography, the labelling requirement for deepfake, and the introduction of post-mortem privacy and data protection (protection against hyperrealistic deepfake avatar created without the consent of the deceased). It has been noted that effective enforcement of these regulations can be burdensome [73]. Also, the creativity and realistic appearance of deepfake may exacerbate debates around the limits of freedom of speech because the hyperrealism of deepfake challenges the traditional cure of dis- and misinformation, counterspeech doctrine [74]. Another development may be further self-regulation of social media platforms due to the mounting pressure of governments to combat deepfake-powered online dis- and misinformation.

Deepfake also challenges the status quo of intellectual property rights. Scraping copyright-protected works from the Internet without permission and compensation to feed generative AI systems may force lawmakers to rethink the limits and protection of intellectual property rights. The last potential legal development concerns the right to privacy. While advancements in content authenticity and provenance technologies are being pursued to counteract nefarious deepfakes, it is essential to acknowledge that these innovations may, in turn, present challenges to the right to privacy [75].

Shifting our focus to the potential societal developments, the pressing topic concerns the health of information ecosystems. Deepfake is expected to further personalized content creation and raise the number of synthetic media. Malicious actors can exploit these advancements to exacerbate dis- and misinformation.

The elevated number and realism of deepfake can contest the established conception that audiovisual media accurately represents reality. This could, in turn, diminish the epistemic value of audiovisual media [76]. It may present challenges for individuals who lack critical thinking skills and basic media, computer, and artificial intelligence literacy. This combined with the problematic business model of social media platforms—which profit takes precedence over all other considerations [77]—may exacerbate the adverse impacts resulting from the misuse of deepfakes.

Looking on the bright side, the novelty of deepfake may erode as society becomes familiar with this technology. Society does not have a fixed ability to perceive reality and has adopted novel media forms on many occasions, and a classic example dates back to the 19th century. [78] recalls the first screening of the short documentary movie *the Arrival of a Train at La Ciotat Station* (1896), which displays the entry of a train to a railway station. The audience—having their first cinema experience—was terrorized during the screening because the projected train looked like it rushed towards them. Today, such a reaction seems strange because the novelty of cinema has worn off. Due to the novelty of deepfake, society is gravely concerned about deepfake, akin to the audience of *the Arrival of a Train at La Ciotat Station*. Nonetheless, the inventor of the technology behind deepfake foretells how society may adjust its norms in the age of deepfake.

Ian Goodfellow is concerned about the threat of malicious deepfakes for the upcoming years; however, he shows a lack concern for the medium and long term. Goodfellow is of the opinion that the epistemic value of audiovisual media is high nowadays due to its realistic appearance. In the next two decades, ‘there will be a bumpy cultural transition,’ and people will understand that synthetic audiovisual media can look as realistic as authentic audiovisual media. Consequently, the epistemic value of audiovisual media will not solely rely on its realistic appearance. Instead, people will question whether audiovisual media is cryptographically signed or whether its authenticity is validated by methods other than deepfake detection [79]. Similarly, Tufekci [80] argues that

‘As we shift from an era when realistic fakes were expensive and hard to create to one where they are cheap and easy, we will inevitably adjust our norms. In the past, it often made sense to believe something until it was debunked; in the future, for certain information or claims, it will start making sense to assume they are fake. Unless they are verified.’

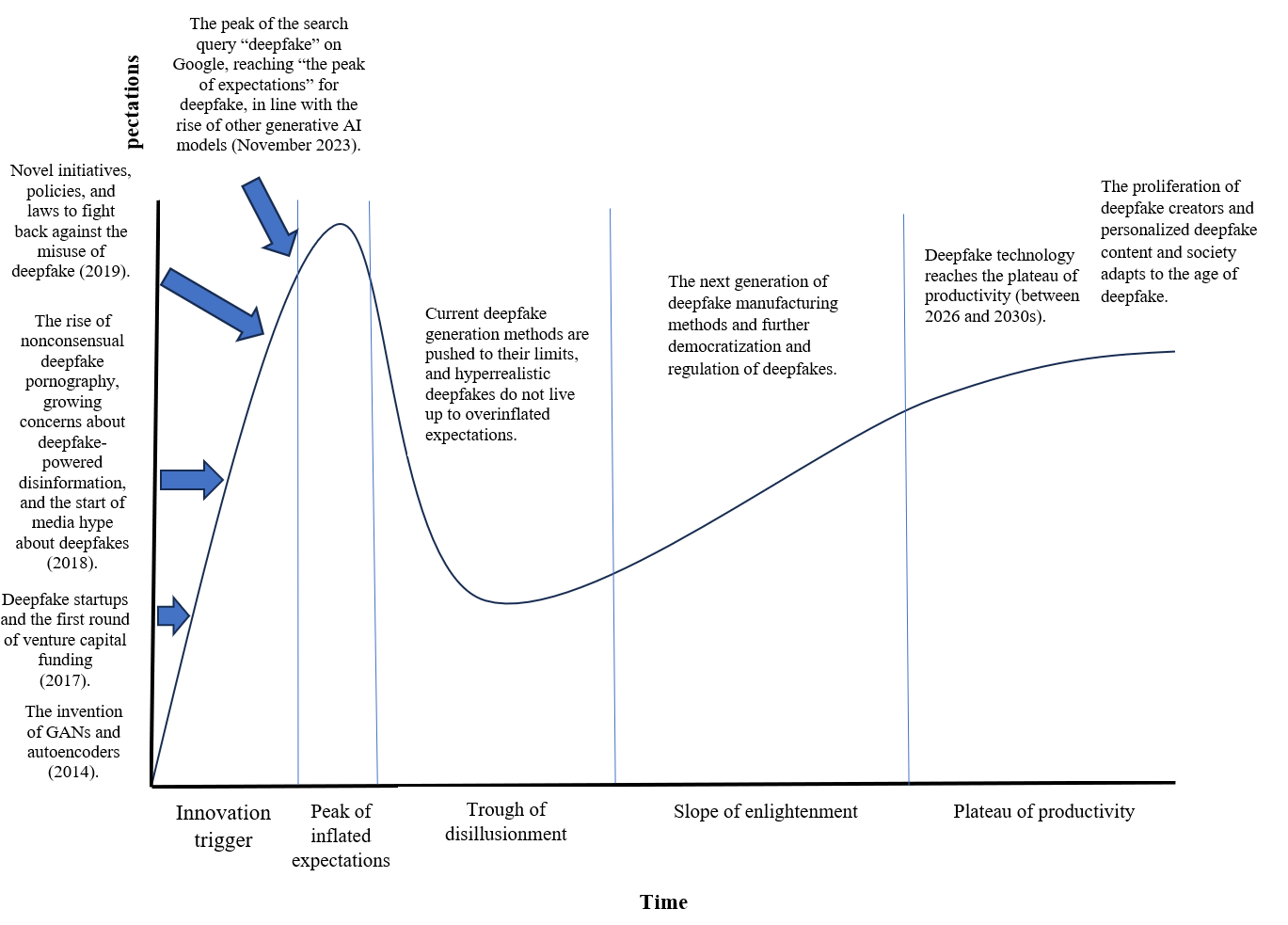
The arguments of Tufekci and Goodfellow recollect Fallis’ viewpoint. ‘Deepfake technology decreases the amount of information that videos carry by increasing the probability that realistic fake videos depicting events that never occurred will be produced’ [76]. Because it will become difficult to validate the authenticity of audiovisual media by visual and auditory senses, it is argued that seeing will not be believing. Eventually, technology may replace our senses to verify the authenticity of audiovisual media.

Sam Gregory, the executive director of WITNESS, takes a more cautious approach to the epistemic value of audiovisual media. Gregory advises caution about the ‘seeing will not be believing because of deepfake’ argument, which could serve as an excuse to dismiss critical viewpoints and authentic information on matters of public interest. Society should continue to defend the epistemic value of audiovisual media instead of abandoning it entirely while adjusting to the age of deepfake. Gregory also acknowledges that seeing is believing may become a more complex scenario in the future. Hence, the law and policymakers should ponder where the responsibility lies to defend the epistemic value of audiovisual media and consider amendments (especially regulating online intermediaries) [81].

In contrast with the media’s eye-catching headlines covering deepfake, the history of audiovisual media and expert statements demonstrate that society can adjust to the age of deepfake. The crucial point is to empower and accelerate this adaptation process by taking necessary measures: investing in deepfake detection, content provenance and authenticity, and providing education to build resilience against dis- and misinformation. In other words, there is a need to narrow the cultural gap that deepfake creates.

**6. Conclusion**

Nonconsensual pornography was the first misuse of deepfake, and there is a growing concern about weaponizing deepfake for dis- and misinformation. Combined with the rapid development and democratization of deepfake and the media’s eye-catching headlines, these created an overhype around deepfake technology, triggered the self-regulation of Internet intermediaries, and invited lawmakers to regulate deepfake technology. In this light, the history of deepfake repeats a well-known story: the rapid advancement of technology was followed by technopanic, ‘an intense public, political, and academic response to the emergence or use of media or technologies [82].’ The industry quickly responded to growing threats with self-regulation, and the law and policymakers lagged behind the developments.

Moving on to the future trajectory of deepfake, the next generation of deepfakes may look more natural and capable, and the democratization of deepfake technology may continue (with a slight delay). It is forecasted that the real-world benefits of deepfake may become visible and accepted, and one-third of the deepfake target audience can adopt this technology between 2026 and 2030s. From a legal perspective, how to effectively safeguard human rights from the misuse of deepfake remains an open question, and there are complex regulation issues around deepfake. Turning to the societal perspective, society may need to rethink the evidential value of audiovisual media in the age of deepfake and build further resilience against dis- and misinformation. **Figure 3** summarizes this study’s take on the history and future of deepfake with the Gartner Hype Cycle.

**Figure 3. Gartner Hype Cycle For Deepfake.**

**Recommendations for Future Research**

Several topics related to the scope of the article deserve further attention. Further studies focusing on the sociolegal impacts of deepfake detection—such as categorizing deepfake detection as a high-risk technology in the EU AI Act—are recommended. The questions of how to achieve widespread access to content authenticity and provenance technologies and their impacts on freedom of speech and the right to privacy remain open. From a societal perspective, how deepfake can impact the epistemic value of audiovisual media and how to narrow the cultural gap that deepfake creates could be scrutinized further.

**Appendix A - List of Interviewed Experts**

1. Noah Giansiracusa, Bentley University, Associate Professor, April 4, 2023
2. Anna Bulakh, Respeecher, Head of Ethics and Partnerships, April 6, 2023
3. Sam Gregory, WITNESS, Executive Director, April 27, 2023

**Acknowledgement**

The author expresses his deepest gratitude to Prof. Dr. Gert Vermeulen for his constructive feedback and insightful suggestions. This work was supported by the Research Foundation – Flanders (Fonds voor Wetenschappelijk Onderzoek – Vlaanderen) under grant number G016620N.

Can Yavuz is a researcher at Ghent University Faculty of Law and Criminology. His research concerns deepfake technology and its socio-legal and human rights impacts. Contact him at can.yavuz@ugent.be.

# References

|  |  |
| --- | --- |
| [1] | M. v. Huijstee, P. v. Boheemen, D. Das, L. Nierling, J. Jahnel, M. Karaboga, M. Fatun, L. Kool and G. Joost, "Tackling deepfakes in European policy," European Parliamentary Research Service, 2021. |
| [2] | N. Schick, Deepfakes: The Coming Infocalypse (e-book), 1st ed., Twelve Hachette Book Group, 2020. |
| [3] | H. Ajder, G. Patrini, F. Cavalli and L. Cullen, "The State of Deepfakes: Landscape, Threats, and Impact," Deeptrace, 2019. |
| [4] | C. Kong, S. Wang and H. Li, "Digital and Physical Face Attacks: Reviewing and One Step Further," 29 September 2022. [Online]. Available: https://arxiv.org/pdf/2209.14692.pdf. [Accessed 8 February 2024]. |
| [5] | Internet Watch Foundation, "How AI is being abused to create child sexual abuse imagery," Internet Watch Foundation, 2023. |
| [6] | Europol’s European Cybercrime Centre & United Nations Interregional Crime and Justice Research Institute & Trend Micro Research, "Malicious Uses and Abuses of Artificial Intelligence," Trend Micro Research, 2020. |
| [7] | D. C. Breen, "Silent No More: How Deepfakes Will Force Courts to Reconsider Video Admission Standards," *Journal of High Technology Law,* vol. 21, no. 1, pp. 122-164, 2021. |
| [8] | Oxford Essential Quotations, "Roy Amara 1925–2007 American futurologist," 2016. [Online]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00018679. [Accessed 13 February 2024]. |
| [9] | S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach (e-book), 1st ed., Pearson, 2020. |
| [10] | F. Chollet, Deep Learning with Python, 1st ed., Manning Publications, 2018. |
| [11] | A. Webb, The Big Nine: How the Tech Titans and Their Thinking Machines Could Warp Humanity (e-book), 1st ed., PublicAffairs, 2019. |
| [12] | M. Mitchell, Artificial Intelligence: A Guide for Thinking Humans (e-book), 1st ed., Farrar, Straus and Giroux, 2019. |
| [13] | O. Dedehayir and M. Steinert, "The hype cycle model: A review and future directions," *Technological Forecasting and Social Change,* vol. 108, pp. 28-41, 2016. |
| [14] | J. Fenn and M. Raskino, Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time, 1st ed., Massachusetts: Harvard Business Review Press, 2008. |
| [15] | J. Fenn and A. Linden, "Understanding Gartner's Hype Cycles," Gartner, 2003. |
| [16] | Gartner, "5 Trends Drive the Gartner Hype Cycle for Emerging Technologies, 2020," 8 March 2020. [Online]. Available: https://www.gartner.com/smarterwithgartner/5-trends-drive-the-gartner-hype-cycle-for-emerging-technologies-2020. [Accessed 27 January 2024]. |
| [17] | A. Zimmermann, "Understand and Exploit GenAI With Gartner’s New Impact Radar," Gartner, 21 December 2023. [Online]. Available: https://www.gartner.com/en/articles/understand-and-exploit-gen-ai-with-gartner-s-new-impact-radar. [Accessed 16 February 2024]. |
| [18] | C. Bregler, M. Covell and M. Slaney, "Video Rewrite: Driving Visual Speech with Audio," *Proceedings of the 24th annual conference on Computer graphics and interactive techniques,* pp. 1-8, 1997. |
| [19] | T. Cootes, G. Edwards and C. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 681-685, 2001. |
| [20] | I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative Adversarial Networks," 10 June 2014. [Online]. Available: https://arxiv.org/pdf/1406.2661.pdf. [Accessed 20 January 2024]. |
| [21] | J. Phillipe Cardenuto, J. Yang, R. Padilha, R. Wan, D. Moreira, H. Li, S. Wang, F. Andaló, S. Marcel and A. Rocha, "The Age of Synthetic Realities: Challenges and Opportunities," 9 June 2023. [Online]. Available: https://arxiv.org/pdf/2306.11503.pdf. [Accessed 13 February 2024]. |
| [22] | Y. Patel, S. Tanwar, R. Gupta, P. Bhattacharya, I. Davidson, R. Nyameko, S. Aluvala and V. Vimal, "Deepfake Generation and Detection: Case Study and Challenges," *IEEE Access,* vol. 11, pp. 143296-143323, 2023. |
| [23] | F.-A. Croitoru, V. Hondru, R. Tudor Ionescu and M. Shah, "Diffusion Models in Vision: A Survey," 10 September 2022. [Online]. Available: https://arxiv.org/pdf/2209.04747.pdf. [Accessed 13 February 2024]. |
| [24] | L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui and M.-H. Yang, "Diffusion Models: A Comprehensive Survey of Methods and Applications," 2 September 2022. [Online]. Available: https://arxiv.org/pdf/2209.00796.pdf. [Accessed 13 February 2024]. |
| [25] | M. Temkin, "VCs try to parse through the 'noise' of generative AI," 22 December 2022. [Online]. Available: https://pitchbook.com/news/articles/generative-ai-venture-capital-investment. [Accessed 20 February 2024]. |
| [26] | S. Cole, "AI-Assisted Fake Porn Is Here and We’re All Fucked," 11 December 2017. [Online]. Available: https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn. [Accessed 20 January 2024]. |
| [27] | S. Cole, "Twitter Is the Latest Platform to Ban AI-Generated Porn," 7 February 2018. [Online]. Available: https://www.vice.com/en/article/ywqgab/twitter-bans-deepfakes. [Accessed 23 January 2024]. |
| [28] | D. K. Citron and R. Chesney, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," *California Law Review,* vol. 107, no. 6, pp. 1753-1820, 2019. |
| [29] | Stanford University Institute for Human-Centered Artificial Intelligence, "Artificial Intelligence Index Report 2023," Stanford University Institute for Human-Centered Artificial Intelligence, 2023. |
| [30] | deeptomcruise, "Sports!," 21 February 2021. [Online]. Available: https://www.tiktok.com/@deeptomcruise/video/6932166297996233989?lang=en. [Accessed 23 January 2024]. |
| [31] | Google Trends, "Deepfake," 14 February 2024. [Online]. Available: https://trends.google.com/trends/explore?date=all&q=deepfake&hl=en-US. [Accessed 14 February 2024]. |
| [32] | J. F. Boylan, "Will Deep-Fake Technology Destroy Democracy?," 17 October 2018. [Online]. Available: https://www.nytimes.com/2018/10/17/opinion/deep-fake-technology-democracy.html. [Accessed 1 February 2024]. |
| [33] | J. Snow, "AI Could Set Us Back 100 Years When It Comes to How We Consume News," 7 October 2017. [Online]. Available: https://www.technologyreview.com/2017/11/07/147903/ai-could-send-us-back-100-years-when-it-comes-to-how-we-consume-news/. [Accessed 1 February 2024]. |
| [34] | O. Schwartz, "You thought fake news was bad? Deep fakes are where truth goes to die," 12 November 2018. [Online]. Available: https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth. [Accessed 3 February 2024]. |
| [35] | J. Delcker, "Welcome to the age of uncertainty," 17 December 2019. [Online]. Available: https://www.politico.eu/article/deepfake-videos-the-future-uncertainty/. [Accessed 14 February 2024]. |
| [36] | R. Toews, "Deepfakes Are Going To Wreak Havoc On Society. We Are Not Prepared," 25 May 2020. [Online]. Available: https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/. [Accessed 14 February 2024]. |
| [37] | C. Watts and T. Hwang, "Deepfakes are coming for American democracy. Here’s how we can prepare.," 10 September 2020. [Online]. Available: https://www.washingtonpost.com/opinions/2020/09/10/deepfakes-are-coming-american-democracy-heres-how-we-can-prepare/. [Accessed 9 February 2024]. |
| [38] | T. Keary, "Why deepfake phishing is a disaster waiting to happen," 9 December 2022. [Online]. Available: https://venturebeat.com/security/deepfake-phishing/. [Accessed 14 February 2024]. |
| [39] | J. Bickerton, "Deepfakes Could Destroy the 2024 Election," 24 March 2023. [Online]. Available: https://www.newsweek.com/deepfakes-could-destroy-2024-election-1790037. [Accessed 8 February 2024]. |
| [40] | D. Philmlee , "Practice Innovations: Seeing is no longer believing — the rise of deepfakes," 2023 July 2023. [Online]. Available: https://www.thomsonreuters.com/en-us/posts/technology/practice-innovations-deepfakes/. [Accessed 9 February 2024]. |
| [41] | M. Blackall, "How deepfake videos could derail democracy at the next general election," 27 October 2023. [Online]. Available: https://inews.co.uk/news/deepfake-videos-derail-democracy-next-general-election-2707917. [Accessed 14 February 2024]. |
| [42] | J. K. Cramer and T. Thrall, "Introduction," in *American Foreign Policy and The Politics of Fear Threat Inflation since 9/11*, J. K. Cramer and T. Thrall, Eds., Routledge, 2009, p. Introduction. |
| [43] | B. Paris and J. Donovan, "Deepfakes and Cheap Fakes. The Manipulation of Audio and Visual Evidence," Data & Society, 2019. |
| [44] | J. Vincent, "‘Deepfake’ that supposedly fooled European politicians was just a look-alike, say pranksters," 30 April 2021. [Online]. Available: https://www.theverge.com/2021/4/30/22407264/deepfake-european-polticians-leonid-volkov-vovan-lexus. [Accessed 23 January 2024]. |
| [45] | T. H. Tran, "Remember That Deepfake Cheerleader Story? Turns Out It Was Probably Nonsense," 28 August 2021. [Online]. Available: https://futurism.com/the-byte/deepfake-cheerleader-video-nonsense. [Accessed 13 February 2024]. |
| [46] | WITNESS, "Prepare, Don’t Panic: Synthetic Media and Deepfakes," [Online]. Available: https://lab.witness.org/projects/synthetic-media-and-deep-fakes/. [Accessed 5 February 2024]. |
| [47] | Content Authenticity Initiative, "Our members," [Online]. Available: https://contentauthenticity.org/our-members. [Accessed 5 February 2024]. |
| [48] | The Coalition for Content Provenance and Authenticity, "What is the Coalition for Content Provenance and Authenticity (C2PA)?," [Online]. Available: https://c2pa.org/faq/. [Accessed 27 January 2024]. |
| [49] | M. Anderson, "Google Has Banned the Training of Deepfakes in Colab," 9 December 2022. [Online]. Available: https://www.unite.ai/google-has-banned-the-training-of-deepfakes-in-colab/. [Accessed 27 January 2024]. |
| [50] | "National Security Challenges of Artificial Intelligence, Manipulated Media, and “Deepfakes”," [Online]. Available: https://www.congress.gov/event/116th-congress/house-event/109620. [Accessed 27 January 2024]. |
| [51] | The Legislative Information System, "HB 2678 Unlawful dissemination or sale of images of another; penalty.," [Online]. Available: https://lis.virginia.gov/cgi-bin/legp604.exe?191+sum+HB2678. [Accessed 27 January 2024]. |
| [52] | W. Corvey, "Media Forensics (MediFor) (Archived)," [Online]. Available: https://www.darpa.mil/program/media-forensics. [Accessed 27 January 2024]. |
| [53] | W. Corvey, "Semantic Forensics (SemaFor)," [Online]. Available: https://www.darpa.mil/program/semantic-forensics. [Accessed 27 January 2024]. |
| [54] | Facebook, "Deepfake Detection Challenge Results: An open initiative to advance AI," 12 June 2020. [Online]. Available: https://ai.facebook.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai/. [Accessed 27 January 2024]. |
| [55] | Texas Legislature Online, [Online]. Available: https://capitol.texas.gov/BillLookup/History.aspx?LegSess=86R&Bill=SB751. [Accessed 27 January 2024]. |
| [56] | California Legislative Information, "AB-730 Elections: deceptive audio or visual media.," [Online]. Available: https://leginfo.legislature.ca.gov/faces/billHistoryClient.xhtml?bill\_id=201920200AB730. [Accessed 27 January 2024]. |
| [57] | The New York State Senate, "Senate Bill S5959D," 2019. [Online]. Available: https://www.nysenate.gov/legislation/bills/2019/S5959. [Accessed 15 February 2024]. |
| [58] | Gov,info, "Public Law 116 - 258 - Identifying Outputs of Generative Adversarial Networks Act'' or the ``IOGAN Act," [Online]. Available: https://www.govinfo.gov/app/details/PLAW-116publ258. [Accessed 27 January 2024]. |
| [59] | China Law Translate, "Provisions on the Administration of Deep Synthesis Internet Information Services," 11 December 2022. [Online]. Available: https://www.chinalawtranslate.com/en/deep-synthesis/. [Accessed 8 February 2024]. |
| [60] | Z. Khanjani, G. Watson and V. P. Janeja, "Audio deepfakes: A survey," *Frontiers in Big Data,* vol. 5, pp. 1-24, 2022. |
| [61] | M. Masood, M. Nawaz, K. M. Malik, A. Javed and A. Irtaza, "Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward Show affiliations," 25 February 2021. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/2103/2103.00484.pdf. [Accessed 21 January 2023]. |
| [62] | deeptomcruise, "deeptomcruise," [Online]. Available: https://www.tiktok.com/@deeptomcruise. [Accessed 13 February 2024]. |
| [63] | AllttA, "Savages," 30 March 2023. [Online]. Available: https://www.youtube.com/watch?v=y7r6PAkFRfU. [Accessed 14 February 2024]. |
| [64] | OpenAI, "Video generation models as world simulators," OpenAI, 15 February 2024. [Online]. Available: https://openai.com/research/video-generation-models-as-world-simulators. [Accessed 16 February 2024]. |
| [65] | J. Vincent, "Tom Cruise deepfake creator says public shouldn’t be worried about ‘one-click fakes’," 5 March 2021. [Online]. Available: https://www.theverge.com/2021/3/5/22314980/tom-cruise-deepfake-tiktok-videos-ai-impersonator-chris-ume-miles-fisher. [Accessed 14 February 2024]. |
| [66] | A. Bulakh, Interviewee, *Personal Interview.* [Interview]. 6 April 2023. |
| [67] | M. Anderson, "The Future of Autoencoder-Based Deepfakes," 11 July 2022. [Online]. Available: https://archive.ph/AUuq8. [Accessed 17 January 2024]. |
| [68] | J. W. Seow, M. K. Lim, R. C. Phan and J. K. Liu, "A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities," *Neurocomputing,* vol. 513, pp. 351-371, 2022. |
| [69] | R. Foster, "Working The S-Curve: Assessing Technological Threats," *Research Management,* vol. 29, no. 4, pp. 17-20, 1986. |
| [70] | D. Collingridge, The Social Control of Technology, Frances Pinter (Eublishers), 1980. |
| [71] | M. Wong, "We Haven’t Seen the Worst of Fake News," 20 December 2022. [Online]. Available: https://www.theatlantic.com/technology/archive/2022/12/deepfake-synthetic-media-technology-rise-disinformation/672519/. [Accessed 27 January 2024]. |
| [72] | V. Riparbelli, "Our Vision for the Future of Synthetic Media," 23 June 2019. [Online]. Available: https://vriparbelli.medium.com/our-vision-for-the-future-of-synthetic-media-8791059e8f3a#:~:text=Synthetic%20media%20will%20significantly%20accelerate,digital%20realm%20begins%20and%20ends.. [Accessed 2 February 2024]. |
| [73] | B. v. d. Sloot, Y. Wagensveld and B.-J. Koops, "Deepfakes: the legal challenges of a synthetic society," Tilburg University, 2022. |
| [74] | M. A. Franks and A. E. Waldman, "Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions," *Maryland Law Review,* vol. 78, no. 4, pp. 892-898, 2019. |
| [75] | WITNESS, "Ticks or It Didn't Happen," WITNESS, 2019. |
| [76] | D. Fallis, "The Epistemic Threat of Deepfakes," *Philosophy & Technology,* vol. 34, p. 623–643, 2021. |
| [77] | S. Aral, The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health--and How We Must Adapt [e-book], 1st ed., Penguin Random House, 2020. |
| [78] | J. Bailenson, Experience on Demand: What Virtual Reality Is, How It Works, and What It Can Do (ebook), W. W. Norton & Company, 2018. |
| [79] | I. Goodfellow, Interviewee, *Ian Goodfellow: Generative Adversarial Networks (GANs) | Lex Fridman Podcast #19.* [Interview]. 18 April 2019. |
| [80] | Z. Tufekci, "The Imperfect Truth About Finding Facts in a World of Fakes," 18 February 2019. [Online]. Available: https://www.wired.com/story/zeynep-tufekci-facts-fake-news-verification/. [Accessed 23 January 2024]. |
| [81] | S. Gregory, Interviewee, *Personal interview.* [Interview]. 27 April 2023. |
| [82] | A. Thierer, "Threat Inflation, and the Danger of an Information Technology Precautionary Principle," *Minnesota Journal of Law, Science & Technology,* vol. 14, no. 1, pp. 309-386, 2013. |