Deepfakes: Deceptions, mitigations, and opportunities

Deepfakes:
Deceptions, Mitigations, and Opportunities

ABSTRACT

Deepfakes—artificial but hyper-realistic video, audio and images created by algorithms—are one of the latest technological developments in artificial intelligence. Amplified by the speed and scope of social media, they can very quickly reach millions of people and result in a wide range of marketplace deceptions. However, extant understandings of deepfake implications in the marketplace are limited and fragmented. Against this background, we develop insights into the significance of deepfakes for firms and consumers – the threats they pose, how to mitigate those threats, and the opportunities that they present. Our findings indicate that the main risks to firms include damage to image, reputation and trustworthiness, and the rapid obsolescence of existing technologies. However, consumers can also suffer blackmail, bullying, defamation, harassment, identity theft, intimidation, and revenge porn. We then accumulate and present knowledge on the strategies and mechanisms to safeguard against deepfake-based marketplace deception. Furthermore, we uncover and report the various legitimate opportunities offered by this new technology, and present an agenda for future research in this emergent and highly critical area.

Keywords: Deepfake; artificial intelligence; machine learning; opportunities; threats; challenges; protection; deception; marketing

1 INTRODUCTION

The "successful" moon mission was a hoax! The "truth" is the Apollo 11 astronauts actually never returned from the moon. In an incredibly realistic video, the then president of the United States, Richard Nixon, delivered a televised speech to the nation in a gloomy voice: *"Fate has ordained that the men who went to the moon to explore in peace will stay on the moon to rest in peace!"* A sad day for humanity! Although the Apollo 11 mission was successful in reality, this "deepfake" video[1] was created by the MIT Center for Advanced Virtuality to generate public awareness of the dangers of this emerging artificial intelligence (AI)-based technology. In the words of Francesca Panetta, the Project Co-Lead and XR Creative Director:

> *"We hope that our work will spark critical awareness among the public. We want them to be alert to what is possible with today's technology, (...) and to be ready to question what they see and hear as we enter a future fraught with challenges over the question of truth*."

Deepfakes are digitally manipulated synthetic media content (e.g., videos, images, sound clips) in which people are shown to do or say something that never existed or happened in the real world (Boush et al., 2015; Chesney & Citron, 2019; Westerlund, 2019). Advances in AI—particularly machine learning (ML) and deep neural networks (DNNs)—have contributed to the development of deepfakes (Chesney & Citron, 2019a; Dwivedi et al., 2021; Kietzmann et al., 2020; Mirsky & Lee, 2021). These look highly credible and "true to life," to the extent that distinguishing them from authentic media can be very challenging for a human (see Figure 1). Thus, they can be used for the purpose of widespread marketplace deception, with a wide range of ramifications for both firms and consumers (Europol, 2022; Luca & Zervas, 2016). In fact, a recent study by scientists from University College London ranks fake audio or video content

[1] https://www.youtube.com/watch?v=2rkQn-43ixs

as the most worrisome use of AI in terms of its potential applications for crime or terrorism (Caldwell et al., 2020). But simultaneously, this emerging technology has the potential to bring forth major business opportunities for content creation and engagement (Etienne, 2021; Farish, 2020; Kietzmann et al., 2020).

[Please Insert Figure 1 About Here]

Deception in the marketplace is ubiquitous, which makes it a fundamental issue in consumer research and marketing (Boush et al., 2015; Darke & Ritchie, 2007; Ho et al., 2016). In general, deception refers to a deliberate attempt to present others with false or omitted information, with the aim of creating a belief that the communicator himself or herself considers to be false (Darke & Ritchie, 2007; Ludwig et al., 2016; Xiao & Benbasat, 2011). Thus, it is an intentional act, accomplished by the manipulation of information, and which has an end goal of creating a false belief in others' minds (i.e., deceiving parties), all of which can be further increased through deepfakes and hurt consumers and firms alike (Xiao & Benbasat, 2011). Deception permeates the marketplace, harms health, welfare and financial resources, and ultimately undermines trust in organizations and the digital marketplace as a whole.

For example, a fake video of a CEO admitting the company has been charged with a large regulatory fine (or class-action lawsuit) could cause severe damage, with a crash in the stock value of the company being one of the first victims. These types of attacks have already begun to occur. According to The Wall Street Journal (Stupp, 2019), in one high-profile case, cybercriminals used "deepfake phishing" to deceive the CEO of a UK energy company into transferring them $243,000. Using AI-based voice spoofing software, the criminals successfully impersonated the head of the firm's parent company, deceiving the CEO into believing he was speaking with his boss. The cybersecurity organization Symantec has stated that it had encountered at least three examples of deepfake-based fraud in 2019, which resulted in millions of dollars being lost (Zakrzewski, 2019). Moreover, consumers can also be

subjected to blackmail, intimidation, sabotage, harassment, defamation, revenge porn, identity theft, and bullying (Chesney & Citron, 2019b; Cross, 2022; Europol, 2022; Fido et al., 2022; Karasavva & Noorbhai, 2021; Whittaker et al., 2020).

Yet at the same time, this emerging technology carries many positive potentials through different forms of commercialization (Johnson & Diakopoulos, 2021; Maksutov et al., 2020), which may even help to change or innovate business models (Kietzmann et al., 2020). The opportunities pertaining to deepfakes are becoming even more relevant as consumers start spending more time in virtual worlds, which will foreseeably attract more attention and investment from firms across the board. For example, Facebook has changed their name to Meta and declared they will be pursuing a virtual reality world called Metaverse, in whose development they announced to invest 10 billion dollars in the fiscal year of 2021 alone, and more in the course of the upcoming years[2]. This virtual world will largely be composed of deepfake objects. Thus, this latest technology will usher in new opportunities, as well as new dangers. Accordingly, this dualistic nature is why, in this article, we investigate the risks and opportunities of deepfakes which are virtually unexplored in the present business literature.

Another critical factor that makes deepfakes relevant by amplifying their impact is their dissemination via the internet and social media – both of which have become integral to people's personal and professional lives, allowing consumers to access easy-to-use platforms for real-time discussions, ideology expression, information dissemination, and the sharing of emotions and sentiments (Perse & Lambe, 2016). Consequently, the scale, volume and distribution speed of deepfakes, combined with the increasing pervasiveness of digital technologies in all areas of society, are going to have profound implications—both positive and negative—in the marketplace (Kietzmann et al., 2020; Westerlund, 2019).

---

[2] https://www.cbsnews.com/news/facebook-earnings-report-2021-q3-metaverse/

However, as deepfakes are an emergent technology and complex in nature (Chesney & Citron, 2019; Dwivedi et al., 2021; Kietzmann et al., 2020; Westerlund, 2019), the current understanding of their implications is scattered, sparse, and at a nascent stage (Botha & Pieterse, 2020; Chesney & Citron, 2019; Kietzmann et al., 2020). As extant literature only offers anecdotal and disparate indications related to the various possibilities of deepfakes for firms and consumers (Chesney & Citron, 2019; Vimalkumar et al., 2021; Wagner & Blewer, 2019), there is a lack of coherent understanding of marketplace deceptions through deepfakes, and also the specific opportunities that they present for both companies and consumers (Chesney & Citron, 2019; Kietzmann et al., 2020; Westerlund, 2019).

To date, marketplace deception has been primarily investigated from the consumer perspective, with a heavy emphasis on how it affects consumers (Taylor, 2021; Xie et al., 2020). The effects of deepfakes on businesses have received scant attention, despite the fact that researchers have noted that firms are not immune to their effects (Chadderton & Croft, 2006; Xie et al., 2020). Moreover, deepfakes have a legitimate side to create business opportunities (Johnson & Diakopoulos, 2021; Kietzmann et al., 2020; Malbon, 2013). Consequently, both consumers and firms must develop their understanding and avoidance capabilities of deepfake deception, to mitigate the harm that deepfakes can create, and enjoy the opportunities that they may offer (Boush et al., 2015; Taylor, 2021).

Against this background, *the purpose of this study is to generate a holistic understanding of deepfakes in relation to marketplace deception and the potential opportunities that they offer.* More specifically, we address the following research questions (RQs):

- **RQ 1:** How may deepfakes contribute towards marketplace deception?
- **RQ 2:** How may firms and consumers avoid the malicious effects of deepfakes?
- **RQ 3:** What opportunities do deepfakes offer firms and consumers?

Through the application of an integrative literature review (ILR) methodology (Toronto & Remington, 2020; Torraco, 2016), we analyze the previous literature to create a comprehensive understanding in relation to our purpose. In addition to business academia, we review literature from multiple research streams with footprints in deepfake research, including communications, computer science, information science, journalism and social sciences, in order to synthesize existing knowledge. Through the current study, we establish a foundational understanding of deepfakes in terms of marketplace deception, for firms and consumers. We also address the protection mechanisms from their harmful effects, and offer insights into the legitimate opportunities that are presented by this emerging technology.

2 CONCEPTUAL UNDERPINNINGS

2.1 Understanding Marketplace Deception

Marketplace deceptions are based on misperception, misprediction, non-perception, or non-prediction (Mechner, 2010; Taylor, 2021). Deception is a common feature of marketplace interactions between business entities, marketers, consumers, and any other party that may seek to gain benefit in an illegal or unethical manner (Boush et al., 2015). Such deceptions can include misrepresentations through numerical information or research results, distraction and information overload, display of false emotions in sales and service delivery situations, brand mimicry, and lying about product features and usage outcomes, to name but a few (Boush et al., 2015; Mechner, 2010; Xie et al., 2020).

The early academic literature in this area focused mainly on deceptions through advertising and marketing communications. As early as 1975, Gardner (p. 42) posited that "If an advertisement […] leaves the consumer with an impression(s) […] different from what would normally be expected if the consumer had reasonable knowledge, and that impression(s) […] is factually untrue or potentially misleading, then deception is said to exist". This argument

emphasizes how a marketer might take advantage of consumers by disseminating false information. Given that such communications are frequently developed and disseminated by professionals, it is reasonable to presume that the false information is created with the intent of profiting at the expense of consumers (Chadderton & Croft, 2006; Xie et al., 2020). Consequently, marketplace deceptions through advertising result in consumers' negative perceptions about advertising and marketing in general, as well as their skepticism of future advertising claims (Darke & Ritchie, 2007).

In the context of e-commerce, Xiao and Benbasat (2011) argue that product-related deceptive information practices can encompass the manipulation of information generation, information content, and information presentation. For example, an e-commerce platform can conceal potentially unfavorable information about a product, or present incorrect information about its contents on its packaging (Román, 2010; Xiao & Benbasat, 2011). Moreover, artificial product recommendation agents—software that mimics individual consumers' product interests or preferences—can manipulate recommendation systems to generate deceptive product recommendations (Román, 2010; Xiao & Benbasat, 2011).

Similarly, because buyers rely on product reviews when making online purchases, businesses can fabricate and distribute fake product reviews to sway buyers' selections (Malbon, 2013; Y. Zhao et al., 2013). Such forms of marketplace deception (also known as opinion spams) can be human-generated or computer-generated (Salminen et al., 2022). Human-generated fake reviews can be sponsored by firms through false online consumer identities (Malbon, 2013; Salminen, Kandpal, et al., 2022). Computer-generated fake reviews use text-generation algorithms to automate fake review creation (Salminen et al., 2022). But irrespective of the mechanisms by which the deceptions are created and distributed, the focus is to deceive consumers, and in some cases, competitors to obtain monetary or economic gain (Luca & Zervas, 2016).

2.2 Marketplace Deception Through Synthetic Media

The use of synthetic media in marketplace deception differs from traditional deception in several ways (Giansiracusa, 2021; Karnouskos, 2020; Mechner, 2010; Mirsky & Lee, 2021; Van Huynh et al., 2021). Synthetic media is an umbrella term that refers to the artificial creation or modification of media by "machines" – more specifically, programs that use AI and ML (CB Information Services, 2021; synthesia, 2020; Taylor, 2021). Today, synthetic media include music composed by AI, text generation, imagery and video generation, and voice synthesis, among other means (CB Information Services, 2021; Karnouskos, 2020). Among these various forms, deepfakes refers to using deep learning to generate artificial content (Chesney & Citron, 2019; Zotov et al., 2020). The term deepfake was coined in late 2017 as a portmanteau of the terms "deep learning" and "fake".

Generally, traditional forms of deception such as deceptive advertising, entail concealing some information and/or presenting false information as true (Ott et al., 2013; Taylor, 2021). The more recent technology-based forms such as opinion spam and fake reviews are generally textual in nature, or may include out-of-context but genuine photographs (Lappas, 2012; Malbon, 2013; Ott et al., 2013). They are also context- and purpose-specific (Lappas, 2012). However, the introduction of synthetic media takes marketplace deception to a whole new level due to its versatile nature, and its higher appeal to human cognitive functions (Taylor, 2021; Wagner & Blewer, 2019). These media are also much more life-like and more appealing, with broad applications in a variety of contexts, all of which make protection from them significantly more difficult (Maksutov et al., 2020).

The presence of visible or nonverbal clues (e.g., facial expression, eye contact) for evaluating a piece of information has become diminished or even nonexistent as a result of recent technological breakthroughs (Maksutov et al., 2020; Ramadhani & Munir, 2020; Tong et al., 2020), bringing the degree of marketplace deception to a whole new level (Ho et al.,

2016; Taylor, 2021). Moreover, as computer-mediated deception has previously been applied to language-action cues such as verbal and nonverbal immediacy, in addition to the superfluous use of words, structured messages or argument development, and has adapted or mimicked interactional exchanges between messages, it has become increasingly challenging to evaluate the truthfulness of the incoming information (Ho et al., 2016; Ludwig et al., 2016). Thus, the recent introduction of deepfakes makes marketplace deception even more damaging, as hyper-realistic videos and other multimedia deepfakes are extremely difficult to differentiate from reality (Boush et al., 2015; Giansiracusa, 2021; Tahir et al., 2021; Zhao et al., 2020).

3 METHODOLOGY

The ILR approach that we have applied in this study is *"a form of research that reviews, critiques, and synthesizes representative literature on a topic in an integrated way such that new frameworks and perspectives on the topic are generated"* (Torraco, 2005, p. 356*).* It is considered as a particular form of a systematic literature review (SLR), as it follows a "systematic" process of sampling the literature (Toronto & Remington, 2020). However, the SLR approach tends to narrowly focus on a specific topic or type of study (Booth et al., 2016). In contrast, the aim of ILR is to be phenomenologically inclusive, placing less emphasis on the type of study, venue, and discipline (Toronto & Remington, 2020; Torraco, 2016).

Our adoption of the ILR approach is influenced by the inadequacy of existing research on deepfakes in the business and marketing domains. As relevant research in other fields such as computer science and political science is relatively more developed compared to the business domain, it is worth pursuing knowledge that has been generated in those fields, while at the same time analyzing any ramifications it may have in a marketing context. As such, the ILR approach allowed us to integrate primary knowledge from various research streams, so as to generate coherent and insightful answers to our research questions (Toronto & Remington,

2020; Torraco, 2016). As described by Tranfield et al. (2003), and following their adaptation by Sivarajah et al. (2017), we applied a three-phase approach, as illustrated in Figure 2:

Phase I – Planning the Review Process: Identifying the critical phenomenon of deepfake, and defining the research aim and scope.

Phase II – Conducting the Review Process: Identifying studies to review, developing an analytical framework, coding and synthesizing the relevant information, and developing the presentation framework.

Phase III – Reporting and Dissemination of the Research Results: Descriptive reporting of results according to the research questions, discussing the findings further, drawing implications from the study, and identifying future research avenues. (Sivarajah et al., 2017).

"Phase I" of the research—identification of the critical phenomenon of deepfakes, and defining the research aim and scope—has already been presented in the introduction section of this article. Next, we offer a description of "Phase II" in detail. "Phase III"—reporting and dissemination the overall results—is presented in Sections 4 and 5.

[Please Insert Figure 2 About Here]

Figure 2: Visual illustration of the process of this study

3.1 Data Collection and Screening (Phase II)

To identify relevant literature, we used three academic databases – Web of Science (WoS), ACM Digital Library, and IEEE Xplore. As a generic database, WoS is the most comprehensive, containing over 12,000 high-impact journals and scientific articles from over 3,300 publishers. The ACM Digital Library and IEEE Xplore databases are specialized databases focusing on technical disciplines. When combined, these three databases offer a large and balanced coverage of the existing literature on deepfakes.

We conducted detailed searches in each of the three databases. Given the nascent stage of deepfake research, we did not want to pre-limit the searches with highly specific keywords

that could result in the omission of important papers. Rather, to identify a wide range of publications that could illuminate deepfake and its implications, we used only the keywords "deepfake*" and "deep fake*" (* denotes plural forms), and manually identified any associated papers. Through searching the three databases, we identified a total of 798 publications (WoS: 362; ACM Digital Library: 177, and IEEE Xplore: 259). For all publications, we recorded the following details of each article in our review database: the title, author(s), publication outlet, year of publication, and abstract.

In order to check whether publications fitted within the scope of our study, we read the title and abstract, and if necessary the introduction and conclusion of each publication (Mustak et al., 2016). First, we included any paper published in international scientific journals or in established conference proceedings, as they tend to present the most up-to-date and established knowledge across scientific disciplines (Mustak et al., 2016). We excluded other forms of publications such as opinion pieces. Second, we excluded papers that were present in multiple databases. For example, the paper titled "Deepfake Portraits in Augmented Reality for Museum Exhibits" by Nathan Wynn, Kyle Johnsen and Nick Gonzalez (2021) was present in both WoS and IEEE Explore. We also excluded papers which had their title/abstract/keywords indexed in English in the databases, but the actual publication was in another language than English.

Finally, from this pool of publications, we selected those that contributed to the aims of this study. Here, any publication that was useful in answering any of our three research questions were retained, and the rest discarded. In doing so, we preferred articles that included literature reviews or which presented conceptual frameworks (Torraco, 2016), as they tend to summarize previous research rather than focusing on a specific aspect of the phenomenon. This 'top-to-bottom' approach was chosen because of the interdisciplinary nature of the phenomenon, so allowing us to capture succinct summaries from multiple fields (Toronto & Remington, 2020; Torraco, 2016). In addition, we included empirical studies that clearly

articulated implications for either consumers (users) or firms (organizations). Our final list included 74 publications (WoS: 42; ACM Digital Library: 14; IEEE Xplore: 18). The details of these papers—including their source database, title, authors, publication outlet, publication year, addressed research questions, form(s) of deepfake addressed, and key findings—are available in the appendix (Table 3).

The 74 publications reviewed in the current study were published in 57 different outlets, indicating that the topic currently attracts the attention of a diverse range of publication outlets and is highly multidisciplinary. In our pool of reviewed literature, only the following outlets published more than one paper on deepfakes: *Convergence: The International Journal of Research into New Media Technologies* (4 papers); *Cyberpsychology, Behavior, and Social Networking* (4 papers); *Communications of the ACM* (3 papers); *IEEE Spectrum* (2 papers); and *IEEE Transactions on Technology and Society* (2 papers). As illustrated in Figure 3, the first paper was published in 2017, and there were none in 2018. But the number of publications has increased significantly since 2019, providing a clear indication of the topic's increasing significance in terms of research. Simultaneously, the dotted line in Figure 3 represents the Google popularity index value (which ranges from 0 to 100, as determined by Google Trends), indicating that both public and academic interest in deepfakes is growing rapidly.

[Please Insert Figure 3 About Here]

Figure 1: Research articles and the Google popularity trend of deepfakes

3.2 Analytical framework, Coding and Synthesizing (Phase II)

Next, we systematically analyzed each article individually. We operationalized and followed an analytical framework with specific questions that were created to address the goals of the current research in a coherent and holistic manner, as suggested in previous methodological literature (Toronto & Remington, 2020; Torraco, 2016). From the research questions, we

derived specific analytical questions (AQs) that were used to analyze the articles. Our analytical framework is presented in Table 1.

Table 1: Analytical framework of this study

[Please Insert Table 1 About Here]

When analyzing the articles, we marked any text related to our analytical questions using short and intuitive codes (Toronto & Remington, 2020). After coding, we categorized the codes and associated texts based on their commonalities in relation to the analytical questions. We then read and analyzed them thoroughly to elucidate appropriate answers. Once we generated answers for each of the AQs, we grouped them according to our RQs. We then read and analyzed the grouped answers again, in order to check whether they coherently addressed the RQs (Torraco, 2016). We then discussed the findings among the research team, critically examined any disagreements in terms of interpretations, corrected any anomalies, and produced a set of answers to the research questions on which all researchers were agreed (Toronto & Remington, 2020; Torraco, 2016).

4 FINDINGS

Based on our detailed analysis of the reviewed literature, we develop a conceptual framework to capture the phenomenon of deepfake in the context of marketplace deception, and also the opportunities that it offers (Figure 4). The framework provides an overview of the phenomena, and simultaneously facilitates an organized presentation of the findings. We conceptualize that this emergent and highly potent technology is dualistic in nature, which can pose radical threats and offer new opportunities simultaneously, both to companies and consumers. The dotted arrows representing threats in Figure 4 indicate that according to our findings, the application of existing protection strategies and mechanisms does not mitigate the harmful effects of deepfakes in a comprehensive manner, and offer only a partial protection. Thus, some harmful

effects can still reach to companies and consumers. The dotted arrows on the right suggest that the effects of deepfake—both positive and negative—do not necessarily remain only within the spheres of companies or consumers. Rather, they often carry spillover effects where effects on companies can also affect consumers, and *vice versa*.

Figure 4: Conceptual framework of this study

[Please Insert Figure 4 About Here]

In line with the conceptual framework and in response to our RQs, next we present the various possible marketplace deceptions associated with deepfakes. Then, we analyze existing knowledge regarding how firms and consumers can safeguard themselves against their malicious effects. Following that, we identify and report the potential opportunities presented by this emerging technology.

4.1 Marketplace Deception through Deepfakes

4.1.1 Threats of Deception for Firms

The existing literature on marketplace deception focuses primarily on consumers who are the victims of deceptive actions and behaviors (Boush et al., 2015; Ludwig et al., 2016). However, our study demonstrates that in comparison to traditional deceptions, the scope of threats posed by deepfakes is significantly greater , as businesses can be harmed in a variety of ways (Johnson & Diakopoulos, 2021; Kietzmann et al., 2020; Zakrzewski, 2019). These include derogatory activities such as defamation and sabotage, and also damage to a firm's image, reputation and trustworthiness (Botha & Pieterse, 2020; Schwartz, 2018; Westerlund, 2019). The proliferation of deepfakes is subjecting companies to derogatory activities such as defamation and sabotage, which can pose a strong threat to a company's reputation and brand image through marketplace deception, resulting in a loss of trust from customers and other stakeholder groups (Di Domenico & Visentin, 2020; Rubin, 2019). Firms can be viciously harmed (e.g., through reputation loss) by adversary-initiated deepfake propagation (Botha & Pieterse, 2020;

Giansiracusa, 2021; Zakrzewski, 2019). It is critical to note here, that as we illustrate in our conceptual framework, these harmful effects often spill over from companies to consumers, and *vice versa*.

An example of harm to a company's reputation and brand image is where a firm's senior executive or figurehead is seen to be making compromising or deeply controversial statements (Westerlund, 2019). The screenshot of the video that we presented at the beginning of this paper (Figure 1) is another example. In a film created by the artists Bill Posters and Daniel Howe —and in collaboration with the advertising business Canny—, Zuckerberg can be seen sitting at a desk and allegedly delivering a menacing speech on Facebook's power (Eadicicco, 2019): "Imagine this for a second: One man, with total control of billions of people's stolen data, all their secrets, their lives, their futures," Zuckerberg's likeness says. "I owe it all to Spectre. Spectre showed me that whoever controls the data controls the future." Considering the controversies surrounding Facebook over the last few years—for example the Cambridge Analytica scandal (Confessore, 2018)—a deepfake video like this carries the potential to cause severe damage to the firm's reputation and brand image.

Another example is a fake news report in which the CEO of Pepsi (Indra Nooyi) was deliberately misquoted as saying that Donald Trump supporters should "take their business elsewhere." This prompted boycott calls and a 3.75 percent decline in PepsiCo's stock price. Thus, misinformation can result in negative financial consequences and diminished brand perceptions (Johnson & Diakopoulos, 2021; Wagner & Blewer, 2019; Zakrzewski, 2019). Similarly, videos that purposefully inflate earnings estimates can depress stock prices or harm a company's reputation, putting stakeholders at risk. Additionally, there are opportunities for algorithmic extortion, where managers may pay a fee to avoid deepfakes being shared, and resulting in public suffering (Kietzmann et al., 2020).

Deepfake technology can also be deployed to cause damage to firms of different capacities and profiles. For example, competitors can use deepfakes to deceive a firm's customers, or to develop negative public opinions or confusion about a rival's products, brands, and services (Zannettou et al., 2019). Additionally, deepfakes can be used to harm a business by creating fake reviews of their products and services. For instance, in a virtual brand community (VBC), the emergence of false but highly realistic deepfake-based reviews (particularly negative reviews) can affect the interactions of individuals with other VBC members as they begin to lose trust in the group, weakening their interest in interacting with other members (Feng et al., 2018). Additionally, if a business develops deepfakes as a means of providing false information (or concealing information) to consumers, this may also increase levels of consumer distrust (Malbon, 2013; Wu et al., 2020).

Along with harming a firm's image, reputation and trustworthiness through various forms of marketplace deception, deepfake technology has the potential to harm business models by disrupting incumbent technologies in certain industries (e.g. entertainment), effectively rendering them obsolete (Kietzmann et al., 2020). However, the situation opposite also exists where such technologies may be used to enhance these industries, as we discuss in Section 4.3.1. For instance, the dubbing and re-voicing industry which previously translated films to ensure that words in another language matched the actor's original lip movements is endangered and at risk of becoming extinct, due to the advancing technological ability to change languages and lips (Giansiracusa, 2021; Johnson & Diakopoulos, 2021; Zakrzewski, 2019). Similarly, deepfake technologies pose a significant threat to biometric authentication technologies, potentially disrupting businesses that provide authentication services (Botha & Pieterse, 2020; Schwartz, 2018; Zotov et al., 2020).

4.1.2 Deepfake Threats for Consumers

Deception via deepfakes can have major negative consequences for consumers that extend beyond the boundaries of firm-customer interactions, as they can be used for a variety of malicious purposes (Whittaker et al., 2020). According to the first report by Europol (the European Union Agency for Law Enforcement Cooperation) on deepfakes, these threats include, but are not limited to, harassing or humiliating individuals online, perpetrating extortion and fraud, facilitating document fraud, falsifying online identities and fooling 'know your customer' mechanisms, non-consensual pornography, online child sexual exploitation, falsifying or manipulating electronic evidence for criminal justice investigations, disrupting financial markets, distributing disinformation and manipulating public opinion, supporting the narratives of extremist or terrorist groups, stoking social unrest, and political polarization (Europol, 2022, p. 10).

Consumers' vulnerability, their chances of being exploited by deepfakes, and their lack of protection are increased further due to humans' limited cognitive abilities and ideological prejudices (Sharma et al., 2019). For instance, a lack of media literacy or familiarity with modern digital technologies may predispose consumers to false or deceptive information (Köbis et al., 2021; Rubin, 2019), stressing a new form of digital divide where consumers that lack the cognitive skills to detect deepfakes are at a structural disadvantage to those that have such skills. In other words, less sophisticated consumers can more easily fall prey to deepfake deception. For instance, more than 70% people in the UK are unaware of deepfakes and their impact (Europol, 2022).

In a similar manner, consumers with an insufficient knowledge of digital technology could be exposed to deepfake technologies and further propagate digital misinformation (Nygren & Guath, 2019). For instance, the website "Random Face Generator (This Person Does Not Exist)" uses AI to artificially generate fake portraits – i.e., people who do not exist

in reality. Figure 5 shows a few examples of such portraits, but not everyone is able to guess that AI could generate such realistic but non-existent faces in less than a couple of seconds. The AI face generator is powered by *StyleGAN*, a neural network from NVidia developed in 2018. According to the website, "AI is so developed that 90% of fakes are not recognized by an ordinary person and 50% are not recognized by an experienced photographer" (*Random Face Generator*, 2022).

[Please Insert Figure 5 About Here]

Figure 5: AI generates fake portraits – none of the people exist in reality

Furthermore, existing research indicates that certain demographic groups are more susceptible to fake contents. According to Guess et al. (2019), Facebook users over the age of 65 shared nearly seven times as many articles from fake news domains as the youngest age cohort. Moreover, the literature suggests online misinformation is associated with the third-person effect (Jang & Kim, 2018). The central tenet of the third-person effect is that people tend to overestimate the influence of media (e.g., deepfakes) on other people's attitudes and behaviors, while underestimating the effect on their own behaviors (Jang & Kim, 2018; Schweisberger et al., 2014).

From a commercial standpoint, deepfake technology has the potential to increase uncertainty in the marketplace and mislead consumers, resulting in their mistrust of businesses and also psychological discomfort (Botha & Pieterse, 2020; Giansiracusa, 2021; Zakrzewski, 2019). This, in turn, can erode consumers' purchasing intentions and impair the accuracy of helpful technologies such as recommendation systems. Additionally, given the rapid development of deepfake technologies that can generate human-like narratives using natural language processing (NLP) such as *GPT-3* (a text-generation model), it is reasonable to expect that the integration of such technologies with deepfakes will only contribute to an increase in marketplace deception (Etienne, 2021; Kietzmann et al., 2020; Westerlund, 2019).

Kietzmann et al. (2020) argue that deepfakes make it more difficult for people to respond to personalized advertisements. For instance, weighing the perceived value of highly personalized advertisements against a perceived violation of personal privacy requires consumers to strike a balance between the personalization of incoming data from deepfakes and the extent to which they can compromise privacy, which can be highly challenging. Additionally, consumers who participate in a variety of virtual communities (e.g., brand communities) frequently share similar ideologies (Zannettou et al., 2019). Accordingly, deepfake technologies may be used to launch inherently disruptive campaigns against such virtual communities, where members of such communities would likely regard the message as truthful as a result of the perceived parallels between the message and their embraced ideology.

Marketplace deceptions through deepfakes can take other forms and shapes, beyond those of firm-customer transactions. For instance, such deceptions might have a detrimental effect on anyone looking for employment (Chesney & Citron, 2019). According to a recent report from Microsoft (Burt & Horvitz, 2020), more than 90% of employers use search results to make decisions about applicants. However, these results have a negative impact in over 77% of cases, as businesses frequently refuse to interview or recruit individuals due to inappropriate images discovered during these searches. The reasons for these findings are rather evident, and hiring candidates who are not stigmatized by perceived negative online reputations is considered to be less risky. In these instances, creating compromising photographs and videos of a person and making them publicly available on the internet will significantly diminish that person's employment prospects. This simultaneously will hurt employers, as they risk missing out on potential talent. But moving beyond employment, various intelligence agencies have expressed concern that by propagating political misinformation and meddling with election campaigns, deepfakes have implications for national security (Europol, 2022; Westerlund, 2019), affecting consumers' ability to stay informed about the true state of affairs.

4.2 Protection from Marketplace Deception through Deepfakes

The magnitude of the threat posed by deepfakes in terms of marketplace deception and malevolent intent necessitates the development and availability of protection mechanisms against them. Next, we offer our findings in this regard. Important to note is that even though we present the protection mechanisms for firms and consumers separately for the ease off presentation and reporting, they are not mutually exclusive (Chesney & Citron, 2019; Europol, 2022; Farish, 2020; Kirchengast, 2020). Thus, protecting firms from deepfakes often means that malicious effects do not spillover to their consumers, and *vice versa* (Vizoso et al., 2021).

4.2.1 Protection for Firms from Marketplace Deception through Deepfakes

Extant studies primarily assume that the application of legal means is the primary—and often sole—protection mechanism from traditional forms of marketplace deception (Chesney & Citron, 2019; Langa, 2021; Ray, 2021). However, our analysis clearly shows that it is extremely difficult to protect firms and also consumers from the malicious effects of deepfakes using legal means alone. Rather, in order to address the concerns posed by deepfakes, three distinct but interrelated sorts of protection mechanisms—market, circulation, technical, and their legal responses—are concurrently needed (Chesney & Citron, 2019; Langa, 2021; Ray, 2021).

For firms, market responses to protect themselves include the mechanisms and methods that they can develop and implement to educate consumers about their products, brands and services, helping them to identify firm-sponsored and credible sources of information (Rubin, 2019). Investments in corporate social responsibility initiatives for improving public media literacy will benefit both the brand, and ultimately the marketplace as a whole (Bulger & Davison, 2018). Such a strategy aims to develop consumer information, media literacy, critical thinking, and evaluation skills that can be applied to assess the credibility and facticity of incoming information or news (Bulger & Davison, 2018; De Paor & Heravi, 2020). Notley and Dezuanni (2019) lamented that designing information literacy interventions requires a broader

disciplinary approach than simply education, and that contributions from economics, social psychology, and legal studies are also required. Furthermore, in designing strategies for improving the deception awareness of consumers, firms can build awareness about opinion-reinforcing versus opinion-challenging information that consumers can use when evaluating contents online (Lee & Shin, 2021). Opinion-reinforcing information is that which confirms or validates existing beliefs or opinions, whereas opinion-challenging information goes against the existing beliefs or opinions of an individual or consumer (Lee & Shin, 2021).

In the market, firms can also take advantage of online brand communities to counter marketplace deception through deepfakes (Wang et al., 2019). These strategies include interacting with online communities that may generate deepfake content, thereby avoiding actions that can make firms vulnerable to deepfake attacks (Giansiracusa, 2021; Johnson & Diakopoulos, 2021; Taylor, 2021; Wagner & Blewer, 2019). In addition, resources can be gathered from user credibility networks, expert group domains, and user ratings, in order to verify and develop the credibility of information being circulated via online channels (Meel & Vishwakarma, 2020). Similarly, firms can devise strategies for managing consumer interactions and feedback to foster protective behaviors within the brand community in response to the reputational dangers posed by deepfakes (Di Domenico & Visentin, 2020). Thus, by collaborating with real-life influential figures and by using deepfake technology, firms can develop what are known as online good nodes (approved artificial accounts of a real person) that can disseminate accurate information to refute or counter deceptive information (Zannettou et al., 2019).

Limiting or strictly regulating the circulation of deepfakes can offer further protection from their potential negative impacts. An outright ban on posting them on social media platforms is also taking place. For instance, TikTok is working on prohibiting “synthetic or manipulated content that misleads users by distorting the truth of events in a way that could

cause harm", through updating their community guideline (TikTok, 2019). Reddit updated their policy around impersonation, and "does not allow content that impersonates individuals or entities in a misleading or deceptive manner" (Reddit, 2020). YouTube has an existing ban for manipulated media, for instance, "Video content that has been technically manipulated (beyond clips taken out of context) to fabricate events where there's a serious risk of egregious harm." (YouTube, 2022). However, many of these rules contain subjectively interpretable terms such as "may cause harm," "misleading or deceptive," and "serious risk", and as a result, they may have loopholes that can be exploited by unscrupulous actors.

Technical responses include limiting access to computing resources that are necessary to develop and produce deepfakes. As an example, Google has banned the training of deepfakes in Google Colaboratory, which is a product from Google Research – a hosted Jupyter notebook service that requires no configuration and provides free access to computational resources, including GPUs (Anderson, 2022). Further research and development (R&D) investments in deepfake detection technologies and their successful deployment are also critical (Pu et al., 2021; Zotov et al., 2020). In making such investments, companies can use algorithm-based, computational detection techniques such as support vector machines and deep learning for detecting and countering the content-, context-, and domain-dependent features of deepfakes (Maksutov et al., 2020; Zotov et al., 2020). For example, Microsoft has introduced the Microsoft Video Authenticator, which can analyze a still image or video to determine the likelihood that it has been intentionally altered. However, it must be noted that these technology-based protections against deepfake deceptions come with specific limitations, simply due to the fast pace of improvement in generating synthetic media (Johnson & Diakopoulos, 2021; Ramadhani & Munir, 2020). For instance, if a method is reliant on the detection of an abnormal reflection of light in the eyes of the synthetic person, the adversarial network-based deep learning algorithms will learn how to overcome any shortcoming very fast

(Ludwig et al., 2016; Zotov et al., 2020). In this machine-versus-machine scenario, the whole detection method then becomes obsolete (Maksutov et al., 2020; Ramadhani & Munir, 2020). Hence, it is highly dependent on whether the detection technology can continuously stay one step ahead of the deepfake generation technology.

As a further measure, firms can deploy professional fact-checking bodies or individuals for verifying and detecting fake news (or deepfakes) that might be propagated against their products, services and brands (Lee & Shin, 2021; Nieminen & Rapeli, 2019; Zannettou et al., 2019). For example, Facebook works with third-party fact-checkers to address content that is reported as inaccurate or misleading, and partner with more than 50 fact-finding organizations, researchers, experts, and policymakers to find potential solutions (Westerlund, 2019). In an effort to increase user responsibility, the company has also developed tools for users to flag fake content, and educates them on how to identify them. Similarly, Google has incorporated fact-checking into its search engine, and Google News to help minimize the spread of false information (Farber, 2017). Wikipedia is also developing a spin-off site (Wikitribune) that employs crowdsourcing to verify the authenticity of news sources (Hern, 2017). Businesses can benefit from adopting and adhering to similar developmental deepfake policies across online platforms. One crucial aspect here is equality, and while larger firms may be able to leverage legal resources to battle deepfakes, smaller ones likely cannot. Therefore, social media platforms need to enable built-in detection and reporting features that make the playing field even for all operators facing a risk of "deepfake hijacking" (e.g., using their brand or people as part of a deepfake production without consent).

In this study, when it comes to legal responses to deception via deepfakes, we found that the legal protection is rather limited in most countries around the world (Karasavva & Noorbhai, 2021; Langa, 2021; O'Donnell, 2021). In December 2019, the US passed its first federal legislation addressing deepfakes (Graham et al., 2021), and in addition, some US states

have enacted their own laws to address the issue. Deepfake victims have a private right of action in New York and California, and Virginia has amended its penal code to make sharing deepfakes with necessary intent and without consent a crime (Graham et al., 2021). Additionally, state laws in the US such as the Illinois Biometric Information Protection Act (Illinois General Assembly, 2008), the California Consumer Privacy Act (State of California Department of Justice, 2018), and the New York SHIELD Act (The New York State Senate, 2019) are designed to safeguard residents' personal information, and may offer protection against deepfakes to some extent. However, as Graham et al. (2021) point out, as deepfake content is fabricated and artificially manufactured, establishing a privacy breach can be highly difficult for victims of deepfake-based deceptions.

In the European Union, the AI regulatory framework proposed by the European Commission will play a key role for law enforcement (European Parliament, 2021). The framework approaches the regulation of AI and its use from a risk-based perspective. Deepfakes are explicitly covered in terms of "AI systems used to generate or manipulate image, audio, or video content", and must adhere to certain minimum requirements such as labeling content as deepfake to make it clear to users that they are dealing with manipulated footage. However, the framework is still at a proposal level and not yet operational. In other major economies such as the UK, no legal means are available that offer direct protection from deepfakes. However, companies and consumers may seek protection through legislation prohibiting fraud, as well as provisions against harassment, defamation, infringement of copyright, as well as data protection laws (Graham et al., 2021). The newly established Civil Code of China, Art. 1019, essentially prohibits the violation of image rights by means of information technology or otherwise (Wei, 2020), which may also offer some degree of protection against deepfakes (Graham et al., 2021).

Businesses are typically unable to dictate rules, regulations, and laws. In this current state of affairs, however, they may monitor and advocate for legislation that protects the rights of organizations targeted by harmful deepfake content. As KPMG (Anderson, 2020) argue: "establishing a governance framework that embraces disruptive technologies and encourages innovation while ensuring risks are identified and managed is essential to an organization's ability to survive and thrive in a digital world." Additionally, firms can collaborate with regulators to develop, implement, and communicate laws or guidelines governing the creation or dissemination of deepfake content (Rubin, 2019).

4.2.2 Protection for Consumers from Marketplace Deception through Deepfakes

Our analysis reveals that little research is available on how consumers may protect themselves from marketplace deception through deepfakes. A phase of "disintermediation" has characterized the deepfake realm. The diversity of sources involved in the distribution of deepfakes, their potential for confidentiality, a lack of information quality requirements, the ease with which material can be manipulated and modified, the lack of contextual clarification, and the absence of credibility assessment objectives (i.e., subject matter, medium, and source) substantially complicate the issue of protecting oneself against deepfakes (Hwang et al., 2021; Viviani & Pasi, 2017).

For consumers in their everyday lives, a rather generalized but crucial protection mechanism is to develop the capabilities necessary for analyzing and interpreting the legitimacy of online content (Bulger & Davison, 2018; De Paor & Heravi, 2020; Viviani & Pasi, 2017). A consideration of the reputation of the information source, the involvement of trustworthy intermediaries such as experts and/or opinion leaders, and also personal confidence based on first-hand experiences will further enhance their protection (Hwang et al., 2021; Viviani & Pasi, 2017; Westerlund, 2019; Whittaker et al., 2020). Additionally, developing or gathering knowledge about products, brands and services by customers will enhance their

possibility for identifying and avoiding misinformation (Lee & Shin, 2021). Here, enhancing analytical thinking capabilities is of the utmost importance for consumers when examining the credibility or facticity of incoming information.

Furthermore, consumers need to be aware of the risks at the core of deepfake technologies. For this to happen, consumers need to be vigilant in the virtual environments in which they constantly interact, and develop a basic understanding (or literacy) about the technology and the existing deepfakes. To this end, online tools are becoming available. For instance, Jevin West and Carl Bergstrom at the University of Washington have created a website called "Which Face Is Real" (https://www.whichfaceisreal.com). All of the images on the site are either computer-generated from thispersondoesnotexist.com using the StyleGAN software, or are actual photographs from the FFHQ dataset of Creative Commons and public domain images. Putting the real and fake photos side by side, the site helps people to learn to be more analytical of potentially false portraits.

At an individual level, a variety of social ties—defined as the diversity of offline groups and contexts represented in one's online social networks (Torres et al., 2018)—can help increase the awareness of fake content. Additionally, the study indicates that increasing consumer awareness of fake content such as deepfakes has a beneficial effect on verification behavior and network trust (Torres et al., 2018). Thus, combating social media's echo chamber effect through an active exposure to diverse perspectives and networks also represents a viable individual-level strategy for addressing the deepfake problem (Cinelli et al., 2021; Gillani et al., 2018). Consumers may even take an offensive coping strategy by refuting the claims portrayed in fake content by searching for and presenting contrary evidence, in an effort to protect other consumers (Roozenbeek et al., 2021).

4.3 Opportunities Offered by Deepfakes

The risks of marketplace deceptions through deepfakes are undeniable for both firms and consumers. However, in comparison to other forms of deception that are used solely for unethical and malicious purposes, the emergence of deepfake technology is unique in that it also brings forth various positive opportunities. Here, we analyze and present the benefits of deepfakes, both for businesses and consumers. As shown in our conceptual framework (Figure 4), similar to threats, the opportunities afforded by deepfake technologies may also carry spill-over effects as well. That means the benefits of the technologies for firms are also likely be advantageous for consumers, and *vice versa*.

4.3.1 Opportunities for Firms

For businesses, opportunities include new forms of marketing campaigns including virtual brand ambassadors, developing cost-effective and accessible learning environments and contents, designing and deploying AI-based solutions to detect and counter deepfakes, and ultimately, developing new offerings and business models supported by deepfakes (Farish, 2020; Johnson & Diakopoulos, 2021; Wagner & Blewer, 2019).

OPPORTUNITY 01: New opportunities for marketing campaigns. Firms can use deepfakes to design and execute appealing marketing campaigns at a low cost by replacing and/or augmenting the role of humans in marketing communications (Farish, 2020; Zakrzewski, 2019). With deepfakes, marketing campaigns do not necessarily need to incorporate real humans; rather, they can create artificial human-like models that can attract and engage many fans and followers (Dwivedi et al., 2021). Furthermore, deepfake can easily assist in the removal of language barriers, allowing for the creation of multilingual marketing campaigns by dubbing videos in different languages, and artificially matching lip movements and facial expressions accordingly (Johnson & Diakopoulos, 2021; Kietzmann et al., 2020). This enables company executives and celebrities to speak directly to individuals using tailored

messages, even addressing customers by name. Deepfakes can also be used to add audiovisual elements to user-generated contents, such as textual customer reviews and testimonials (Wagner & Blewer, 2019; Westerlund, 2019).

OPPORTUNITY 02: Virtual brand ambassadors. Another way for businesses to use deepfakes in marketing is to create virtual brand ambassadors. For example, the Instagram account @lilmiquela (shown in Figure 6) depicts Lil Miquela, a fictitious idol created using deepfake technology. Created by Brud, a Los Angeles-based startup specializing in robotics and AI (Hsu, 2019; Koh & Wells, 2018), Lil Miquela is an artificial social media marketer, and a virtual influencer embodying the appearance and personality traits of a human (Hsu, 2019). Despite not being real, with more than three million followers as of November 2021, Lil Miquela has become one of the top influencers on the platform (Blanton & Carbajal, 2019; Drenten & Brooks, 2020).

[Please Insert Figure 6 About Here]

Figure 6: Lil Miquela – an artificial social media marketer with more than 3 million followers on Instagram, created using deepfake technology (Source: Instagram account @lilmiquela)

Lil Miquela exemplifies how brands can develop their own virtual brand ambassadors for brand sponsorship, disseminating their desired message through a digital avatar. For the new generations of consumers that enjoy an immersion into social media and virtual reality, artificially created content may not be categorically less valuable than "real" content, especially if it satisfies their entertainment needs, or other experiential purposes. This is alluded to by the fact that virtual influencers can garner large audiences, and for example, another virtual influencer Lu do Magalu boasts more than 14 million followers on Facebook, close to 6 million followers on Instagram, more than 2.5 million YouTube subscribers, and more than 1 million followers on TikTok and Twitter, respectively. We did not find any academic studies on the effectiveness of virtual influencers for brands, which therefore remains as a future research

topic. However, the fact that several virtual influencers have millions of followers suggests that deepfake technology can create artificial characters that consumers find interesting enough to follow.

OPPORTUNITY 03: Developing cost-effective and accessible learning environment and content offerings. According to the literature, deepfake technology provides a variety of opportunities to firms that create educational content, including the ability to provide learners with knowledge in more convincing ways than traditional approaches (Westerlund, 2019; Whittaker et al., 2020). This technology enables relatively inexpensive and easily accessible video production that creates new films or shows, or adapts old ones to convey various pedagogical perspectives (Chesney & Citron, 2019). Also, celebrity voices can be used to narrate books, memoirs can be read by the author, and historical figures can recount their stories in their own voices using AI voice cloning software (Martin, 2020). As a result, the listener has an immersive, high-quality listening experience. Moreover, because increasing information literacy has been considered as a means to mitigate the negative consequences of misinformation, the technology itself can be used for education and interventions that are specifically designed to address the challenges posed by deepfakes (Hollis, 2019; Notley & Dezuanni, 2019).

OPPORTUNITY 04: Designing and deploying AI-based solutions to detect and counter deepfakes. Addressing the surge of algorithm-generated misinformation opens up a new field of business for developing AI-based solutions and services that detect synthetic content from human-generated content, and provide consumers with warnings when confronted with marketplace deception or suspicious content (Maksutov et al., 2020; Torres et al., 2018; Zotov et al., 2020). Consequently, this opens the possibility to create and sell services designed to protect companies and consumers from deepfake deception (Chesney & Citron, 2019). Such technologies also have the potential to expand on a number of services that have

emerged in recent years as a result of consumer concerns about identity theft (Liere-Netheler et al., 2019).

OPPORTUNITY 05: Developing new offerings and business models supported by deepfakes. The literature highlights the possibility that the application of deepfakes may enable firms to develop new offerings, or even entirely new business models (Dwivedi et al., 2021). The technology can act as a valuable personalization tool for products, brands, and services (Dwivedi et al., 2021; Farish, 2020; Wagner & Blewer, 2019). For example, news organizations are currently examining ways to improve their efficiency and engagement through the use of video synthesis and other synthetic media technologies (Reuters, 2020). As an example, the South Korean television channel MBN presented viewers with a deepfake of its own news anchor Kim Joo-Ha, a snapshot[3] of which is seen in Figure 7. The broadcaster told viewers ahead of time that the newsreader would be fake, and Kim Joo-Ha is still employed. The firm behind the deepfake, DeepBrain AI, has stated that it is searching for media customers in China and the United States, and MBN has stated that it will continue to use the deepfake for breaking news reports (Foley, 2022). Extending this concept, certain aspects of visual illustration such as animated cartoons, comic books and political cartoons can be streamlined or even completely automated using image synthesis tools (Grubb, 2017). Further, as the automation process eliminates the need for teams of designers, artists, and others involved in the entertainment production process, product costs are mitigated, enabling individuals to produce content that is indistinguishable from that of the highest budget productions, for little more than the cost of operating their computer (synthesia, 2020).

[Please Insert Figure 7 About Here]

Figure 7: Deepfake of news anchor Kim Joo-Ha of the Korean television channel MBN

[3] https://www.youtube.com/watch?v=IZg4YL2yaM0

Innovative applications can open new doors in the fields of augmented and virtual reality, as well as enable value creation in cyber-physical systems. The technology can be used to create "digital humans" – artificial, lifelike personas that are both interactive and communicative. This possibility has been utilized in the concept of profound resurrection, and has already been demonstrated in the tourism sector with tourist sites such as the Salvador Dalí Museum in St. Petersburg, Florida having adopted advanced technology to bring the late Spanish surrealist (who passed away in 1989) back to life[4]. After visitors click a button adjacent to a life-sized screen, the deepfake-based avatar leaves his easel and approaches them, offering information about his artwork and the museum. Dalí reintroduces himself to tourists as they exit the museum, inquiring if they would like a selfie with him (Mihailova, 2021; Whittaker et al., 2020). As another example, "Digital Einstein"[5] embodies the personality of the actual scientist, and can answer daily quizzes about his life and work, answering scientific questions using the WolframAlpha knowledge engine. Thus, deepfakes can revolutionize customer experience with artificial human personages, for instance, in the form of a digital customer assistant, sales concierge, financial advisor, or healthcare coach (Digital Humans, 2021).

4.3.2 Opportunities for Consumers

Similar to firms, deepfake technology has been indicated to offer various opportunities for consumers. In this study, we identified two specific opportunities – the enhancement of the digital customer experience, and social good and medical usage.

OPPORTUNITY 06: Enhancement of the digital customer experience. Deepfakes carry the potential to enhance the digital customer experience (Whittaker et al., 2020). Merging

[4] Behind the Scenes: Dalí Lives. Url: https://www.youtube.com/watch?v=BIDaxl4xqJ4
[5] https://einstein.digitalhumans.com/?_ga=2.133820942.1293455835.1637135902-101885267.1637135902

deepfakes with synthetic AI models brings forward a high degree of personalization for online consumer interactions, such as online clothes-shopping (Kietzmann et al., 2020; Zakrzewski, 2019). For instance, customers will be able to input their primary physical characteristics into an online clothing store, which will then be able to generate life-like avatars to use for online shopping (Whittaker et al., 2020).

Thus, deepfakes may be used to create highly tailored material that transforms people into models, allowing them to virtually try on an outfit before purchasing it. Furthermore, targeted fashion advertising can be created that differs according to time, weather, and audience (Westerlund, 2019). The Japanese AI firm "Datagrid" has developed an AI engine that helps achieve these purposes, and automatically generates virtual models for advertising and fashion. This technology is called systematic model generation, and can be used by fashion advertisers or a wide range of communicators in the virtual sphere. A possible advantage of this type of application is consumers perceiving artificial content as being catchy, entertaining, or even emotionally attaching, thus deriving experiential value from deepfakes (Choi & Lim, 2019).

OPPORTUNITY 07: Social good and medical usage. Deepfakes can also be deployed for social good. For instance, consumers will benefit from their use in removing the language barriers that frequently impede the delivery of cross-cultural content, and which would typically require subtitle reinforcement. The technology will also provide a voice to people who have lost their own voice due to medical conditions such as motor neuron disorders. For example, Project Revoice (https://www.projectrevoice.org) employs deep learning principles to create video deepfakes with customized synthetic voices, based on voice samples provided by vocally paralyzed people (Whittaker et al., 2020).

In another example, Amazon has released an experimental Alexa capability that allows the AI assistant to impersonate the voices of users' deceased relatives. This capability was shown at the company's annual MARS conference, with a video depicting a child asking Alexa to read

a bedtime story in the voice of his deceased grandmother (Vincent, 2022). “As you saw in this experience, instead of Alexa’s voice reading the book, it’s the kid’s grandma’s voice,” said Rohit Prasad, Amazon's lead scientist for Alexa AI. He began the video by stating that adding “human attributes” to AI systems was becoming increasingly vital “in these times of the ongoing pandemic, when so many of us have lost someone we love.” “While AI can’t eliminate that pain of loss, it can definitely make their memories last,” Prasad added (Vincent, 2022).

5 DISCUSSION

5.1 General Discussion

Deepfakes are highly realistic synthetic media generated by algorithms (Chesney & Citron, 2019; Maksutov et al., 2020), and then typically distributed as social media content. They carry the potential to create marketplace deceptions for both firms and consumers. But simultaneously, deepfakes also offer various opportunities for both entities (Chesney & Citron, 2019; Dwivedi et al., 2021; Kietzmann et al., 2020; Westerlund, 2019). The current knowledge on deepfake is scant and dispersed (Maksutov et al., 2020; Zotov et al., 2020). In this study, we reviewed and analyzed 74 papers related to deepfakes from the fields of business, communications, computer science, information science, journalism, and social sciences, in order to generate insights regarding their implications for firms and customers. We provide an objective assessment of the risks that deepfake-induced marketplace deceptions pose to firms and consumers, the protection strategies and mechanisms against harmful effects, as well as the opportunities that deepfake technology presents.

Deepfakes can spread exponentially in an era where a large swathe of customers increasingly uses social media as a source of information. In contrast to the “offline” world where individuals have historically minimized credibility uncertainty based on either the reputation of the knowledge source (e.g., experts and/or opinion leaders) or personal first-hand

experiences, making an evaluation in the digital domain is frequently more complex (Viviani & Pasi, 2017). The multiplicity of sources involved in the distribution of deceptive content, the absence of information quality requirements and evaluation, the ease of manipulating and altering information, the lack of contextual clarification, and the existence of several potential credibility evaluation objectives (i.e., content, source, and medium) make deepfakes a very real and potent threat (Viviani & Pasi, 2017). As artificial contents blend seamlessly with authentic content in the digital environment, the terms *reality* and *truth* may become less relevant in comparison to how we humans understand these concepts. Similar to the arguments presented by Xiao and Benbasat (2011), deepfakes can be used to deceive the marketplace by manipulating information content, information presentation, and information generation.

The problem is not only that deepfake technology is improving at a very fast pace (Johnson & Diakopoulos, 2021; Schwartz, 2018). Rather, it is that the social processes through which we collectively acquire knowledge and determine whether something is *genuine* or *deceptive* are under threat, and the very definition of reality is a critical concern (Hwang et al., 2021; Schwartz, 2018). This is a phenomenon where frequent exposure to false information causes people to lose faith in what they see and hear. In other words, the danger is not that necessarily that people will be deceived just in the marketplace, but that they will come to regard everything as deception, and lose faith in the marketplace (Kirchengast, 2020; Schwartz, 2018; Tong et al., 2020). While consumers may accept content that supports their worldviews (even if the content is fabricated), they may lose interest in facts and develop a postmodernist cynicism in which "what is pleasurable is genuine". These effects of the erosion of trust and the muddying of the borders between real and artificial have left marketers wary. According to recent polls, trust in major institutions and the media is eroding (Ognyanova et al., 2020), and this trend is likely to be exacerbated by the proliferation of deepfakes if appropriate controls are not put in place (European Parliament, 2021; Langa, 2021; Schwartz, 2018).

The rise of marketplace deception through deepfakes, if not successfully addressed, may lead to a further erosion of consumer trust in business in general, and marketing in particular (Di Domenico & Visentin, 2020; Kietzmann et al., 2020). Deception protection and preparedness are crucial for consumers, firms, and the overall marketplace, to the extent that it has been labeled as a "critical life skill" (Boush et al., 2015, p. 1). However, most marketing textbooks and articles on marketplace deception treat it as a legal topic of interest, primarily addressed to corporate attorneys, judges, juries, and government regulators (Boush et al., 2015a; Farish, 2020; Langa, 2021; O'Donnell, 2021; Ray, 2021). Furthermore, research from technical disciplines such as computer science or data science focuses on technology as the primary solution for deception protection (Ramadhani & Munir, 2020; Schwartz, 2018; Zhao et al., 2020; Zotov et al., 2020). However, our research shows that protecting against deepfake-based marketplace deception cannot be accomplished through solely legal or technical means, and it necessitates combining market, circulation, technical, and legal responses, as well as educating and improving individuals' abilities to distinguish *truth* from *deception*.

The marketplace is a critical context in which to study deception, particularly in the face of the emergence of new and potent technologies (Boush et al., 2015; Schwartz, 2018; Xie et al., 2020). Through this study, we contribute to the marketplace deception literature by extending the overall understanding concerning deepfakes (Boush et al., 2015; Darke & Ritchie, 2007). Indeed, the findings of this work may have broader implications for comprehending deception beyond the marketplace. For instance, the general public is continuously being exposed to news about politicians, celebrities and influencers engaging in misleading behavior, which is an issue that will only become more pronounced through the use of deepfakes (Chadderton & Croft, 2006; Xie et al., 2020). In this regard, our study deepens the existing understanding of various forms of deception, their effects, and the protection mechanisms involved, from the perspective of society as a whole.

Most previous studies on deepfakes have focused on a particular industry, product or service. While this approach has yielded valuable insights into several key domains of deepfakes, there is a clear need for research that zooms out, so as to examine the implications of deepfakes from a broader perspective (Dwivedi et al., 2021; Vimalkumar et al., 2021). Moreover, most studies have perceived deepfakes as a grave danger (e.g., Giansiracusa, 2021; Graham et al., 2021; Maksutov et al., 2020). This is understandable, as deepfakes can undeniably present a serious threat to firms and consumers. Furthermore, the technology may appear mystical and incomprehensible to the regular person with a non-technical background, eliciting responses of intimidation and fear (Giansiracusa, 2021; Graham et al., 2021; Wagner & Blewer, 2019). Nonetheless, we have aimed to highlight the dualistic nature of deepfakes (see Figure 4), as we investigate the potential opportunities presented by this emerging and critical technology. As a result, our research is among the first to generate and present a balanced understanding of the phenomenon, that takes into account the perspectives of both firms and consumers, and combines the perspectives of multiple stakeholder groups.

Concerning the novelty of deepfakes in relation to other forms of market deception, the technological advancements regarding their ease of creation and diffusion makes synthetic content more commonplace than previous market deception manifestations. As a result, firms and consumers are transitioning into a mixed reality, where components of real and fake merge and fuse. This change has been characterized as the *post-truth society*, and forms a more pervasive transformation than the previous environment of deception, in that despite presenting complex schemes and forms of deception, the previous environment was still technologically limited and not omnipresent in people's lives in the same way that deepfakes will be. Notably, deepfakes seem to be part of the transition to a higher degree of digitality in people's lives, which involves an increasing amount of time spent in virtual reality and augmented reality. This mélange of realities stresses the need for new skills both from firms and consumers, in

order to cope with object detection and veracity judgments – cognitive skills that were not required previously. Paradoxically, part of the deepfake appeal is also the entertainment aspect of deepfakes, to the point that people might, to some extent, enjoy the deception in that it has a certain sense of a magic that both amuses and surprises.

5.2 Managerial Implications

Our study carries several implications for firms and managers. Deepfake technologies make it easier for criminals to perpetrate marketplace deceptions while remaining undetected. This study offers a comprehensive picture for firms regarding the severity of such threats. Especially, deepfake-based deceptions can result in direct financial damage, and negative and predatory deepfake campaigns can destroy a company's reputation, brand image, and stakeholder trust.

Therefore, firms need to invest in developing resources and capabilities to protect themselves from marketplace deceptions carried out through deepfakes. This includes investing in technology that enhances a firm's deepfake detection and avoidance competencies. At the same time, they should invest in human resources to enhance their capabilities of successfully countering the potential malicious effects of deepfake technology. In addition, managers need to pay attention to any potential harm that their consumers may suffer, and take preventive measures to safeguard them.

However, we also suggest managers pay close attention to the various commercial opportunities that are presented by technology, and be prepared to capitalize on them. For companies, non-deceptive, value-adding applications of deepfake-based marketing contents and campaigns can be highly beneficial, and deepfake technologies can also provide advantages in advertising, brand personification, and customer services. Moreover, we suggest that in addition to videos, managers should be aware of and benefits from other formats of synthetic media in their businesses. To this end; based on the insights offered by CB

Information Services (2021), we show a range of applications of synthetic media for brands and retailers in Table 2.

Table 2: Applications of synthetic media for brands and retailers (Source: CB Information Services, 2021)

[Please Insert Table 2 About Here]

In light of the findings of this study, and considering the rapid evolution of the technology landscape, adopting a proactive rather than reactive strategy is strongly recommended. Importantly, in addition to developing new offerings, deepfakes carry the potential of disrupting entire business models, and many firms may suddenly find themselves taken by surprise if they do not take these aspects into account. While the application of deepfakes is currently focused on entertainment and humoristic jokes, the historical trajectory of technology development has shown that performance of a given technology tends to change from a joke to a usable one; such a trajectory may also take place for deepfakes. The entertainment vs. value ratio might therefore change going forward.

5.3 Limitations and Suggestions for Future Research

As with any research, our study has certain limitations. First, we only investigated scientific papers indexed in three specific databases (Web of Science, ACM Digital Library, and IEEE Xplore). Despite the depth and breadth that they offer in terms of literature coverage, we have inevitably missed some valuable knowledge available in other databases. Second, we only focused on papers published in English, thereby omitting knowledge that will likely be available in other languages. Accordingly, any future research that widens this coverage and includes of this type of literature will enhance our knowledge base further. Third, we chose the conceptual lens of marketplace deception to approach the deepfake phenomenon. However, there could be alternative conceptual and theoretical frameworks that can further help to increase our understanding of deepfakes, such as market orientation and innovation (Atuahene-

Gima, 1996), and ethical marketing (Chonko & Hunt, 1985). But as these fall outside the scope of the current paper, we leave these alternative perspectives for future research.

Considering that we are in the early stages of deepfake research, particularly in the business domain, a lot remains to be investigated. Here, we make some recommendations for further research in critical areas. Overall, academics, firms, and consumers may benefit from studies that examine the origins and antecedents of deepfake. Academic and managerial relevance will also accrue from research aimed at determining the factors that contribute to the visibility of deepfakes on online platforms – i.e., how content recommendation and newsfeed ranking processes interact with deepfake content.

As our review suggests, consumer skills and aptitudes differ in terms of the ability to detect fake content, as do their attitudes toward artificial contents in general. Future research should delve further into these distinctions, in order to gain a better understanding of consumers' nuanced actions and attitudes based on deepfakes, and to make more precise recommendations for consumer education. Similarly, it is self-evident that some ethical rules for deepfake-based marketing are necessary, but they are currently missing from the marketing literature. As a primer on this topic, we propose that the criteria for ethical deepfake use involves usage in a way that is *non-deceptive* (i.e., making it clear that the content is artificial and not real), *transparent* (i.e., identifying the source authority and data from which the content originates), *fair* (i.e., does not violate the rights of third parties, whether they are a firm, consumer, or group of consumers), and *accountable* (i.e., consumers should be able to opt-out of fake content if desired).

The motivations of actors creating deepfakes require further scrutiny, including distinguishing between benevolent and malicious actors. As generally with AI technologies, identifying and assessing the moral standings of the users of a deepfake technologies remains a vexing challenge, as these technologies can be used for multiple purposes.

The legal implications of the technology also deserve additional scrutiny. Presently, legal scholars have urged that legislation be amended to encompass libel, defamation, identity theft, and impersonating government officials (Langa, 2021; Ray, 2021; Westerlund, 2019). Here, the critical issue to address is whether and how regulations or enforcements will be normatively appealing and acceptable (Chesney & Citron, 2019a; Europol, 2022; Farish, 2020).

Finally, given the opportunities presented by deepfakes, additional research on how to harness the technology for constructive purposes is clearly necessary. These explorations would benefit from exploring different content modalities. Currently, the focus of deepfakes is on video content but there are other content modalities, such as voice, that have potential business value. For example, synthetic voice creation is already offered as a service by some deep-learning companies (e.g.., Overdub). For example, one can type a text that he or she wants to speak, and let the ML model trained on one's own voice do the speaking based on a written script. This leads to interesting implications of hybrid forms of communication, where the author uses a replica (or a deepfake persona) of themselves to communicate. This and other effects of deepfakes on business processes in areas like sales and customer service open a fruitful avenue for experimental research.

REFERENCES

Anderson, J. (2020). *Are you ready for the next big wave? - KPMG Global*. KPMG. https://home.kpmg/xx/en/home/insights/2018/01/are-you-ready-for-the-next-big-wave.html

Anderson, M. (2022, May 28). Google Has Banned the Training of Deepfakes in Colab. *Unite.AI*. https://www.unite.ai/google-has-banned-the-training-of-deepfakes-in-colab/

Atuahene-Gima, K. (1996). Market orientation and innovation. *Journal of Business Research*, *35*(2), 93–103.

Blanton, R., & Carbajal, D. (2019). Not a Girl, Not Yet a Woman: A Critical Case Study on Social Media, Deception, and Lil Miquela. In *Handbook of Research on Deception, Fake News, and Misinformation Online* (pp. 87–103). IGI Global.

Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic Approaches to a Successful Literature Review*. SAGE.

Botha, J., & Pieterse, H. (2020). Fake news and deepfakes: A dangerous threat for 21st century information security. *ICCWS 2020 15th International Conference on Cyber Warfare and Security*, 57.

Boush, D. M., Friestad, M., & Wright, P. (2015a). *Deception in the marketplace: The psychology of deceptive persuasion and consumer self-protection*. Routledge.

Boush, D. M., Friestad, M., & Wright, P. (2015b). *Deception in the marketplace: The psychology of deceptive persuasion and consumer self-protection*. Routledge.

Bulger, M., & Davison, P. (2018). The promises, challenges, and futures of media literacy. *Journal of Media Literacy Education*, *10*(1), 1–21.

Burt, T., & Horvitz, E. (2020, September 1). *New Steps to Combat Disinformation*. Microsoft On the Issues. https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/

Caldwell, M., Andrews, J. T. A., Tanay, T., & Griffin, L. D. (2020). AI-enabled future crime. *Crime Science*, *9*(1), 14. https://doi.org/10.1186/s40163-020-00123-8

CB Information Services. (2021, June 30). *Should Brands And Retailers Adopt Synthetic Media—The AI For Digital Content?* CB Insights Research. https://www.cbinsights.com/research/what-is-synthetic-media/

Chadderton, C., & Croft, R. (2006). Who is kidding whom? A study of complicity, seduction and deception in the marketplace. *Social Responsibility Journal*, *2*(2), 207–215. https://doi.org/10.1108/eb059274

Chesney, B., & Citron, D. (2019a). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, *107*, 1753.

Chesney, B., & Citron, D. (2019b). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, *107*, 1753–1819.

Chesney, B., & Citron, D. (2019c). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, *107*, 1753–1819.

Chonko, L. B., & Hunt, S. D. (1985). Ethics and marketing management: An empirical examination. *Journal of Business Research*, *13*(4), 339–359.

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9), 1–8.

Confessore, N. (2018, April 4). Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. *The New York Times*. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

Cross, C. (2022). Using artificial intelligence (AI) and deepfakes to deceive victims: The need to rethink current romance fraud prevention messaging. *Crime Prevention and Community Safety*, *24*(1), 30–41.

Darke, P. R., & Ritchie, R. J. (2007). The defensive consumer: Advertising deception, defensive processing, and distrust. *Journal of Marketing Research*, *44*(1), 114–127.

De Paor, S., & Heravi, B. (2020). Information literacy and fake news: How the field of librarianship can help combat the epidemic of fake news. *The Journal of Academic Librarianship*, *46*(5), 1–8. https://doi.org/10.1016/j.acalib.2020.102218

Di Domenico, G., & Visentin, M. (2020). Fake news or true lies? Reflections about problematic contents in marketing. *International Journal of Market Research*, *62*(4), 409–417.

Digital Humans. (2021). *Digital humans: Conversational AI solutions beyond just chatbots | UneeQ*. Digital Humans. https://digitalhumans.com/

Drenten, J., & Brooks, G. (2020). Celebrity 2.0: Lil Miquela and the rise of a virtual star system. *Feminist Media Studies*, *20*(8), 1319–1323. https://doi.org/10.1080/14680777.2020.1830927

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Eadicicco, L. (2019). *There's a fake video showing Mark Zuckerberg saying he's in control of "billions of people's stolen data," as Facebook grapples with doctored videos that spread misinformation*. Business Insider. https://www.businessinsider.com/deepfake-video-mark-zuckerberg-instagram-2019-6

Etienne, H. (2021). The future of online trust (and why Deepfake is advancing it). *AI and Ethics*, *1*(4), 553–562. https://doi.org/10.1007/s43681-021-00072-1

European Parliament. (2021). *Tackling deepfakes in European policy*. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2021)690039

Europol. (2022). *Europol (2022), Facing reality? Law enforcement and the challenge of deepfakes, an observatory report from the Europol Innovation Lab, Publications Office of the European Union, Luxembourg.* https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes

Farish, K. (2020). Do deepfakes pose a golden opportunity? Considering whether English law should adopt California's publicity right in the age of the deepfake. *Journal of Intellectual Property Law & Practice*, *15*(1), 40–48. https://doi.org/10.1093/jiplp/jpz139

Feng, N., Su, Z., Li, D., Zheng, C., & Li, M. (2018). Effects of review spam in a firm-initiated virtual brand community: Evidence from smartphone customers. *Information & Management*, *55*(8), 1061–1070.

Fido, D., Rao, J., & Harper, C. A. (2022). Celebrity status, sex, and variation in psychopathy predicts judgements of and proclivity to generate and distribute deepfake pornography. *Computers in Human Behavior*, *129*, 107141.

Foley, J. (2022). *14 deepfake examples that terrified and amused the internet*. Creative Bloq. https://www.creativebloq.com/features/deepfake-examples

Gardner, D. M. (1975). Deception in Advertising: A Conceptual Approach: Deception in advertising needs further definition and procedures for measurement—Gardner's conceptual approach offers suggestions for both. *Journal of Marketing*, *39*(1), 40–46.

Giansiracusa, N. (2021). Deepfake Deception. In *Deepfake Deception. In: How Algorithms Create and Prevent Fake News* (pp. 41–66). Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-7155-1_3

Gillani, N., Yuan, A., Saveski, M., Vosoughi, S., & Roy, D. (2018). Me, my echo chamber, and I: Introspection on social media polarization. *Proceedings of the 2018 World Wide Web Conference*, 823–831.

Graham, N., Hedges, R., Chiu, C., de La Chapelle, F., & van de Graaf, A. (2021, October 12). How can businesses protect themselves from deepfake attacks? *Business Going Digital*. https://www.businessgoing.digital/how-can-businesses-protect-themselves-from-deepfake-attacks/

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1).

Hern, A. (2017, April 24). *Wikipedia founder to fight fake news with new Wikitribune site*. The Guardian. http://www.theguardian.com/technology/2017/apr/25/wikipedia-founder-jimmy-wales-to-fight-fake-news-with-new-wikitribune-site

Ho, S. M., Hancock, J. T., Booth, C., & Liu, X. (2016). Computer-mediated deception: Strategies revealed by language-action cues in spontaneous communication. *Journal of Management Information Systems*, *33*(2), 393–420.

Hollis, H. (2019). Information literacy and critical thinking: Different concepts, shared conceptions. *Proceedings of the Conceptions of Library and Information Science 10th International Conference (CoLIS 2019)*. Conceptions of Library and Information Science 10th International Conference, Ljubljana, Slovenia.

Hsu, T. (2019, June 17). These influencers aren't flesh and blood, yet millions follow them. *The New York Times*. https://www.nytimes.com/2019/06/17/business/media/miquela-virtual-influencer.html

Hwang, Y., Ryu, J. Y., & Jeong, S.-H. (2021). Effects of disinformation using deepfake: The protective effect of media literacy education. *Cyberpsychology, Behavior, and Social Networking*, *24*(3), 188–193.

Illinois General Assembly. (2008). *740 ILCS 14/ Biometric Information Privacy Act.* https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57

Jang, S. M., & Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions. *Computers in Human Behavior*, *80*, 295–302. https://doi.org/10.1016/j.chb.2017.11.034

Johnson, D. G., & Diakopoulos, N. (2021). What to do about deepfakes. *Communications of the ACM*, *64*(3), 33–35.

Karasavva, V., & Noorbhai, A. (2021). The real threat of deepfake pornography: A review of Canadian policy. *Cyberpsychology, Behavior, and Social Networking*, *24*(3), 203–209.

Karnouskos, S. (2020). Artificial intelligence in digital media: The era of deepfakes. *IEEE Transactions on Technology and Society*, *1*(3), 138–147.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, *63*(2), 135–146.

Kirchengast, T. (2020). Deepfakes and image manipulation: Criminalisation and control. *Information & Communications Technology Law*, *29*(3), 308–323.

Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *Iscience*, *24*(11), 103364.

Koh, Y., & Wells, G. (2018). *The Making of a Computer-Generated Influencer*. The Wall Street Journal. https://www.wsj.com/articles/the-making-of-a-computer-generated-influencer-11544702401

Langa, J. (2021). Deepfakes, real consequences: Crafting legislation to combat threats posed by deepfakes. *Boston University Law Review*, *101*, 761.

Lappas, T. (2012). Fake reviews: The malicious perspective. *International Conference on Application of Natural Language to Information Systems*, 23–34.

Lee, E.-J., & Shin, S. Y. (2021). Mediated misinformation: Questions answered, more questions to ask. *American Behavioral Scientist*, *65*(2), 259–276. https://doi.org/10.1177/0002764219869403

Liere-Netheler, K., Gilhaus, L., Vogelsang, K., & Hoppe, U. (2019). A Literature Review on Application Areas of Social Media Analytics. *International Conference on Business Information Systems*, 38–49.

Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, *62*(12), 3412–3427.

Ludwig, S., Van Laer, T., De Ruyter, K., & Friedman, M. (2016). Untangling a web of lies: Exploring automated detection of deception in computer-mediated communication. *Journal of Management Information Systems*, *33*(2), 511–541.

Maksutov, A. A., Morozov, V. O., Lavrenov, A. A., & Smirnov, A. S. (2020). Methods of deepfake detection based on machine learning. *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, 408–411.

Malbon, J. (2013). Taking fake online consumer reviews seriously. *Journal of Consumer Policy*, *36*(2), 139–157.

Martin, K. (2020). What is Voice Cloning? *ID R&D*. https://www.idrnd.ai/what-is-voice-cloning/

Mechner, F. (2010). Anatomy of deception: A behavioral contingency analysis. *Behavioural Processes*, *84*(1), 516–520.

Meel, P., & Vishwakarma, D. K. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, *153*, 112986.

Mihailova, M. (2021). To dally with Dalí: Deepfake (Inter) faces in the art museum. *Convergence*, *27*(4), 882–898.

Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys (CSUR)*, *54*(1), 1–41.

Mustak, M., Jaakkola, E., Halinen, A., & Kaartemo, V. (2016). Customer participation management: Developing a comprehensive framework and a research agenda. *Journal of Service Management*, *27*(3), 250–275. https://doi.org/10.1108/JOSM-01-2015-0014

Nieminen, S., & Rapeli, L. (2019). Fighting misperceptions and doubting journalists' objectivity: A review of fact-checking literature. *Political Studies Review*, *17*(3), 296–309. https://doi.org/10.1177/1478929918786852

Notley, T., & Dezuanni, M. (2019). Advancing children's news media literacy: Learning from the practices and experiences of young Australians. *Media, Culture & Society*, *41*(5), 689–707. https://doi.org/10.1177/0163443718813470

Nygren, T., & Guath, M. (2019). Swedish teenagers' difficulties and abilities to determine digital news credibility. *Nordicom Review*, *40*(1), 23–42.

O'Donnell, N. (2021). Have we no decency? Section 230 and the liability of social media companies for deepfake videos. *University of Illinois Law Review*, 701.

Ognyanova, K., Lazer, D., Robertson, R. E., & Wilson, C. (2020). Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-024

Ott, M., Cardie, C., & Hancock, J. T. (2013). Negative deceptive opinion spam. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 497–501.

Perse, E. M., & Lambe, J. (2016). *Media effects and society*. Routledge.

Pu, J., Mangaokar, N., Kelly, L., Bhattacharya, P., Sundaram, K., Javed, M., Wang, B., & Viswanath, B. (2021). Deepfake videos in the wild: Analysis and detection. *Proceedings of the Web Conference 2021*, 981–992.

Ramadhani, K. N., & Munir, R. (2020). A Comparative Study of Deepfake Video Detection Method. *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*, 394–399.

Random Face Generator. (2022). ThisPersonDoesNotExist - Random AI Generated Photos of Fake Persons. https://this-person-does-not-exist.com/en

Ray, A. (2021). Disinformation, deepfakes and democracies: The need for legislative reform. *Iversity of New South Wales Law Journal*, *44*(3), 983–1013.

Román, S. (2010). Relational Consequences of Perceived Deception in Online Shopping: The Moderating Roles of Type of Product, Consumer's Attitude Toward the Internet and Consumer's Demographics. *Journal of Business Ethics*, *95*(3), 373–391. https://doi.org/10.1007/s10551-010-0365-9

Roozenbeek, J., Maertens, R., McClanahan, W., & van der Linden, S. (2021). Disentangling item and testing effects in inoculation research on online misinformation: Solomon revisited. *Educational and Psychological Measurement*, *81*(2), 340–362. https://doi.org/10.1177/0013164420940378

Rubin, V. L. (2019). Disinformation and misinformation triangle: A conceptual model for "fake news" epidemic, causal factors and interventions. *Journal of Documentation*, *75*(5), 1013–1034. https://doi.org/10.1108/JD-12-2018-0209

Salminen, J., Kandpal, C., Kamel, A. M., Jung, S., & Jansen, B. J. (2022). Creating and detecting fake reviews of online products. *Journal of Retailing and Consumer Services*, *64*, 102771.

Salminen, J., Mustak, M., Corporan, J., Jung, S., & Jansen, B. J. (2022). Detecting Pain Points from User-Generated Social Media Posts Using Machine Learning. *Journal of Interactive Marketing*, 10949968221095556. https://doi.org/10.1177/10949968221095556

Schwartz, O. (2018, November 12). You thought fake news was bad? Deep fakes are where truth goes to die. *The Guardian*. https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth

Schweisberger, V., Billinson, J., & Chock, T. M. (2014). Facebook, the third-person effect, and the differential impact hypothesis. *Journal of Computer-Mediated Communication*, *19*(3), 403–413.

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(3), 1–42.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286. https://doi.org/10.1016/j.jbusres.2016.08.001

State of California Department of Justice. (2018, October 15). *California Consumer Privacy Act (CCPA)*. State of California - Department of Justice - Office of the Attorney General. https://oag.ca.gov/privacy/ccpa

Stupp, C. (2019, August 30). Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. *Wall Street Journal*. https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

synthesia. (2020). *The Future of (Synthetic) Media*. https://www.synthesia.io/post/the-future-of-synthetic-media

Tahir, R., Batool, B., Jamshed, H., Jameel, M., Anwar, M., Ahmed, F., Zaffar, M. A., & Zaffar, M. F. (2021). Seeing is believing: Exploring perceptual differences in deepfake videos. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16.

Taylor, B. C. (2021). Defending the state from digital Deceit: The reflexive securitization of deepfake. *Critical Studies in Media Communication*, *38*(1), 1–17. https://doi.org/10.1080/15295036.2020.1833058

The New York State Senate. (2019, June 14). *NY State Senate Bill S5575B*. NY State Senate. https://www.nysenate.gov/legislation/bills/2019/s5575/amendment/b

Tong, X., Wang, L., Pan, X., & gya Wang, J. (2020). An Overview of Deepfake: The Sword of Damocles in AI. *2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, 265–273.

Toronto, C. E., & Remington, R. (2020). *A step-by-step guide to conducting an integrative review*. Springer.

Toronto, C., & Remington, R. (2020). *Step-by-Step Guide to Conducting an Integrative review*. Springer. https://doi.org/10.1007/978-3-030-37504-1

Torraco, R. J. (2016a). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review*, *15*(4), 404–428.

Torraco, R. J. (2016b). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review*, *15*(4), 404–428. https://doi.org/10.1177/1534484316671606

Torres, R., Gerhart, N., & Negahban, A. (2018). Epistemology in the era of fake news: An exploration of information verification behaviors among social networking site users. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *49*(3), 78–97.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*(3), 207–222.

Van Huynh, N., Hoang, D. T., Nguyen, D. N., & Dutkiewicz, E. (2021). DeepFake: Deep Dueling-based Deception Strategy to Defeat Reactive Jammers. *IEEE Transactions on Wireless Communications*.

Vimalkumar, M., Sharma, S. K., Singh, J. B., & Dwivedi, Y. K. (2021). 'Okay google, what about my privacy?': User's privacy perceptions and acceptance of voice based digital assistants. *Computers in Human Behavior*, *120*, 106763. https://doi.org/10.1016/j.chb.2021.106763

Vincent, J. (2022). *Amazon shows off Alexa feature that mimics the voices of your dead relatives*. The Verge. https://www.theverge.com/2022/6/23/23179748/amazon-alexa-feature-mimic-voice-dead-relative-ai

Viviani, M., & Pasi, G. (2017). Credibility in social media: Opinions, news, and health information—a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(5), e1209.

Vizoso, Á., Vaz-Álvarez, M., & López-García, X. (2021). Fighting deepfakes: Media and internet giants' converging and diverging strategies against Hi-Tech misinformation. *Media and Communication*, *9*(1), 291–300.

Wagner, T. L., & Blewer, A. (2019). "The word real is no longer real": Deepfakes, gender, and the challenges of ai-altered video. *Open Information Science*, *3*(1), 32–46.

Wang, Y., McKee, M., Torbica, A., & Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. *Social Science & Medicine*, *240*. https://doi.org/10.1016/j.socscimed.2019.112552

Wei, C. (2020, May 21). *2020 NPC Session: A Guide to China's Civil Code (Updated)*. NPC Observer. https://npcobserver.com/2020/05/21/2020-npc-session-a-guide-to-chinas-civil-code/

Westerlund, M. (2019). The emergence of Deepfake technology: A review. *Technology Innovation Management Review*, *9*(11), 40–53. https://doi.org/10.22215/timreview/1282

Whittaker, L., Kietzmann, T. C., Kietzmann, J., & Dabirian, A. (2020). "All Around Me Are Synthetic Faces": The Mad World of AI-Generated Media. *IT Professional*, *22*(5), 90–99.

Wu, Y., Ngai, E. W., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems*, 113280.

Xiao, B., & Benbasat, I. (2011). Product-related deception in e-commerce: A theoretical perspective. *Mis Quarterly*, 169–195.

Xie, G.-X., Chang, H., & Rank-Christman, T. (2020). Contesting Dishonesty: When and Why Perspective-Taking Decreases Ethical Tolerance of Marketplace Deception. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-020-04582-6

Zakrzewski, C. (2019). Analysis | The Technology 202: Businesses should be watching out for deepfakes too, experts warn. *Washington Post*. https://www.washingtonpost.com/news/powerpost/paloma/the-technology-202/2019/12/13/the-technology-202-businesses-should-be-watching-out-for-deepfakes-too-experts-warn/5df279f1602ff125ce5b2fe7/

Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019). The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *Journal of Data and Information Quality*, *11*(3), 1–37.

Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2013). Modeling consumer learning from online product reviews. *Marketing Science*, *32*(1), 153–169.

Zhao, Z., Wang, P., & Lu, W. (2020). Detecting deepfake video by learning two-level features with two-stream convolutional neural network. *Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence*, 291–297.

Zotov, S., Dremliuga, R., Borshevnikov, A., & Krivosheeva, K. (2020). DeepFake Detection Algorithms: A Meta-Analysis. *2020 2nd Symposium on Signal Processing Systems*, 43–48.