Artificial Intelligence in Digital Media: The Era of Deepfakes

Stamatis Karnouskos
SAP, Germany

***Abstract*—The recent practical advances realized by Artificial Intelligence, have also given rise to the phenomenon of deepfakes, which can be considered as a form of fake news. Deepfakes is the phenomenon of creation of realistic digital products, and a plethora of videos have emerged over the last two years in social media. Especially the low technical expertise and equipment required to create deepfakes, means that such content can be easily produced by anyone and distributed online. The societal implications are significant and far-reaching. This work investigates the deepfakes via multi-angled perspectives that include media and society, media production, media representations, media audiences, gender, law, and regulation, as well as politics. Some key implications of these viewpoints are identified and critically discussed. The results indicate that as a society, we are not ready to deal with the emergence of deepfakes at any level. That we have not witnessed any severe impacts so far is due to their early stage of development, that shows imperfections To address the issue, a combination of technology, education, training, and governance are urgently needed.**

***Index Terms*—Deepfakes, Artificial Intelligence, Digital Media, Society.**

I. Introduction

Digital media dominance characterizes our era, where digital information can easily be created, communicated, and read globally. While this has increased access to information, its plethora also means that it has become increasingly challenging for the citizens, to verify and trust such information. The recent practical advances in Artificial Intelligence (AI), have a profound impact on a variety of domains, including that of digital media overall, and with critical implications. AI is considered as a paradigm-change technology, due to its manifold practical applications that have been demonstrated in the last years, most of which are attributed to a sub-field of AI named "Deep Learning" [1].

Some of the most recent tangible applications demonstrated show interdisciplinary and widespread areas of applicability, e.g., self-driving cars, super-human performance in games, human-level spoken interaction, intelligent robots, creation of cures to diseases, creation of new works such as texts, paintings, films, etc. When the sophisticated capabilities of Deep Learning are applied, images, text, and voices can be created or altered in a highly realistic way [2]. As such, Deep Learning has also enabled the creation of fake texts, fake voices, fake videos, and fake photos, all of which may at first look appear strikingly genuine and realistic, while they are not. AI has demonstrated recently, the capability of creating realistic fake videos, as it can, e.g., take an existing video and superimpose a person's photo on the face of the main character, or alter the voice of someone to say or do things that do not adhere to reality and never were performed [2]–[4]. Such technological capabilities have the potential to reshape the digital media, while the societal implications could also be severe, as they undermine the public confidence on what is seen, heard, and eventually believed to be true.

Several popular deepfakes circulate the Internet and can be found on popular websites such as YouTube. One of the first and probably the most famous deepfake, is the 2018 video of Barack Obama [5] where ironically, he warns about the dangers of deepfakes, something that Obama never actually did. Some other deepfakes are celebrity porn videos, as well as a misrepresentation of politicians e.g., replacing the face of Angela Merkel's with Donald Trump's, or Donald Trump's with that of Mr. Bean. Popular deepfakes at the time of writing, include an alternative version of some parts of the 1976 iconic American psychological thriller film "Taxi Driver", where the main actor's face, i.e., Robert De Niro's, has been replaced with that of Al Pacino's [6], and real-time creation of avatars for popular teleconferencing systems from a photo [7].

Deepfakes have the capability not only to impose photos in a video but also to generate new content. This, for instance, means realistic human face photos, as shown in [8], as a result of AI studying other human photos [9]. Available software even lets users create in real-time deepfake avatars for Skype and Zoom teleconference tools, simply from a photo [7]. Similar technology (deep learning) can be used to create fake CVs, as shown in [10]. While both of these may still seem like an "innocent" playground, consider the scenario where the human resources department of a major company is overwhelmed with realistically looking CVs and photos of potential hiring candidates, or with fake users impersonating in real-time others in business teleconferences (where the majority of the world conducts business during the COVID-19 pandemic). Today, there is simply no sophisticated defense against such kind of deepfake attacks.

While the technology is complex, its complexity is hidden behind common easy-to-use tools [11] and services that are available to the general public. The tools to create deepfakes have low technical requirements, which usually imply conventional home PCs that are equipped with gaming graphics cards (which have fueled the deep learning advances). For instance, the Taxi Driver deepfake was easily created with low-cost hardware and publicly available software DeepFaceLab [12] that utilizes AI to replace faces in videos. Because of the low learning-curve, public access to the technology, deepfakes can be created easily even by home users and without the need

for deep technical expertise. This can lead to the creation of realistic fake content that may be hard to verify its authenticity, and which in combination with its distribution to social media, can put existing fake news actions on steroids. Deepfakes are not a "one day, this might be possible" technology, but they have already been used in practice. Fabrication of photos and in some cases video is not something new, but the easiness by which this can be achieved, coupled with the realistic results, is something that is new and exciting but also worrying.

As with any disruptive technology, mixed feelings accompany it, that focus on the strengths or the weaknesses of the technology, and often attempt to make predictions about the future and how it will impact individuals and society. However, such views, especially in the media, are often asymmetrically focusing on some aspects of the technology and its impacts, and as such, may guide the public perception towards supporting or rejecting its applications. Deepfake is also such technology, for which some of its results have been featured in mainstream media. Therefore, it is of interest to see how AI, its applications, and impacts are communicated via modern media and how these are perceived as a threat or an opportunity by the public.

Digital media have several key characteristics that influence their nature and the way they are utilized. Although they form a continuation of traditional media, their capabilities make the information copying and transmission very easy, and as such, the target audiences can be huge masses without limitations on location, time, or content size. The utilization of digital media as a communication medium over the Internet has enabled instant communication, global audiences, and interactivity, which have already been capitalized, e.g., in social revolutions such as the Arab spring [13]. Deepfakes are strongly connected with digital media, and especially social media, via which they reach a wide audience. Since the text, images, videos, sound, are the key elements of interaction and communication within the public sphere, its implications for affecting it, are significant. While the power and capabilities of AI are still to be investigated, and far from being regulated, such efforts are reported in media with diverging views, and their portrayed implications range from the extinction of the human race to the cornerstone to its survivability.

The intersection of digital media and AI is of interest since it binds the cornerstone of media communication in the modern era, with the latest technological developments in AI, which blurs the boundaries of reality and information dissemination. The result may be an amplification of digital media impacts on an unprecedented scale, which has the potential to bring new benefits to society but also be misused to guide public opinion. With deepfakes, "the ability to distort reality has taken an exponential leap forward" [14]. Therefore it is of interest to investigate this intersection from different perspectives.

II. Methodology

The aim of this work is to better understand the intersection of digital media and AI and its implications in modern society. To that extent, a variety of methodologies [15] can be utilized. Document analysis is selected, and it was verified that sufficient material and access to it was available. This is challenging, considering that the phenomenon of deepfakes is very new (approx. 2-3 years old), and as such, peer-reviewed literature that deals with it is limited. In the core literature explicitly dealing with deepfakes was considered, e.g., [1]–[4], [16]–[18], which was complemented with the analysis of several other documents including technical reports, press releases, legislation proposals, and audiovisual material available on freely accessible Internet platforms (e.g., YouTube). The aim is to synthesize findings, views, examples from different perspectives, as discussed in section III – section IX and in addition to critically reflect on the phenomenon overall as discussed in section X. For each perspective, additional literature is considered that reflects key aspects so relevant to that viewpoint's context.

The different angles selected to investigate this phenomenon, are inspired by intersectionality, an approach that can be used to investigate how categories are inter-/intra-connected and how they interact at different levels. Via intersectional analysis, we can attempt to identify and understand the impact in different domains of digital media and implicated social phenomena such as injustice, inequality, political influence, etc. To investigate a new and complex phenomenon such as deepfakes, multiple perspectives are needed in order to be able to capture its essence. Therefore, it was decided to start from the perspectives that digital media can offer, i.e., media and society, media production, media representation, and media audiences. However, since the societal aspects are in focus, additional perspectives were deemed necessary, i.e., gender, politics as well as law & regulation. While these may not be exhaustive, we consider that they do provide a sufficient basis for the discussions that shed some light on the phenomenon and enable us to discuss its implications.

It must be pointed out though, that although inspired by intersectionality, it is more freely followed rather than strictly implemented, and this assists towards the aim of providing a multi-angled view and analysis of an everyday phenomenon from different perspectives. The undertaking of identifying implications in a methodological manner is not trivial, as it is methodologically difficult to isolate and identify the effects of media across the social space [19].

III. Media and Society Perspective

Culture constantly changes [20], and its impact in society is intertwined with the way people interact and use modern technologies. Mediatization captures such aspects as it "tries to capture long-term interrelation processes between media change, on the one hand, and social and cultural change on the other" [21]. Overall, mediatization is a process but "can also be seen as a container in which observations can be collected" [22], and as "a dynamic process of increasing media influence, cannot be regarded as a deterministic and linear development" [23]. Mass media have evolved from newspapers and TV in the past, to digital networks and platforms that are perceived as a "shared space of increased visibility and connectivity" [24]. Overall, mass media have effects that lead to "change in an outcome within a person or social entity that is due to mass

Preprint version of doi:10.1109/TTS.2020.3001312

media influence following exposure to a mass media message or series of messages" [25]. As such, content communicated over any kind of mass media, for instance, deepfakes over social media, raises some concerns.

In line with mediatization, the attempts to envision and capture the effect of AI on the life of a community, especially from a social, political, or ethical point of view, have already been an issue in science fiction and novels. There, AI is often embodied in the existence of intelligent machines that autonomously manage trivial tasks and, therefore, free up time for people to pursue more spiritual and fun activities [26]. On the basis of this context, social processes and their regulation need to be seen in a new perspective. Today, regulation of deepfakes hardly exists (see also discussion in section VIII), while in social processes, they fall under the larger category of Internet memes, fun, or simply fake news. Fiction can be seen as a form of interrogation that can deliver different aspects than the traditional journalism or academic writing [27], and in the digital media era, this is well represented in literature, films, games, etc. However, we are now entering a new era, where this interactivity is becoming much more real, as AI systems play and win sophisticated games against humans (e.g., the board game Go, strategy game StarCraft II), while they are able to achieve super-human performance on specific tasks (e.g., image classification). This implies that soon enough, what is up to now discussed only in fiction, may step-over to the real-world realm, where it will have an impact on existing societal processes.

The first impacts of AI are already visible, as several processes in economy and business are enhanced via AI, and more advanced products can be offered as companies can manage profitability and risks more efficiently [28]. In modern digital media, large amounts of digital information can be collected and analyzed, while processes can be automatized and highly customized towards even individuals. For instance, many social media platforms and other service providers feature user-targeted marketing campaigns based on individual user history, actions, and preferences. Such customized messages selected by sophisticated AI algorithms also include political messages as well as views of other like-minded citizens in social media. The utilization of deepfakes in these processes significantly extends the effectiveness and outreach of such actions, as they may reinforce beliefs or provoke actions. While today social platforms engineer sociality by enabling and forging connections, and emerge as active mediators between users, technologies and content [29], in the future, empowered with AI, they will be even more well-integrated in such processes, eventually creating "individual perception bubbles" that could potentially be misused to manipulate individuals and eventually the public opinion. As such, politics and digital media are strongly coupled with AI overall, and the deepfakes could be an additional enabler towards manipulating the real-world.

Mediatization is "constituted in the mutually influencing and molding relationship between institutions and the actors that reproduce, maintain, and develop them through their agency" [30]. In this context, the impact of new technologies (such as AI) and society influence each other, as "technology also influences the way we think about the social and the political" [31]. In such assessments, the focus is often put on the individual, and as such, only a specific view of the social is evaluated, while for a more holistic viewpoint and understanding of the interplay, it needs to be extended to the community and society [31]. The latter larger context, where the interplay of technology and society evolves, still needs to be better addressed, and this holds true also for deepfakes in digital media. The direct effect of deepfakes is just the first layer that needs to be considered, while the real danger is the distrust on organizations, processes, and people that are induced. Looking at the AI from the media and societal perspective is interesting, especially considering that today many situations involving deepfakes are insufficiently addressed due to lack of expertise and difficulty in managing complex social, conflict, and real-world aspects [32]. AI can enable us to understand people and group views better, combine their knowledge in a timely fashion, simulate reactions, and find negotiation win-win for complex situations that are beneficial for all [32]. However, for that to be materialized, such actions need to be done over a healthy and truthful basis, something that deepfakes could potentially jeopardize.

Not everything is put in a positive light in this interplay between AI and society. Many times new technologies that are not well understood have been misused, as there is a lack of appropriate regulatory frameworks in place, e.g., for robotics [33], [34]. Especially in the press, as well as in social media, often articles have circulated that paint a dystopian future, where AI is controlling everything and enforcing or enabling questionable practices and ethics in humans. Deepfakes enable individuals and organizations to create content that could be utilized for nefarious purposes. The interplay of AI digital media and society is complex, and although some of the issues analyzed do provide some insights, it would be of interest to investigate how political-economic structures in industry forge or hinder individual agency around the utilization of AI applications.

IV. Media Production Perspective

One of the key areas affected by AI in digital media is that of media production, where a bidirectional power relationship between production and consumption exists. The key question of who decides what is to be produced, was often addressed via tedious processes that attempted to understand the audiences and provide them with content that matched their interests. In the digital media era however, such processes are automatized and provide instant insights at unparalleled detail. AI algorithms analyze in real-time massive amounts of data (big data) and derive detailed profiles on users, their interests, their needs, and their satisfaction [35]. This, coupled with recommendation engines, can automatize the profiling efforts [36] in media organizations and enable the generation of content specifically for selected target groups based on very fine and sometimes very personal details. In this manner, deepfakes further amplify this targeting, as the content can be shaped in a way that is more appealing to the specific characteristics of an individual, rather than that of a mass. For instance, a message could be produced in support of a

Preprint version of doi:10.1109/TTS.2020.3001312

political candidate, featuring a specific actor or leader that the user finds trustworthy and in a language that s/he is susceptible to believe. Nowadays, business organizations rely on social media for their decision-making processes, and when this is coupled with the learning abilities of AI, new media intelligence approaches can emerge that are social and multimodal [37].

News is a product, and how it is produced can be understood by looking at cultural, economic, political-economic, or other power relationships [38]. The interplay of media and journalism is very complex, with multiple angles, and as such, it may not be easy to understand its full extent. For this to happen, the full nature of journalism needs to be understood, i.e., its focus on making news in the digital era, while also considering the observed trends, e.g. (i) the use of the affordances of news websites, (ii) radical commercialization, (iii) participating audience and (iv) the multi-skilling and de-skilling of journalists [30]. To this end, the participation of the audience, as well as utilization of such journalism over social media platforms, its impacts, and implications are not well understood, as evidenced by the fake news and Facebook contemporary discussions pertaining to the 2016 US election [39]. Hence, there is a need to approach it at the system level, as "systems theory is analytically powerful in describing the changing social power of journalism and mass media" [40]. Especially since journalism is now highly dependent on the public attention [40], and in the era of social media, this is done in real-time, the power relationships between media platforms, social platforms, and other stakeholders need to be investigated. This also implies that journalists need to be able to verify the authenticity of claims, and as such, they need to have the necessary tools and potentially the expertise to spot deepfakes.

Another key aspect is the automated production of media that is fueled by AI it has demonstrated its capability to create text, pictures, and videos, based on what it learns from available sources in digital platforms, e.g., social media such as Facebook, Twitter, Instagram, etc. Deep Learning has enabled this significant progress in Natural Language Processing (NLP) advances, including the capability of summarizing texts [41], as well as the reproduction of voices (e.g., narrating) based on limited available samples (e.g., a 5-second speech sample) [42], via which deepfakes can be created, at a level that even humans find it convincing [43]. As such, in an era where digital information is easily accessible and can be copied, the creation of automated news (including spoken language) is possible. This has significant effects on media production, as complex content can now be created by AI, given adequate sources e.g., a publicly available voice sample from a speech, and a photo from social media, can lead to the creation of a realistic deepfake video [44].

With global audiences, the production of news needs to increase. Automatization in production is a step towards this direction, and is also associated with less human involvement. In a recent survey [45], some have a negative feeling that robots are likely to damage journalism's value, while others also see the positive sides. Such robot journalism raises several challenges, including questions of bias, ethical considerations, as there is still no clear understanding of the interworkings in many of the AI algorithms and the decisions taken by them [46]. Deepfake-driven content may easily find its way through an automatized process that has too few "controls" in place to check for veracity and ethical issues, that might (still) be detectable by humans. In addition, robot journalism has "significant practical, sociopolitical, psychological, legal and occupational implications for news organizations, journalists and their audiences" [47] as for instance, there are discrepancies between authorship and credits for the produced material. Although today there is a need for employees that have multimodal media production skills [48], if such tasks can be automatized and (even partially) delegated to AI-empowered processes, the need of such skills might not be necessary or will have to shift focus, e.g., towards collaboration with AI robots on content creation.

Journalism is nowadays not only produced by journalists but also individual citizens. Citizen-driven journalism is evident in digital media, and even at the mainstream media, where we increasingly see user-generated content (which may include deepfakes). This is often also presented to wider audiences, e.g., via TV or social channels of large organizations as such, which, however, relaxes the veracity checks that are applied and the responsibility is shifted to a large extent to those individuals, e.g., with a simple remark such as "citizen captured video or photo". However, democracy needs high-quality investigative journalism, and not only creative variations of text, or even worse, creative untruthful content, both of which can be easily created with deepfake technology. In addition, in the age of the market-driven operations, the production of news supported by deepfakes may incentivize unethical and profit-maximization actions that contradict institutional ethics of professionalism [49]. As such, the emergence of deepfakes has a significant impact on media as its production will certainly be misused according to the agenda of its creators.

V. Media Representations Perspective

The representation of AI overall and its capabilities in digital media is of interest as it shows how such technological advances and their impacts are communicated to the public. It is pointed out that "culture is central to shaping collective perceptions and the dynamics of media representation reproduces forms of symbolic power" [50]. AI has been in most cultures linked to tangible physical objects, mostly as robots or highly advanced computer systems that have a human-like logic and conversational capabilities. This is evidenced in media, including fiction, images, and films, eventually creating a robot culture over the last decades. However, with the modern AI systems, a physical embodiment is optional, while AI overall is evidenced in everyday life artifacts and processes, even when not directly recognizable, e.g., interaction with voice assistants in smartphones.

The perception of AI in the general public is that of (sometimes humanoid) robots that follow stereotypes, e.g., are either villains or saviors [51]. Such stereotypes pose a way of structuring our understanding of the world, our values, and experiences, and position the robots in a specific

Preprint version of doi:10.1109/TTS.2020.3001312

context; therefore, also our expectations and interactions with them. How media nowadays represents AI is often biased and unbalanced, and the same holds for deepfakes. General AI is often portrayed in the media as "a mixture of flawed entertainment and fear" [52]. For instance, deepfakes are shown as a result that a machine created, and which may raise a wow effect, but may also be coupled with some often unsubstantiated implications, e.g., what would happen when machines could imitate everyone or how such actions would mean the financial catastrophe of film and media industry. Such representations, e.g., in popular culture, may have an effect on public attitude and lead towards a specific view on deepfakes and the associated technologies. Such actions have been demonstrated in the past, for instance, while science fiction may "harden anti–killer robot attitudes among that portion of the population who consume a lot of science fiction" the same can not be claimed for the general public and its opposition to autonomous weapons [53]. The "uncanny valley" hypothesis [54], suggests that end-products that closely resemble human behavior (but not exactly) can show uncanny or strangely familiar feelings of eeriness. However, there is an indication that science fiction can reduce the eeriness of robots [55], something that, in the future, we might also witness for deepfake products.

Deepfakes can be generated with the use of AI, and while "robot" or "robotic processes" are used in popular media sometimes, they simply want to piggy-back on the general notion of AI and robotics, denoting non-human intelligence hosted somewhere. Due to the robotic culture developed over the decades, it is easy to associate AI overall and deepfakes to robots or machines that are intelligent enough to produce new content and fool humans. However, deepfakes merely rely on the application of sophisticated algorithms to content, in order to create high-quality video, sound, and images, but do not act autonomously and not intelligently, even when they outperform humans in specific tasks, e.g., object detection. Nevertheless, their representation as such in media is done to ease their introduction to the public and due to lack of real expertise in the area.

Deepfakes, similar to fake news, have also been portrayed in digital media, with social media being their prime channel of distribution, but also traditional news media. How these are represented there, however, differs. In social media, much of the content so far had entertainment character (e.g., replacing faces of celebrities), and it was easily identified as fake. In traditional media, it has been portrayed as an example of what technology can do, and often put in the context of larger social questions and dilemmas. Media representations play a role in how the phenomenon of deepfakes is perceived, but up to now, this has mostly relied on the fiction and public perception of fake news. However, with the rapid advances in the technology behind deepfakes, such issues will need to be revisited and be better understood.

VI. Media Audiences Perspective

Digital media have certainly changed the way people interact, and now with AI on the rise, it seems that such relationships are again going to be significantly affected. A mass media effect is a change in people or overall in society as a result of the exposure to mass media influence [25]. When considering the capabilities of deepfakes that combine (i) the creation of new content and (ii) the easy dissemination of it via digital media, it is evident that this will be a game-changer and have an interdisciplinary impact. For instance, people have different behaviors when interacting with AI, e.g., robots [56], and it has also been observed that "social surrogates have the potential to cause psychological harm" [57]. While deepfake similar studies are not available, existing ones on the large domain of fake news [58] also exemplify the potential of deepfakes to cause harm and empower existing fake news techniques.

Social media in the modern context are mass media that have the potential to interfere with societal actions, e.g., social movements [59]. Nowadays, more than ever, a single individual can have a significant impact, since if his/her message goes viral, it gets to be seen by millions of people, which affects them. Deepfakes, especially portraying "new facts" or controversial issues, have such potential to become viral, as people might be less reluctant to check their veracity, in light of the time or subject sensitivity of the issue. As an example, political messages in times of conflict among nations can spread uncontrollably and lead to irrational and emotional reactions.

Fandom may be another aspect that is affected. Today "media convergence, new technologies, and transmedia marketing have all created new types of fans" [60], and deepfakes can potentially be another enabler. By lowering the capability of creating media products, deepfakes have the potential to enable communities to emerge more easily, and the generated deepfake content could more easily attract fans, something that implies that industry may no longer solely control such spaces. This is an angle that can further act as an enabler in the scope of the "convergence culture", where mass media are seen as a form of it [61].

Since "uses and gratifications is a media-effects perspective", the exposure to a medium is typically captured via "measures of one or more types of audience activity such as selectivity, media and content preferences, level of attention, and involvement with content" [62]. Due to the AI technology advances, two things can be realized which are game-changing (i) access at mass to people, their individual perceptions/habits/views, etc. as these are captured by digital media, and (ii) instant and continuous evaluation of the available data, as well as correlation at global levels. As such, studies may be easier to carry out and may reveal new insights based on the more detailed data as well as their potential for longitudinal realization. The results can then flow from the audiences to the systems and approaches that create, manage, and operate the digital media services and content, therefore providing a better match between user needs and end-products. To this end, deepfakes may be utilized as enablers and multipliers of the efforts done to increase the gratification aspects and engage more with the audiences.

Preprint version of doi:10.1109/TTS.2020.3001312

VII. Gender Perspective

Deepfakes have a gendered angle. Feminism can be seen as ”an emancipatory, transformational movement aimed at undoing domination and oppression” [63]. In the modern era, different forms of feminism have emerged that utilize modern media [24]. When looking collectively towards media effects and feminism, the existence of ”feminist philosophies, concepts, and logics articulating feminist principles and concepts to media processes such as hiring, production, and distribution; to patterns of representation in news and entertainment across platforms; and to reception” can be evidenced [63].

The question that is raised is what may be the interplay of deepfakes and feminism. Since contemporary feminism, also known as ”hashtag feminism”, ”takes place online and, at times, exclusively through social media platforms” [59], it is heavily susceptible to the deepfakes. It should not be forgotten that the initial deepfake material that appeared portrayed predominantly fake female celebrity pornographic videos and revenge porn on females [2], [4]. Although it was argued often in media that this was created for ”entertainment” and was addressing mainly the male audience, they were also made available to the wider public in well-known pornographic sites, effectively attacking the identity and moral stands of those targeted. While such content is illegal, and various websites make active efforts to remove it once detected, often such actions come too late or are not efficient.

Of particular importance to the gendered angle of deepfakes, is the revenge porn, which is an evolution of the existing non-consensual image sharing (e.g., nude photos and videos). Now realistic videos with matching voices can be created and distributed to online audiences easily. To exemplify the issue, already one in twenty-five Americans has been a victim of ”revenge porn” [58], something that is expected to increase considering the high-quality as well as easiness that deepfakes bring into the table.

In literature there have been observed factors that limit the benefits of feminists; more specifically: ”feminists experience new forms of exclusion of access to publicity and recognition, as digital networks can be, at the same time, spaces of uncertainty and empowerment, depending on skills, resources, and age” [24]. Deepfakes have the potential to increase such uncertainty and limit empowerment, as they can easily create and propagate discriminatory content.

It should be pointed out that the gendered aspects do not mean that women are targeted, and men are not; similar actions can, of course, be realized against men; however, the majority of cases so far have been against women and in specific roles. Overall, gender should be approached as a social construct, and gendered aspects, in conjunction with deepfakes, should be looked upon in the wider area of feminist theory and gender studies.

VIII. Law and Regulation Perspective

Several countries have laws and a regulatory framework dealing with digital media and their processes. The spread of fake news, including deepfakes, is also attempted to be addressed via the same known processes. For instance, in the US, the proposed Malicious Deep Fake Prohibition Act of 2018 ”establishes a new criminal offense related to the creation or distribution of fake electronic media records that appear realistic” [64]. In addition, an accountability act was followed up in 2019, in order to ”combat the spread of disinformation through restrictions on deep-fake video alteration technology” [65].

However, even with such legislative actions in place, while the problem is recognized, its effective addressing is challenging. AI and its implications are not well understood, and therefore assumptions are made, decisions are taken, while their applicability on AI is questionable. For instance, current laws cannot handle the complexity of AI.

In the case of the exemplified prohibition act [64], what is proposed is to toughen the consequences, at the federal level, for a practice that is already unlawful. As such, a traditional approach is taken, which, however, does not lower the risks associated with deepfakes. The accountability act [65] goes a step further and provides a better understanding of the area and lays out potential actions that need to be undertaken. However, some of these actions are unrealistic. For instance, it would require watermarks and clear labeling on deepfake content, something that surely the creators of deepfakes, especially those with nefarious intents, will not abide to. As such, its effectiveness is seen as limited. In addition, there are also some concerns raised for some of its exclusion aspects, as for instance, these conditions would not apply in specific cases of public safety or national security if this is government-generated, i.e. ”produced by an officer or employee of the United States, or under the authority thereof, in furtherance of public safety or national security” [65].

The discrepancy between deepfakes and its specific audience of regulators has tangible impacts on society as specific actions may be very difficult to be enforced. Without proper legislative capturing of deepfakes, both the executive branch that carries out the law and the judicial branch that interprets the laws will face challenges. As it can be seen, the full context of deepfakes is not well understood, and there lies the danger of (i) not addressing it in an effective manner but only superficially, (ii) introduce actions whose implications are not well understood and weaken civil rights, that may lead to long-term societal impacts.

IX. Political Perspective

Technology influences the way we think about the social and the political [31]. Especially with the latest advances in AI, not only fake news, photos, and videos (Deepfakes) can be created, but also fake reviews, convincingly realistic text, and even conversations in real-time [41]. The political mobilization of the masses is now possible via social media such as Facebook [66], and instant messaging applications such as WhatsApp, make global audiences reachable around the clock and in a personalized manner. The power of these media has already been shown in their role in recent social movements such as the Arab Spring in some developing countries [67]. Disinformation in such media e.g., via deepfake generated political videos [68] have the potential to raise uncertainty and reduce the trust placed in the news on social media.

Preprint version of doi:10.1109/TTS.2020.3001312

Both social media and instant messaging applications are seen as "fertile ground for circulating deepfakes with explosive implication for politics" [1]. The fake news issue becomes imminent as, in some developing countries, social media penetration is so high, that it is often considered as the main source of information as well as fact-checking. The problem is that fake news reduces the trust of even legitimate sources and enables misinformation or disinformation tactics, which can have detrimental effects on societal operations [68], [69]. People may react emotionally or be guided to actions that are totally the creation of a computer program and may pose a distorted view of reality.

Actions that lead to the weakening of critical functions are opposed to the fundamental rights in the social welfare state [70], and deepfakes do pose such a threat, as they have the potential to contribute to this weakening by lowering the trust on the public bodies and entities as well as the associated processes.

X. Discussion

The era of deepfakes where sensational, dishonest, or even fabricated content propagates mostly through social media is already here. With some healthy portion of skepticism and cross-referencing, one might still be able to navigate through it. Traditional ways of thinking, captured via popular sayings such as "seeing is believing", "I trust what I see", "a picture is worth a thousand words", are going to be increasingly challenged. Fabrication of photos and videos has always been challenging and could be realized only with significant efforts and expertise, but this is not state of the art anymore. Deepfakes have demonstrated convincingly, the easy access to the capability of creating realistic fake videos [11] and as such, media production aspects are significantly affected both on "how" as well as on "what" is produced.

To deal with the deepfakes phenomenon, we need to position it in the public sphere, where it mostly takes place and also where its effects can be observed. As public sphere, we consider the realm of social life where public opinion can be formed [70], which is the normative basis for a deliberate democracy. In this context, media is seen as a platform for inclusive discussions and is linked to democracy and society, since individuals and groups mobilize via it their support for their perspectives. The role of media is crucial to the governance and democratization since aspects such as community and social media (where deepfakes might be utilized) affect key areas such as poverty, inequality, and society overall [71]. Social media platforms are where deepfakes are predominantly distributed, and as such, they become an integral part of their complex dynamics. As such, deepfakes have large implications since social platforms engineer sociality [29]. One can approach such dynamics via "connecting ANT's [Actor Network Theory] recognition of the interdependence of technical, social and cultural aspects, and Castells' political analysis of the economic-legal-political stratum" [29].

There have been different media-technological innovations, from which social media and smartphones are key in the digital era [72]. In this era, media falsification is not new [3], but the fact that anyone with low technical skills can do it easily, poses new challenges, and affects the media and societal interplay. It directly impacts the classical sociological dichotomy [73] between individual agency (freedom and creativity) and structure (technological interface and peer group/community norms and expectations) for online content creation.

Deepfakes constitute a new technology-empowered manifestation of the well-known phenomenon of fake content creation. Technology savvy citizens and experts in specific domains can be used to evaluate the realness of photographs [74]. Forensic technology tools may be developed that help the effort of identification of deepfakes [16], [75]. While the threat can potentially be mitigated with current legal and technological approaches or new ones that may be developed, none of them will solve it [1], [18]. Furthermore, while the Global North may be more familiar with the cutting edge technologies and have more independent media, many developing countries in the Global South, due to the digital divide or the lack of independent media plurality, have citizens that fall often pray to fake news.

The inquisitive nature of the user who is skeptical of the communication s/he receives is seen as potentially beneficial in the effort of fighting deepfakes. A survey [76] identified that a big part of falling for fake news is due to the "general tendency to be overly accepting of weak claims", as public susceptibility and lack of awareness are seen as a problem in the identification of fake news overall [77]. Fact-checking is considered critical in the identification of fake news [78], and people that cross-check their sources have fewer chances of falling for fake material and further propagating it. There are several websites that do fact-checking [79], and people with sufficient skepticism could verify the information received prior to trusting it or propagating it in social media. However, with deepfakes, things are more challenging and complicated, especially due to their realistic nature.

Social media literacy so far, with fake news overall, may enable users to not fall for deepfakes. If they are critical of texts, photographs, and other material already, the presence of deepfakes, although more challenging, could still be addressed. News literacy [80], [81], and overall technology literacy of the citizens are seen as fundamental, and reasoning in social media environments is seen as a critical skill that, e.g., students should be taught [82]. Digital literacy [83] can help to adopt a more healthy approach in social media and act as an enabler for combating the propagation of fake information, including deepfakes. Therefore, there is an imminent need for education and training so that digital media literacy increases. However, even technology-savvy and social media literate people, loose confidence when they are confronted with the results of deepfakes.

Education and upskilling of the citizens is not the only potential line of action. In addition, news agencies must adhere to high-quality standards, and this may also be enforced via legislative actions to penalize media misinformation distribution. Transparency of media online news and source checking should be the norm [80]. In addition, new tools that can enable both citizens and journalists to identify and check the authenticity of information and pinpoint the source of

Preprint version of doi:10.1109/TTS.2020.3001312

it are needed. There needs to be a combination of human-driven analysis [79] as well as automation of it, e.g., via AI approaches [16] that can act in real-time.

Deepfakes have far-fetching implications, especially in the era where information with great easiness propagates social networks, gets communicated, read, and acted upon all over the world. While social media are fertile ground for circulating deepfakes [1] utilizing them in a specific context, e.g., politics has the potential to disrupt societal processes. The misuse potential becomes evident if one couples a deepfake photo with an appropriately crafted message, in order to create ripples in society. Deepfakes could pose as a new form of contemporary psychological warfare [84] and individual or group manipulation tool.

In political news journalism, i.e., news media coverage of politics, journalists try to be in control of political stories rather than passively report on what is promoted by political actors [23]. To do so, however, they must be able to not fall for deepfake content that is produced by stakeholders that favor such political actors directly or indirectly. For instance, in the context of social movements and activism [67], such new media technology could be used maliciously [84]. Combining a deepfake video with a serious message that fits a specific political agenda will have an impact on people. In the heat of the movement's actions, such videos can be used to discredit the opposition, create reality-near photos that would outrage citizens, and even create a fake temporal reality with events and images that support it, and by the time it is revealed to be fake, it would have served its purpose.

Trust in media, processes, and people is a major challenge, as removing trust from the news, images, videos, basically removes trust in social interactions and structures, and effectively leaves open the door for doubt everywhere, even in legitimate cases. Therefore, misinformation or disinformation tactics, can impact society and its processes [68], [69]. Truth decay coincides with trust decay [14], which raises new concerns since society can no longer share and act on accurate perceptions of reality.

Capturing the long term implications of deepfakes, and comparing it to those of the fake news may also need to be addressed. Also, since not only people have agency, but objects (e.g., algorithms) do also [29], it is relevant to investigate how technology is part of the process and how it influences and gets influenced by the ongoing processes in the area of digital media and deepfakes. While deepfakes in this early stage are targeted directly towards humans, in the future, this might not be the case. The increasing reliance on AI-empowered cyber-physical systems, e.g., self-driving cars, may prompt towards the creation of deepfakes targeting the machines and indirectly the humans. Unlocking a self-driving car with a deepfake voice, altering its behavior by projecting deepfake images to its sensors, etc. could affect the designed behavior of the car and lead it towards unpredictable decisions (e.g., sudden braking) which could harm the humans. More empirical research is needed, that is also bound to the appropriate theoretical frameworks, and provide support or not for them.

Deepfakes are a demonstration of what is possible even by home users with moderate means, and one can only imagine what can be done and at a larger scale, where resources of large enterprises are available (not to mention nation-wide resources). Oppressing or manipulative governments and organizations can utilize practices such as astroturfing [85] to tighten their grip and shape public opinion. Fighting misinformation, may also be used as a Trojan horse to bypass user privacy and undermine freedom of speech [18].

The threats posed by deepfakes need to be addressed via a combination of technology, regulation, and education. The informed and intellectual citizenry is required [86] in order to push for reform-based political and social changes. Recent research [75] portrays AI-based solutions that can detect deepfakes, including in many cases the software that was used to create them. Others have proposed complementary technologies such as Blockchain [17] in order to link videos to trusted/reputable entities. While technical solutions for identification, verification, and removal of such fakes are underway [16], [17], [75], the problem is not strictly a technological one, but one of trust to processes and stakeholders, e.g., to journalism which operates responsibly and provably.

Deepfakes and its underlying technology, pose not only threats but also opportunities. AI algorithms utilized in deepfakes have a wide range of capabilities and can create new text, voice, video, works of art, etc. In addition, even the core deepfake technology can be seen as having beneficial effects in simulation and training of personnel in customized/personalized realistic scenarios that would have been otherwise too costly or impossible to realize. There are also potential uses e.g., in the lawful provision of deception content and tactics against criminals, terrorists, and other adversaries acting against the public good. The latter was attempted to be captured in the US accountability act [65], as discussed in section VIII.

Deepfakes demonstrate a powerful technology in an emerging AI era, and as it is the case with all paradigm-shifting technologies, its use is not determined only by its capabilities, but also the regulatory framework, ethics, culture, and other societal norms [46]. Therefore, as avenues of future research, one can consider several aspects that have been indicated in this work, which, however, need to be addressed more in-depth. Such aspects include a diligent approach to the relationship between deepfakes and society, as well as its impacts. This should include how they manifest as well as their behavior over time. In addition, beneficial would be a detailed intersectional approach that covers in detail the identities, e.g., gender, race, class, sexuality, disability, and their role and impact for discrimination and social injustice. The interplay with media, culture, and society is challenging, and at this stage, empirical research, in combination with good positioning in theoretical frameworks, is lacking. Empirical research that links concrete theoretical frameworks with the utilization of deepfakes and impact in representative use cases, and verifies or disproves proposed theoretical contexts is necessary. Efforts should also be directed to technology-driven identification of deepfake materials, e.g., video, voice, text, and how these efforts may result in tools that can be utilized by the stakeholders, e.g., journalists, citizens, etc. Approaches that enhance trust in digital media sources and dependent processes are also needed, and research could be

Preprint version of doi:10.1109/TTS.2020.3001312

devoted to constructing real-world platforms, services, and tools that enable it. Since deepfakes intersect with several areas of modern life, it is also important to investigate the ethical side of it, as well as the areas related to safety and security in the societal context. Finally, research needs to be devoted to training and educational aspects of affected stakeholders, e.g., citizens as well as those involved in governance, e.g., legislative, executive, and judicial branches.

XI. Conclusions

Understanding the intersection of digital media and AI, in the case of deepfakes, and the effect on modern society is imperative if this technology is to be properly put in context and the challenges it raises are effectively addressed. From the discussions, it is evident that the intersection of digital media and deepfakes, has several impacts on the individuals as well as the society overall. Understanding deepfakes in modern digital media, as well as the processes it affects and its overall implications, is seen as challenging, and therefore they should be investigated from multiple angles that need to be considered (including a temporal aspect). To do so, however, appropriate dimensions need to be defined (which is largely not the case today), and these should be sufficient to capture all the factors involved in the interplay. This work has only revealed some high-level aspects, and a much deeper investigation is needed. There is the inherent danger that the society will no longer be able to credibly recognize in a timely fashion true and fake aspects, which might lower trust in stakeholders, processes, and journalism, and "everything is fake" motto may prevail. While technical solutions for identification, verification, and removal of such fakes is needed, the problem is not strictly technological, but should also involve regulatory measures as well as educational aspects of users.

References

[1] R. Chesney and D. Citron, "Deepfakes and the new disinformation war: The coming age of post-truth geopolitics." *Foreign Affairs*, vol. 98, no. 1, pp. 147–153, 2019.

[2] M. Westerlund, "The emergence of deepfake technology: A review," *Technology Innovation Management Review*, vol. 9, no. 11, pp. 39–52, Jan. 2019.

[3] J. Fletcher, "Deepfakes, artificial intelligence, and some kind of dystopia: The new faces of online post-fact performance," *Theatre Journal*, vol. 70, no. 4, pp. 455–471, 2018.

[4] J. Kietzmann, L. W. Lee, I. P. McCarthy, and T. C. Kietzmann, "Deepfakes: Trick or treat?" *Business Horizons*, vol. 63, no. 2, pp. 135–146, Mar. 2020.

[5] "You Won't Believe What Obama Says In This Video," YouTube, 2018. [Online]. Available: https://www.youtube.com/watch?v=cQ54GDm1eL0

[6] "Taxi Driver starring Al Pacino [DeepFake]," YouTube, 2019. [Online]. Available: https://www.youtube.com/watch?v=9NkKj0aNB0s

[7] "Avatarify: Avatars for Zoom, Skype and other video-conferencing apps," GitHub, 2020. [Online]. Available: https://github.com/alievk/avatarify

[8] "This person does not exist," website, 2019. [Online]. Available: https://thispersondoesnotexist.com

[9] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[10] "This resume does not exist," website, 2019. [Online]. Available: https://thisresumedoesnotexist.com

[11] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe, "First order motion model for image animation," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 7137–7147.

[12] "DeepFaceLab: the leading software for creating deepfakes," GitHub, 2019. [Online]. Available: https://github.com/iperov/DeepFaceLab

[13] A. Smidi and S. Shahin, "Social Media and Social Mobilisation in the Middle East: A Survey of Research on the Arab Spring," *India Quarterly: A Journal of International Affairs*, vol. 73, no. 2, pp. 196–209, Jun. 2017.

[14] R. Chesney and D. K. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," *SSRN Electronic Journal*, 2018.

[15] L. Given, *The SAGE Encyclopedia of Qualitative Research Methods*. SAGE Publications, Inc., 2008.

[16] M.-H. Maras and A. Alexandrou, "Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos," *The International Journal of Evidence & Proof*, vol. 23, no. 3, pp. 255–262, Oct. 2018.

[17] H. R. Hasan and K. Salah, "Combating deepfake videos using blockchain and smart contracts," *IEEE Access*, vol. 7, pp. 41 596–41 606, 2019.

[18] J. Pitt, "Deepfake videos and DDoS attacks (deliberate denial of satire) [editorial]," *IEEE Technology and Society Magazine*, vol. 38, no. 4, pp. 5–8, Dec. 2019.

[19] N. Couldry and A. Hepp, "Conceptualizing mediatization: Contexts, traditions, arguments," *Communication Theory*, vol. 23, no. 3, pp. 191–202, Jul. 2013.

[20] S. Schech, *Companion to Development Studies*. Hoboken : Taylor and Francis, 2014, ch. Culture and development, pp. 42–46.

[21] A. Hepp, S. Hjarvard, and K. Lundby, "Mediatization: theorizing the interplay between media, culture and society," *Media, Culture & Society*, vol. 37, no. 2, pp. 314–324, Feb. 2015.

[22] D. Deacon and J. Stanyer, "Mediatization: key concept or conceptual bandwagon?" *Media, Culture & Society*, vol. 36, no. 7, pp. 1032–1044, Aug. 2014.

[23] K. Falasca, "Political news journalism: Mediatization across three news reporting contexts," *European Journal of Communication*, vol. 29, no. 5, pp. 583–597, Jul. 2014.

[24] A. Fotopoulou, "Digital and networked by default? women's organisations and the social imaginary of networked feminism," *New Media & Society*, vol. 18, no. 6, pp. 989–1005, Sep. 2014.

[25] W. J. Potter, "Conceptualizing mass media effect," *Journal of Communication*, vol. 61, no. 5, pp. 896–915, Oct. 2011.

[26] Y. Rumpala, "Artificial intelligences and political organization: An exploration based on the science fiction work of iain m. banks," *Technology in Society*, vol. 34, no. 1, pp. 23–32, Feb. 2012.

[27] O. Hemer, *Fiction and truth in transition: writning the present past in South Africa and Agentina*, ser. Freiburg studies in social antropology: Band 34. Wien : LIT, 2012.

[28] C. Dirican, "The impacts of robotics, artificial intelligence on business and economics," *Procedia - Social and Behavioral Sciences*, vol. 195, pp. 564–573, Jul. 2015.

[29] J. van Dijck, "Facebook and the engineering of connectivity," *Convergence: The International Journal of Research into New Media Technologies*, vol. 19, no. 2, pp. 141–155, Sep. 2012.

[30] A. Kammer, "The mediatization of journalism," *MedieKultur: Journal of media and communication research*, vol. 29, no. 54, p. 18, Jun. 2013.

[31] M. Coeckelbergh, "Technology and the good society: A polemical essay on social ontology, political principles, and responsibility for technology," *Technology in Society*, vol. 52, pp. 4–9, Feb. 2018.

[32] D. J. Olsher, "New artificial intelligence tools for deep conflict resolution and humanitarian response," *Procedia Engineering*, vol. 107, pp. 282–292, 2015.

[33] C. Holder, V. Khurana, F. Harrison, and L. Jacobs, "Robotics and law: Key legal and regulatory implications of the robotics age (part i of II)," *Computer Law & Security Review*, vol. 32, no. 3, pp. 383–402, Jun. 2016.

[34] C. Holder, V. Khurana, J. Hook, G. Bacon, and R. Day, "Robotics and law: Key legal and regulatory implications of the robotics age (part II of II)," *Computer Law & Security Review*, vol. 32, no. 4, pp. 557–576, Aug. 2016.

[35] R. Alharthi, B. Guthier, and A. E. Saddik, "Recognizing human needs during critical events using machine learning powered psychology-based framework," *IEEE Access*, vol. 6, pp. 58 737–58 753, 2018.

[36] P. Ducange, R. Pecori, and P. Mezzina, "A glimpse on big data analytics in the framework of marketing strategies," *Soft Computing*, vol. 22, no. 1, pp. 325–342, Mar. 2017.

[37] L. Degerstedt and S. Pelle, "More media, more people – on social & multimodal media intelligence," *Human IT*, vol. 13, no. 3, pp. 54–84, 2017.

Preprint version of doi:10.1109/TTS.2020.3001312

[38] M. Schudson, "The sociology of news production," *Media, Culture & Society*, vol. 11, no. 3, pp. 263–282, Jul. 1989.

[39] M. J. Kushin, M. Yamamoto, and F. Dalisay, "Societal majority, facebook, and the spiral of silence in the 2016 US presidential election," *Social Media + Society*, vol. 5, no. 2, p. 205630511985513, Apr. 2019.

[40] R. Kunelius and E. Reunanen, "Changing power of journalism: The two phases of mediatization," *Communication Theory*, vol. 26, no. 4, pp. 369–388, Jul. 2016.

[41] P. Janaszkiewicz, J. Krysińska, M. Prys, M. Kieruzel, T. Lipczyński, and P. Różewski, "Text summarization for storytelling: Formal document case," *Procedia Computer Science*, vol. 126, pp. 1154–1161, 2018.

[42] Y. Jia, Y. Zhang, R. J. Weiss, Q. Wang, J. Shen, F. Ren, Z. Chen, P. Nguyen, R. Pang, I. L. Moreno, and Y. Wu, "Transfer learning from speaker verification to multispeaker text-to-speech synthesis," *CoRR*, 2018.

[43] I. Solaiman *et al.*, "Release strategies and the social impacts of language models," OpenAI, techreport, Nov. 2019. [Online]. Available: https://d4mucfpksywv.cloudfront.net/papers/GPT_2_Report.pdf

[44] J. Thies, M. Elgharib, A. Tewari, C. Theobalt, and M. Nießner, "Neural voice puppetry: Audio-driven facial reenactment," *arXiv 2019*, 2019.

[45] D. Kim and S. Kim, "Newspaper journalists' attitudes towards robot journalism," *Telematics and Informatics*, vol. 35, no. 2, pp. 340–357, May 2018.

[46] S. Karnouskos, "Self-driving car acceptance and the role of ethics," *IEEE Transactions on Engineering Management*, vol. 67, no. 2, pp. 252–265, May 2020.

[47] T. Montal and Z. Reich, "I, robot. you, journalist. who is the author?" *Digital Journalism*, vol. 5, no. 7, pp. 829–849, Aug. 2016.

[48] T. Hopp and H. Gangadharbatla, "Examination of the factors that influence the technological adoption intentions of tomorrow's new media producers: A longitudinal exploration," *Computers in Human Behavior*, vol. 55, pp. 1117–1124, Feb. 2016.

[49] E. Freidson, *Professionalism : the third logic*. University of Chicago Press, 2001.

[50] S. Hall, *Representation : cultural representations and signifying practices.*, ser. Culture, media, and identities. London ; Thousand Oaks, Calif. : Sage in association with the Open University, 1997., 1997.

[51] C. Bartneck, "Robots in the theatre and the media," in *Design & Semantics of Form & Movement (DeSForM), Wuxi, China*, 2013, pp. 64–70.

[52] S. L. Epstein, "Wanted: Collaborative intelligence," *Artificial Intelligence*, vol. 221, pp. 36–45, Apr. 2015.

[53] K. L. Young and C. Carpenter, "Does science fiction affect political fact? yes and no: A survey experiment on "killer robots"," *International Studies Quarterly*, vol. 62, no. 3, pp. 562–576, Aug. 2018.

[54] K. F. MacDorman and H. Ishiguro, "The uncanny advantage of using androids in cognitive and social science research," *Interaction Studies*, vol. 7, no. 3, pp. 297–337, Nov. 2006.

[55] M. Mara and M. Appel, "Science fiction reduces the eeriness of android robots: A field experiment," *Computers in Human Behavior*, vol. 48, pp. 156–162, Jul. 2015.

[56] Y. Mou and K. Xu, "The media inequality: Comparing the initial human-human and human-AI social interactions," *Computers in Human Behavior*, vol. 72, pp. 432–440, Jul. 2017.

[57] K. Nash, J. M. Lea, T. Davies, and K. Yogeeswaran, "The bionic blues: Robot rejection lowers self-esteem," *Computers in Human Behavior*, vol. 78, pp. 59–63, Jan. 2018.

[58] A. Lenhart, M. Ybarra, and M. Price-Feeney, "Nonconsensual image sharing: One in 25 americans has been a victim of "revenge porn"," Data & Society Research Institute, and Center for Innovative Public Health Research (CiPHR), USA, Tech. Rep., 2016. [Online]. Available: https://datasociety.net/pubs/oh/Nonconsensual_Image_Sharing_2016.pdf

[59] R. Clark, ""hope in a hashtag": the discursive activism of #WhyIStayed," *Feminist Media Studies*, vol. 16, no. 5, pp. 788–804, Feb. 2016.

[60] K. Busse and J. Gray, "Fan cultures and fan communities," in *The Handbook of Media Audiences*. Wiley-Blackwell, Apr. 2011, pp. 425–443.

[61] J. A. Brown, "#wheresRey: feminism, protest, and merchandising sexism in star wars: The force awakens," *Feminist Media Studies*, vol. 18, no. 3, pp. 335–348, Apr. 2017.

[62] P. Haridakis, *The International Encyclopedia of Media Studies*. Blackwell Publishing Ltd, 2013, ch. Uses and Gratifications: A Social and Psychological Perspective of Media Use and Effects.

[63] L. Steiner, "Feminist media theory," in *The Handbook of Media and Mass Communication Theory*. John Wiley & Sons, Inc., Mar. 2014, pp. 359–379.

[64] B. Sasse, "Malicious deep fake prohibition act of 2018," Dec. 2018. [Online]. Available: https://www.congress.gov/115/bills/s3805/BILLS-115s3805is.pdf

[65] Y. D. Clarke, "H. R. 3230, To combat the spread of disinformation through restrictions on deep-fake video alteration technology," Jun. 2019. [Online]. Available: https://www.congress.gov/116/bills/hr3230/BILLS-116hr3230ih.pdf

[66] V. P. Miletskiy, D. N. Cherezov, and E. V. Strogetskaya, "Transformations of professional political communications in the digital society (by the example of the fake news communication strategy)," in *2019 Communication Strategies in Digital Society Workshop (ComSDS)*. IEEE, Apr. 2019.

[67] M. Ancelovici, P. Dufour, and H. Nez, Eds., *Street Politics in the Age of Austerity: From the Indignados to Occupy*. Amsterdam University Press, Dec. 2016.

[68] C. Vaccari and A. Chadwick, "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news," *Social Media + Society*, vol. 6, no. 1, p. 205630512090340, Jan. 2020.

[69] Z. Tufekci, *Twitter and tear gas : the power and fragility of networked protest*. New Haven London: Yale University Press, 2017.

[70] J. Habermas, S. Lennox, and F. Lennox, "The public sphere: An encyclopedia article (1964)," *New German Critique*, no. 3, pp. 49–55, 1974.

[71] M. Scott, *Media and development.* Zed Books, 2014.

[72] N. Couldry and A. Hepp, *The Mediated Construction of Reality : society, culture, mediatization.* Polity Press, 2016.

[73] V. Kalmus, P. Pruulmann-Vengerfeldt, A. Siibak, and P. Runnel, "Mapping the terrain of "generation c": Places and practices of online content creation among estonian teenagers." *Journal of Computer-Mediated Communication*, vol. 14, no. 4, pp. 1257–1282, 2009.

[74] M. Gates, "Is seeing still believing: Factors that allow humans and machines to discriminate between real and generated images," *SMPTE Motion Imaging Journal*, vol. 127, no. 9, pp. 70–78, Oct. 2018.

[75] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, "FaceForensics++: Learning to detect manipulated facial images," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2019.

[76] G. Pennycook and D. G. Rand, "Who falls for fake news? the roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking," *Journal of Personality*, vol. 88, no. 2, pp. 185–200, Apr. 2019.

[77] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, and Y. Liu, "Combating fake news," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 3, pp. 1–42, Apr. 2019.

[78] L. Whitney, "How to spot fake news online." *PC Magazine*, pp. 155–159, 2019.

[79] J. Hyman, "Addressing fake news: Open standards & easy identification," in *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*. IEEE, Oct. 2017.

[80] J. M. Pérez Tornero, S. Samy Tayie, S. Tejedor, and C. Pulido, "How to confront fake news through news literacy? state of the art." *Doxa Comunicación*, pp. 211–235, 2018.

[81] G. Lotero-Echeverri, L. M. Romero-Rodríguez, and M. A. Pérez-Rodríguez, ""fact-checking" vs. "fake news": Periodismo de confirmación como recurso de la competencia mediática contra la desinformación," *Index Comunicación*, vol. 8, no. 2, pp. 295–316, May 2018.

[82] S. McGrew, T. Ortega, J. Breakstone, and S. Wineburg, "The challenge that's bigger than fake news: Civic reasoning in a social media environment." *American Educator*, vol. 41, no. 3, pp. 4–9, 2017.

[83] S. Horn and K. Veermans, "Critical thinking efficacy and transfer skills defend against 'fake news' at an international school in finland," *Journal of Research in International Education*, vol. 18, no. 1, pp. 23–41, Feb. 2019.

[84] K. A. Pantserev, "The malicious use of AI-based deepfake technology as the new threat to psychological security and political stability," in *Advanced Sciences and Technologies for Security Applications*. Springer International Publishing, 2020, pp. 37–55.

[85] C. H. Cho, M. L. Martens, H. Kim, and M. Rodrigue, "Astroturfing global warming: It isn't always greener on the other side of the fence," *Journal of Business Ethics*, vol. 104, no. 4, pp. 571–587, Jul. 2011.

[86] B. Mutsvairo, *Digital activism in the social media era : critical reflections on emerging trends in Sub-Saharan Africa*. Cham, Switzerland: Palgrave Macmillan, 2016.

Preprint version of doi:10.1109/TTS.2020.3001312