Aug 10th, 12:00 AM

Towards a Theory of Digital Stigma and Deep Fake Video Technology Stigmatization in a Digitally Mediated Environment

Fatima Mohammed
University of North Carolina at Greensboro, f_mohammed@uncg.edu

A. F. Salam
UNC Greensboro, amsalam@uncg.edu

Towards a Theory of Digital Stigma and Deepfake Video Technology: Stigmatization in a Digitally Mediated Environment

Emergent Research Forum (ERF)

Fatima Mohammed
The University of North Carolina at Greensboro
ISSCM Department
f_mohammed@uncg.edu

A. F. Salam
The University of North Carolina at Greensboro
ISSCM Department
amsalam@uncg.edu

Abstract

As social structures and IT becomes intertwined, there is a need for new theoretical perspectives that advance understanding of this intertwinement and its implications for IT and social actors, institutions, and society. This research attempts to develop a theory that conceptualizes the role of technology in creating and manipulating stigma symbols and the subsequent consequences of the resultant symbolic interaction on the social actors, technology, and society. Understanding the changing nature of stigma due to technology is essential in understanding resulting societal manifestations and aids in the timely management of societal implications of the dark side of technology.

Keywords

Digital stigma, symbolic interactionism, IS theories, deepfake.

Introduction

Vignette

After a teacher's deepfake videos surfaced depicting her in sexually inappropriate activity, she was dismissed even after convincing evidence that the video was fake. The harm had been done already. Her society including the school board, her colleagues, and the parents of her students, are concerned about being associated with such a "stained" reputation if she continues to be part of the school. Society has categorized her as "not normal," a "stigmatized individual" due to the depiction of "her" in an activity that departs from society's approved standards of behavior or, in this case, the category for being a teacher.

Stories like that described in the vignette are increasingly becoming common. A recent study found that 96% of deepfake videos are about women depicted in various socially unacceptable behavior, such as in fake pornography videos or fake sexually explicit images (Ajder et al., 2019). This misuse of the technology is associated with consequences for both the victims of deepfake videos and the audience and society. For example, the teacher described here in the vignette lost her job and had to sever her relationships with students' parents, colleagues, and maybe even close friends – with no fault of her own. In addition to the harm done to the teacher, there are several elements of such stories pointing to the changing fabric of society due to technology that has been ignored. Unfortunately, studies on deepfakes have primarily focused on the automatic generation and detection of deepfakes while ignoring the detrimental social implications of the dark side of this technology (Hancock & Bailenson, 2021) for both the individual victim and the proximal society of the victim. While deepfake technology presents immense potential in various industries, e.g., healthcare, entertainment, and tourism, this study focuses on the dark side of the technology, looking at its detrimental impact on the manipulation of social stigma symbols and related consequences.

With the rapid advancement of technology and its mediation in societies, sociological concepts that are rooted in our interactions with society, such as stigma, have taken different shapes and forms. As social structures and IT become intertwined, there is a need for new theories that embrace this intertwinement to advance understanding of the implications of IT for users, societies, and institutions (Floridi, 2010; Schwarz, 2021). In recent years, the need for such theories has garnered increasing interest among IS researchers resulting in imperative theories such as the IT identity (M. Carter & Grover, 2015), which seeks to understand the role of technology in relation to identity. This study seeks to add to such conversation to better understand IT's complex emerging nature and implications in today's digitally mediated world by articulating the construct of digital stigma. Digital Stigma is defined as the discrepancies that emerge between the society's assumptions and expectations of what an individual ought to be (social norms and expectations based on established categories such as teacher, lawyer, doctor, etc.) and what they are depicted to be, in a socially unacceptable behavior or action, in a digitally manipulated deepfake video. Given that stigma is a critical concept associated with labeling, stereotyping, separation, status loss, discrimination, and even power (Link & Phelan, 2001), understanding the role of deepfake video technology in relation to stigma is essential for understanding the negative impact of such technology on individuals and society.

We study digital stigma in relation to deepfake technology. As seen in the opening vignette, deepfake technology has the ability to subvert the basic human understanding of reality (Conwell, 2020). In order to understand the larger dark side effect of deepfake video technology, we need to broaden our conceptualization to include deepfake video-induced stigma and its symbolic impact. We draw from symbolic interaction theory and stigma theory to develop a theory of digital stigma and symbol. We then analyze the implications of this new construct, the individual victim, and society in the context of a digitally mediated society and, in the process, show how pervasive deepfake video can be.

Background

Recent IS research on stigma has studied the use of technologies by a stigmatized population and the stigmatization of IT professionals (e.g., (Wade et al., 2020). Yet these studies look at IT differently from the dynamics of stigma, its creation, and its implications in a digital world. In a study of the stigmatization of employees in an organization, (Kulik et al., 2008) emphasize the need to examine the processes that lead to stigmatization. This is because the process, while normally subtle, may result in devastating consequences. Carter and Grover's (2015) work on IT identity (the extent to which IT is seen as an integral part of self) points IS researchers to the importance of studying sociological processes as it provides grounds for studying people's behavior and how they view themselves in relation to IT. Similarly, generating a digital stigma construct allows for a comprehensive understanding of the implications of technologies such as deepfake videos and the manifestations of society towards its use.

With this in mind, we develop a theory that conceptualizes the role of technology in creating and manipulating stigma symbols and the subsequent consequences of the resultant symbolic interaction on the user and the society. The digital stigma concept we construct has implications for understanding several core issues in IS and stigma research, ranging from the definition of the concept, identification of new forms and directions, increasingly complex levels of stigma, and the rippling consequences on other related social concepts such as identity.

Theoretical Underpinning of Deepfake and Digital Stigma

Stigma

In Goffman's early elaboration of stigma, he classically defined the concept as "an attribute that is deeply discrediting". Stigma has largely been studied on the micro-level, highlighting and addressing the treatment and wellbeing of stigmatized individuals. More recently, sociologists have investigated the macro-level dimensions of stigma, taking into consideration structural causes, population-level consequences, and collective responses (Clair, 2018). Stigma has been used to study the association of socially adverse characteristics such as ethnicity, mental health, physical differences, discrimination, and social class (Brunson, 2007; Harrison & Gill, 2010; Madera & Hebl, 2012; Varni et al., 2012). Streams of

research regarding stigma have advanced to address the impact of stigma and how it can be managed at several levels, including individual, occupational, organizational, and industry (Zhang et al., 2021).

Stigma, in general, has not received much attention within the IS domain. Mainly, IS research has investigated information system use by stigmatized identities (e.g., Wade et al., 2020) or the role of IS in addressing stigma issues, but not the role of technology in the stigmatization process.

Symbolic interaction as a framework

Symbolic interaction (SI) is a theoretically robust framework to study how society is created and maintained through face-to-face, repeated, and meaningful interaction between individuals (M. J. Carter & Fuller, 2016). It focuses on the interaction between individuals and the social structure and situations and hence an appropriate lens to study the technology-mediated interactions. Fundamental to symbolic interaction is the ontological assumption that reality is "socially constructed." Hence what society believes to be true is based on symbols and shared meanings. Reality, then, is based on observations, interpretations, perceptions, and conclusions society can agree on and consequently act upon. SI also asserts that humans have the capacity to internally manipulate symbols with the hopes of manipulating actions (Stryker, 2008). This is very much clarified in Goffman's "The Presentation of Self in Everyday Life," which explains that when an individual comes into contact with someone, they control their setting, appearance, and manner based on their perceptions of the people they interact with the aim of controlling others' impressions of them (Goffman, 1978). Concurrently, the people they interact with perceive both the manner the individual tries to appear, and their actual appearance and act based on their conclusions. As such, when interacting, people strive to convey an identity consistent with the expectations formed by the audience and with the situation that frames the interaction (Robinson, 2007). Their success (or failure) in performing affects how people react towards them.

Symbols and Digital Symbols

Symbols are social objects that have shared meanings created and maintained in social interactions. Symbols are very important because the meanings attached to symbols inform how people act. Symbols are, therefore, means by which we categorize people and relate to people. IS literature has largely studied the important role of technology in social interactions and structures (e.g., (M. Carter & Grover, 2015)). Underlying most of these studies is the symbolic interactionist assumption that reality is "socially constructed. As such, these studies have been largely based on the premise of self-regulation of symbols, in that; individuals control their behaviors to affect the meanings and actions towards them.

This phenomenon becomes limited when digital technologies can create or manipulate social symbols. The creation of a deepfake video involves manipulating personal data objects in a particular conduct determined by the deepfake creator rather than the one who is being portrayed. Society's perception is based on meanings derived from the symbols and not necessarily the creation of these symbols, as fundamentally, identity symbols are self-regulated, as seen in assumptions of SI. In the beginning vignette, for example, the teacher has no way of changing her depiction in the video. As identity has been objectified, the video remains not only in the memory of those that have seen it, but it remains in circulation and stored for future references and interactions.

Fake symbols created by deepfake are from personalized data objects, based on which society and algorithms make meanings from and act accordingly, like the video of the teacher going viral and her school dismissing her as a societal action. Deepfake, therefore, mediates society's assumptions of what an individual ought to be (social norms) and what "they actually are or manipulated to be" – a discrepancy that can lead to stigmatization. This study is therefore built on contending the principal assumption of SI by extending the role of symbol creation; by proposing that human behavior can be digitally regulated.

Stigma as a Symbol: Conceptual Definition of Digital Stigma

According to Goffman, symbols carry social information, either positive (status symbols) or negative (stigma symbols)(Goffman, 2009). Stigmatized symbols communicate a certain meaning and reduce the valuation of the individual, hence affecting how people see and act towards the individual. Underlyingly, stigma symbols can serve as instigators of certain discriminatory behaviors. Stigma is a social construct

that emerges from some discrepancies between society's assumptions of what an individual ought to be (social norms) and what they actually are. Stigma, therefore, arises from differences between two different selves (actual self and virtual self) (Goffman, 2009). A recent study on deepfakes revealed that amidst the alarming increase in the number of deepfake videos online, an overwhelming majority of these videos (96%) depicted acts categorized by society as discrediting (Ajder et al., 2019). Deepfakes technologies are therefore manipulating identity symbols so that they deviate from what society terms "normal." This element of symbol manipulation introduces another "self"; what an individual has been made to be.

Digitalized stigma symbols are, therefore, digitally mediated symbols created from personal data objects whose meanings facilitate the stigmatization of victims without the victim's control. Digital stigma, therefore, underlines discrepancies that emerge between society's assumptions of what an individual ought to be (social norms) and what they are digitally manipulated to be (e.g., deepfake videos). Advancements in technology such as deepfake have aided the creation of false symbols that contain social information that leads to the stigmatization of an individual.

What digital stigma is not?

As we define digital stigma as the discrepancies that emerge between society's assumptions of what an individual ought to be (social norms) and what they are digitally manipulated to be, it is important to clarify what the concept is not to help deepen conceptual boundaries. The concept of digital stigma is not the mere act of stigmatization on online platforms or the use of digital technologies by a stigmatized or a non-stigmatized group, or the broad role of technology in stigmatization. It is the intricate role that technology plays in manipulating social symbols resulting in a change in the reality of the society by affecting their interpretations and actions towards those manipulated symbols.

Meaning Making and Action

From a symbolic interactionist perspective, the basic building blocks of interaction are meaning-making and action. According to Hennestad, (1987), most studies ignore the "deeper structure of meanings and symbols" in information technology and thus miss the expressive world of organizational or social actors. We, therefore, continue to delve deeper to attempt to discuss facets of the digitally mediated environment that affect the meaning-making and actions towards digital stigma symbols.

Objectification of Interactions

Digital mediated interactions have facilitated the creation of secondary order interactions due to the objectification of interactions. Most forms of digitally mediated interactions are self-documenting, hence leaving behind an objectified version of the interaction (Schwarz, 2021). For instance, a shared deepfake video allows secondary order interaction with anyone the media is shared with. Anytime someone sees the content, an interaction is renewed, no matter how old the content has been in circulation. The audience continues to make meanings from the symbols and acts accordingly. Subsequently, this implies that actions following a particular interaction are recurring. Due to the remolding of the availability and accessibility of the interaction to others, objectification of interactions, therefore, has daring consequences (Schwarz, 2021).

Space and Time

Long before computer-mediated interactions, space and time were essential to communication. As such, early research on interaction has been built on the necessary assumption of physical presence and how the contemporaneous physical presence of others influences individual actions; termed copresence by Goffman (2008). Physical space and time define situations in which actors interpret symbols and actions bound to occur. Yet, digitally mediated interactions challenge these assumptions of copresence. The creator of the false symbols does not have to see or know the victim. Similarly, society does not need to establish physical contact with the victim or know them personally. People could still access digital content and can make meaning and act based on the content.

Future Research Directions

As stigma is conceptualized as a process based on the social construction of identity, this study will continue by looking at the effects of digital stigma symbols on identity construction and related consequences. Specifically, we look at the implications of the manipulated stigma symbol on the personal and social identity of the individual. This will be done by drawing from the rich identity theories in relation to the concurrence of stigma (such as discrimination, labeling, power) and its consequences (including social, economic, emotional, and physiological). The results of this study will extend our understanding of a new form of stigma (i.e., digital stigma), its sources, and its implications.

REFERENCES

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). *The State of Deepfakes: Landscape, Threats, and Impact*. (p. 27). Deeptrace.

Brunson, R. (2007). 'Police Don't Like Black People': African-American Young Men's Accumulated Police Experiences. *Criminology & Public Policy*, *6*, 71–101. https://doi.org/10.1111/j.1745-9133.2007.00423.x

Carter, M., & Grover, V. (2015). Me, My Self, and I(T): Conceptualizing Information Technology Identity and its Implications. *MIS Quarterly*, *39*(4), 931–957. https://doi.org/10.25300/MISQ/2015/39.4.9

Carter, M. J., & Fuller, C. (2016). Symbols, meaning, and action: The past, present, and future of symbolic interactionism. *Current Sociology*, *64*(6), 931–961.

Clair, M. (2018). *Stigma: Forthcoming in Core Concepts in Sociology*.

Conwell, J. (2020, January 29). 2020: A year of deepfakes and deep deception. *Help Net Security*. https://www.helpnetsecurity.com/2020/01/29/deepfake-technology/

Floridi, L. (2010). *Information: A very short introduction*. OUP Oxford.

Goffman, E. (1978). *The presentation of self in everyday life* (Vol. 21). Harmondsworth London.

Goffman, E. (2009). *Stigma: Notes on the management of spoiled identity*. Simon and Schuster.

Hancock, J. T., & Bailenson, J. N. (2021). The Social Impact of Deepfakes. *Cyberpsychology, Behavior, and Social Networking*, *24*(3), 149–152. https://doi.org/10.1089/cyber.2021.29208.jth

Harrison, J., & Gill, A. (2010). The experience and consequences of people with mental health problems, the impact of stigma upon people with schizophrenia: A way forward. *Journal of Psychiatric and Mental Health Nursing*, *17*, 242–250. https://doi.org/10.1111/j.1365-2850.2009.01506.x

Hennestad, B. rn W. (1987). Organizations and the computer culture: The mismanagement of meaning. *Organization of Innovation: East-West Perspectives*, 137–147.

Kulik, C. T., Bainbridge, H. T., & Cregan, C. (2008). Known by the company we keep: Stigma-by-association effects in the workplace. *Academy of Management Review*, *33*(1), 216–230.

Link, B. G., & Phelan, J. C. (2001). Conceptualizing Stigma. *Annual Review of Sociology*, *27*(1), 363–385. https://doi.org/10.1146/annurev.soc.27.1.363

Madera, J. M., & Hebl, M. R. (2012). Discrimination against facially stigmatized applicants in interviews: An eye-tracking and face-to-face investigation. *Journal of Applied Psychology*, *97*(2), 317.

Robinson, L. (2007). The cyberself: The self-ing project goes online, symbolic interaction in the digital age. *New Media & Society*, *9*(1), 93–110.

Schwarz, O. (2021). *Sociological Theory for Digital Society: The Codes that Bind Us Together*. John Wiley & Sons.

Stryker, S. (2008). From Mead to a structural symbolic interactionism and beyond. *Annu. Rev. Sociol*, *34*, 15–31.

Varni, S. E., Miller, C. T., McCuin, T., & Solomon, S. (2012). Disengagement and engagement coping with HIV/AIDS stigma and psychological well-being of people with HIV/AIDS. *Journal of Social and Clinical Psychology*, *31*(2), 123–150.

Wade, J., Dinger, M., & Thatcher, J. B. (2020). A Great Escape: The Effect of Negative Public Affiliation on Belongingness to Virtual Communities. *Communications of the Association for Information Systems*, *47*(1), 45.

Zhang, R., Wang, M. S., Toubiana, M., & Greenwood, R. (2021). Stigma beyond levels: Advancing research on stigmatization. *Academy of Management Annals*, *15*(1), 188–222.