ARE DEEPFAKES A THREAT? REDEFINING DEEPFAKE-AI THROUGH POPULAR CULTURE & THE EVERYDAY

By

Alice Somogyi

Submitted to

Central European University

Department of International Relations

In partial fulfillment of the requirements for the degree of

Masters of Arts in International Relations

Supervisor: Professor Michael Merlingen

Vienna, Austria

June 2023

Abstract

While deepfakes have gained momentum as a potentially dangerous technology, the current deepfake literature has largely discarded contributions of the everyday (wo)man. To bridge this gap, this thesis employs the theoretical lenses of popular culture in world politics and the everyday in International Relations (IR) for understanding the deepfake knowledge spectrum, that is by complementing dominant academic knowledge with everyday knowledge generated by "ordinary" people. It thus explores the forms of knowledge produced on Reddit—a social media platform that hosts an active deepfake community—about deepfakes and how they relate to dominant knowledge. In doing so this research aims to display everyday knowledge production by adopting innovative insights that shed light on the complex interplay between dominant and creative knowledge within the micro-level deepfake realm. Applying congruence and discourse analysis and using two case studies allowed us to retrieve marginalized perspectives and diverse knowledge production processes. We find that Reddit serves as a popular site where both dominant and alternative knowledge are (re)produced. Notably, some Reddit users generate novel insights by normalizing deepfakes, considering it an extension of normality, while others challenge dominant disourses through the adoption of progressive and regressive Foucauldian counter-conducts, or what we introduce as counter-knowledge. This exploratory thesis encourages a more holistic comprehension of this emerging AI phenomenon, while highlighting the need for further examination of deepfake knowledge production within everyday contexts.

Acknowledgement

I would like to extend my gratitude to Professor Michael Merlingen, Professor Erzsebet Strausz and Professor Vera Eliasova for their support and guidance provided throughout the course of this project.

And *Bisou, Bisou* to friends and family.

Table of Contents

"It went from 'You can't believe everything you see on the internet' to 'You can't believe *anything* you see on the internet' pretty quick."

— Reddit user Synyster328

Introduction

Former German Chancellor Angela Merkel in pornographic settings (Benve 2022), actress Emma Watson reading Mein Kamp (Sellman 2023), or Ukrainian President Zelensky declaring defeat in the Russian-Ukrainian war (Byman et al. 2023, 2): all are audio-visuals that have been circulating in recent years. But are they real?

As one online user put it, "[i]t went from 'You can't believe everything you see on the internet' to 'You can't believe *anything* you see on the internet' pretty quick" (u/Synyster328, March 27, 2023, 04:27 a.m., comment on KingArthursLance 2023). Since 2017 surprisingly realistic audios, images and videos have emerged online, sparking significant and widespread concerns. Yet these are in fact deepfakes, or convincingly manipulated audio-visual content generated through the latest advancements in Artificial Intelligence (AI) technology. With the surge in disinformation and the exponential growth in easy-to-use AI softwares, deepfakes have permeated the online landscape, blurring the boundaries between reality and fabrication. "Dangerous", "risky", "threatening," are typically employed to describe deepfakes in scholarly discourses, thus revealing the dominant understanding of deepfakes as inherently bad and highlighting their potential to undermine trust, fuel disinformation, and erode the very fabric of society.

However, this prevailing understanding of deepfakes as exclusively menacing fails to capture the nuanced web of knowledge production surrounding this emerging AI phenomenon. While academia has largely shaped the discourse, an important avenue remains unexplored— the role of everyday discussions and popular culture in complementing the deepfake knowledge

1

spectrum. To address this gap, this thesis employs the theoretical lenses of the everyday in International Relations (IR) and popular culture in world politics to delve into the realm of the "ordinary" people so as to illuminate the multifaceted nature of deepfake knowledge production since popular culture and the everyday are sites for the reproduction of dominant knowledge and for creatively being, doing, and knowing otherwise.

At the center of this study lies Reddit, a social media platform that has emerged as a unique hub for open and transparent deepfake creation and discussion. Since deepfakes' first appearance on the platform in 2017, Reddit has hosted a community of users creating, sharing, and debating deepfake content. But interestingly academics have not yet turned towards Reddit despite it being the deepfake popular site *par exellence* and birthplace of this new AI.

Drawing on theoretical frameworks grounded in the everyday in IR and popular culture, we embark on an exploratory qualitative research on the forms of everyday knowledge produced about and with deepfakes by Reddit users, and how this ties in with academic knowledge production. Through a combination of congruence and discourse analysis and using two deepfake case studies, we delve into this digital realm, investigating the intricate interplay between dominant scholarly knowledge and the marginalized understandings generated within the Reddit community. Our empirical exploration seeks to test our hypotheses that Reddit users both reproduce and challenge dominant scholarly knowledge on deepfakes, while generating what may be termed counter-knowledge, a form of Foucauldian counter-conduct that creatively resists dominant knowledge so as to further progressive agendas. The overall findings show Reddit to be a site where dominant knowledge is indeed reproduced, and where individuals produce knowledge otherwise by normalizing deepfakes, in addition to express limited forms of progressive but also regressive counter-knowledge. Interestingly, the empirical findings also uncover the surprising limits of creativity despite the ambitious deepfake/human imagination intersection.

2

This exploratory thesis serves as a call for a more holistic understanding of deepfakes, moving beyond the predominant narrative of danger and recognizing the unique insights of popular culture and the everyday in complementing the current deepfake knowledge production. By unraveling the diverse knowledge spectrum among Reddit users, our study does not aim to make generalizations but is a first step towards acknowledging the complex intersection of technology, popular culture, and everyday experiences, thus encouraging more research on deepfakes within broader popular culture sites by bridging the gap between dominant scholarly knowledge and the rich tapestry of alternative understandings offered by IR.

Chapter 1 begins by presenting the everyday in IR and popular culture as important sites of knowledge production, a crucial theoretical framework that is followed by an overview of the current limited deepfake literature which will inform the rest of our research. Building upon these literatures, we then transition to a detailed exploration of Reddit as a crucial site for popular deepfake knowledge production in Chapter 2, which includes an overview of our specific research question and aim. Chapter 3 then displays the qualitative methodology employed, that is congruence and discourse analysis as analytical frameworks that allow us to uncover diverse perspectives and knowledge production processes within deepfake discourses, before ending this thesis with a comprehensive conclusion in Chapter 5.

Chapter 1. Everyday IR and Popular Culture: Interconnected Sites of Digital Knowledge Production

1.1.   The Everyday and Popular Culture

It must be noted that the political extends beyond parliaments, summits, revolutions, riots or conferences. They are far more implicit, permeating our everyday lives on a daily basis: through memes, advertisements, Twitter-posts, movies, comic books, etc. So far, nothing new: Roland Barthes already applied this rationale back in 1957. Amongst the first academics to understand the importance of the everyday, it is in his book *Mythologies* that Barthes treated banal practices and artifacts as crucial sites of research (Barthes 1957). The most trivial objects seemed of peculiar importance: images, advertisement, wine, soap… all became worthy of analysis for the author who perceived these objects as carrying value and language that transpired beyond its mere physical existence. But while many scholars now consider *Mythologies* fundamental to the study of contemporary culture (Easthope 1991, 140; Storey 1993, 77; Stivale 2002, 458), Barthes' departure from more traditional objects of research was a new approach that scholars still had to digest.

To this day, some still flinch at the idea that cultural artifacts could be read in terms of their supplementary meaning, not to mention International Relations (IR), a field which also still struggles with the politics of the everyday (Crilley 2021, 166). Indeed, despite popular culture being an increasingly studied site of world politics those past two decades, its inclusion has not always been welcomed with open arms (Shepherd et al. 2016, 4). Popular culture has, and still is, met with resistance, including in the field of IR. "Popular culture is ostensibly everything that world politics is not: fiction, entertainment, amusement," (Rowley 2015, 361) as one would argue. This 'stuckness' in a narrow definition of IR leads to popular culture being considered as unworthy of IR's attention because it is categorized as "low politics, domestic politics, or

not political at all" (Weldes and Rowley 2015, 1). This reflects how world politics and popular culture are conceptualized within IR as "potentially interconnected but ultimately separate domains" (Grayson, Davies, and Philpott 2009, 155). One domain is the traditional one perceived as the most trusted method to uncover knowledge, the other domain suffers from "a question of knowledge" as Warner (1990, 727) puts it, explaining that popular culture, due to its variety, unboundedness and ambient, challenges various academic discourses, thus seeing its legitimacy suffer from it.

But the tide has turned, and a growing number of scholars see a great deal of potential in the study of popular culture, identified as "an important site where power, ideology and identity are constituted, produced and/or materialized" (Grayson, Davies, and Philpott 2009, 155-156). More importantly, popular culture is "a site of cultural negotiation" where lively exchanges of understandings on various topics take place (Hasselstein, Ostendorf and Schneck 2001, 333). With popular cultural objects revealing the rich connotations encapsulated in everyday life, it is not only productive of local and global politics, but also of identities and agency (Dixit 2012; Pusca 2015; Innes and Topinka 2016; Wedderburn 2019).

Mark Sachleben's (2014, 2) was right, "[i]f you stop and think about it, it is remarkable how much of our everyday lives are shaped by popular culture." This realization is translated into the close association of popular culture and world politics with everyday IR literature. Crilley (2021, 167) admits that a change is occurring in the discipline thanks to the increasing attention to everyday narratives within IR where scholars draw on popular culture and focus groups in an effort to better understand the interrelatedness of both strands. As attested by recent works (Pears 2016; da Silva and Crilley 2017; Bos 2018), both paths of everyday IR and popular culture in world politics are closely linked to one another. Hence, we are now encouraged to see international relations in the everyday, by reading politics in familiar places (Grayson, Davies, and Philpott 2009, 160).

This comes hand in hand with our argument on the dialectical idea that the everyday and popular culture are sites for both the reproduction of dominant knowledge and for creatively being, doing, and knowing otherwise. These literatures are of interest to IR in that the everyday and popular culture can reproduce, and in doing so normalize dominant knowledge. As such, "normalization is a form of power" (Kiersey and Neumann 2015, 75). The dominant knowledge is adopted into the everyday discourses and in popular culture of ordinary people's lives, and undergoes a process of internalization. Dominant knowledge in the everyday and popular culture are then constructive: "it is the result of a great number of decisions and selective incorporations of previous ideas, beliefs and images" (Studley 1998, 5).

But at the same time Studley (1998, 5) explains that knowledge also appears as "destructive of other possible frames of conceptualization and understanding." Although these literatures may appear as "the insignificant and the banal" (Lefebvre 1988, 78), they are in fact the "fertile soil" on which the "flowers or magnificent woods" of creative human activity grow (Lefebvre 1991, 87). Accordingly, they are capable of putting an end to the aforementioned normalization process of dominant knowledge by challenging what would otherwise be taken for granted (Kiersey and Neumann 2015, 75). This enables the blooming of new types of knowledges which go beyond dominant frameworks and engender a process of knowing *otherwise*, of knowing *creatively*, which unfolds when digging into sites of knowledge production that would traditionally be discarded. These sites are retrieved in both the everyday of "ordinary" people, but also in popular culture and can question predominant understandings. As such, artifacts or sites of popular culture in our everyday can redefine and transform what is so far considered as 'known' (Kiersey and Neumann 2015, 75) since they enable new creative ways of being, doing, and knowing otherwise.

This is best expressed in digital culture: "the unprecedented growth of digital technology and its embeddedness in the everyday lives of billions of people across the planet means that

6

digital technology now impacts PCWP [Popular Culture in World Politics] in multiple ways"
(Crilley 2021, 170). Popular culture has created new definitions of knowledge and its social
locations. Since the emergence of the Internet, these locations have been further extended, thus
expanding the boundaries of knowledge production to the digital space. This has led to everyday
practices juggling between offline and online modus, with the latter gaining more adherents
over the years. This trend continued with new technologies such as virtual reality, robots,
drones, social media platforms, or Artificial Intelligence (AI) serving as new sites of
knowledge. Hence, popular cultural manifestations are increasingly present in digitized media,
further extending the virtual playground and its joint, creative cultural production.

This subsequently opens the path for people to be, do, and know differently and in very
original ways, since new options for the public's expression and agency are offered, away from
traditional physical spaces, with online users now contributing to a broader discourse digitally.
Digital culture can reconfigure dominant knowledge present in the everyday and popular culture
since online users and creators apply individual constructions and interpretation, whether
through editing choices, censorship or modified representations. As such, the public has new
tools in hands to confront prevailing observations: individuals can now articulate their
understanding, frustrations or dissent towards dominant knowledge in digitally creative ways.
Consequently, digital tools like AI for instance fuel users' technological curiosity and grounds
digital popular culture even more. The digital world then appears as an enabler of new sites of
knowledge. To point this out is to highlight its significance as part of the everyday and popular
culture in the field of IR and how it opens up new paths of knowing *differently*. As summarized
by Schmid (2020, 698) "[i]f we see world politics and popular culture as co-constitutive and
images as a strong, omnipresent means to bridge the two, then the digital realm reveals itself to
be the battleground for the future of ideologies, identities, and values." Knowledge is then not

only produced in academia, but also in *digital* popular culture and in the everyday by "ordinary" people.

Conclusively, both literatures are the sites of dominant knowledge reproduction, but also the sites for creatively being, doing, and knowing otherwise since knowledge "that is to be made into popular culture must, then, contain both the forces of domination and the opportunities to speak against them, the opportunities to oppose or evade them from subordinated, but not totally disempowered, positions" (Fiske 2011, 21). This dialectical idea produces nomadic consumers and producers of knowledge who can move freely around this grid of dominant knowledge and knowing otherwise and realign their knowledge allegiances according to their individual values, ideas and beliefs.

1.2. *Foucauldian Counter-Conduct*

These aforementioned reformulations of knowledge "are made within a structure of power relations," since "all social allegiances have not only a sense of *with whom*, but also of *against whom*" (Fiske 2011, 20). Hence, it appears that the relationship of dominant knowledge and knowing 'otherwise' echoes the power-resistance binary, more specifically Foucault's counter-conduct dialectic with power (Foucault 2007, 194. Both literatures on popular culture in world politics and everyday IR have strong affinities with Foucault's idea of counter-conduct which is defined as *any* type of resistance and, "[a]s such, it politicises the everyday and locates politics 'everywhere'" (Demetriou 2016, 218). Counter-conducts then become part of the everyday politics of resistance, which can manifest itself in *everything*. Popular culture scholarship addresses and resists forms of power as well, especially through the transformation of our perceptions of resistance (Crilley 2021, 166). Counter-conduct then appears as a tool in this practice of creatively resisting and revealing itself in popular culture sites or artifacts.

The term counter-conduct emerged in Security, Territory, Population as a new concept to rethink the problem of resistance (Foucault 2007, 200). Other terms such as dissidence, revolt, disobedience and the like could not quite grasp what Foucault tried to pinpoint as "the struggle against the processes implemented for conducting others" (Foucault 2007, 201) since counter-conducts are "much more diffuse and subdued forms of resistance" (Foucault 2007, 200). The notion was used within the author's broader work on governmentality: counter-conduct is not to be equaled to a general rejection of the government, rather it provides for the perpetual questioning of the "preoccupation about the way to govern and the search for the ways to govern," the idea being "how not to be governed like that, by that, in the name of those principles, with such and such objective in mind" (Foucault 2007, 44).

Yet, some citizens do not get governed so easily. Indeed, while an individual could be conducted in a specific way, the latter has also the ability to conduct him/herself in a variety of ways (Lorenzini 2016, 7) because resistance is not the passive underside or the mere negative to a phenomenon (Foucault 1990, 195; 1994, 126–127). As opposed to a 'great Refusal' to being conducted, resistances are expressed in different more subtle ways, from "necessary, improbable, spontaneous," to "savage, solitary, concerted, rampant, violent, irreconcilable" (Foucault 1990, 95-96). It follows that the core of resistance practices lies not so much in opposing established power and power relations, but rather in its creative component that surpasses the mere reaction (Malvig 2016, 6). Counter-conduct manifests itself creatively beyond visible and politically discernible practices of resistance that enable the reproduction of dominant knowledge. Instead, it appears in less visible practices of dissent, in unexpected physical or digital places, but also among individuals who avoid public or academic registers. Hence, new spaces for practicing dissent open in "discursive ways, that is, by engendering a slow transformation of values" (Bleiker 2000, 276). Additionally, counter-conducts are creative, subtle forms of expressing resistance that enjoy an "ethos of novelty and

dangerousness" (Malmvig 2016, 274). The elements of novelty and risk lay in the fact that counter-conducts may form a new wave of reactions aimed at eliciting the rupture of sedimented and institutionalized forms of powers (Davidson 2011, 28-19).

By way of conclusion, knowledge is not only produced in academia, but also in *digital* popular culture and in the everyday by "ordinary" people. More importantly, popular culture and the everyday are not only the sites of dominant knowledge reproduction, but also the sites for creatively being, doing, and knowing otherwise which ties in with Foucault's concept of counter-conduct. Hence, this paper turns its attention to everyday online discussions of new AI technologies, specifically deepfakes, as a popular site for knowledge production and counter-conduct in the everyday and popular culture since both literatures are characterized by their potential in stressing the emancipatory, creative, resisting ideas and worldviews, that is the hidden potentialities that emerge in the "ordinary" person's daily life.

1.3. Deepfakes

Deepfakes are a new type of AI that enables the creation of pictures, audios or videos of people doing or saying things they never did, nor said. The deepfake looks as real as regular audio-visual footage but is in fact fake content manipulated by the latest AI technologies. These usually have a mocking, sexual or deceiving character and circulate freely on the web, especially on social media platforms and pornographic websites. Deepfakes employ "deep-learning technology, a branch of machine learning that applies neural net simulation to massive data sets" (Kerner and Risse 2021, 82). After extracting a sufficient amount of visual and audio data from person X, the latter's face and voice are superimposed onto person Y's body and face, thus giving life to an audio-visual that has been created from scratch without person X's involvement, and usually without his/her consent (Iacobucci et al. 2021, 194).

The first deepfake appeared in December 2017 on the social media website Reddit, when a user applied deep learning technique to create deepfakes of famous people in pornographic settings (Mirsky and Lee 2020, 1). It went viral as a result of its shocking, yet realistic character, but was quickly debunked and taken off the platform (Bohácek and Farid 2022, 1). However, the proliferation of computer-generated altered videos on social media ensued, with deepfakes being created, shared, commented, and reproduced online, and in particular on Reddit which became the Holy Grail of deepfakes (Spivak 2019, 339).

1.3.1. Mainstream AI-Deepfake Literature

AI does not receive extensive scholarly attention as a site of knowledge production. Instead, what researchers are interested in are the entailed risks of AI (Braunschweig and Ghallab 2021, 3). Some argue that AI is "opaque" and "prone to bias" due to its lack of transparency, explainability, verifiability, or security (Chatila *et al.* 2021, 13). Others have cautioned about AI's potential to "widen inequality, create a global underclass" or even to "launch 'new' arms races, and bring about bloodier wars" (Katz, 2020, 2). Also, its threat to diplomacy, politics and more specifically democracy is not denied: it is said that AI could possibly obstruct political and democratic procedures such as voting systems or elections for instance (Djeffal 2019, 260; See also Pantserev 2020; Chesney and Citron 2018; Diakopoulos and Johnson 2020).

The same caveats resonate in the mainstream literature on deepfake technology. Since their first appearance in 2017 and their continuous exponential increase (Sensity 2021, 85), they gradually gained more scholarly attention, although still limited. Screening the academic literature on deepfakes, Carjaval and Iliadis (2020, 3) observe that Humanities and Social Sciences predominantly inform users about deepfakes and their negative consequences, while a minority of the literature focuses on monitoring the production and detection of this new type of technology. This is especially true as deepfakes primarily target political figures and porn

(Gamage et al. 2022, 6). Based on these findings, deepfake literature can be divided into three strands: the dangers of political deepfakes, the risks of pornographic deepfakes, and the monitoring and detecting of deepfakes overall.

1.3.1.1. *Political Deepfakes*

The key feature of the first identified strand in the literature is the latter's focus on deepfakes' negative effects on politics. With titles such as "Deep fakes: a looming challenge for privacy, democracy, and national security" (Chesney and Citron 2019) ; "Protecting World Leaders Against Deep Fakes" (Agarwal et al. 2019) ; "The Malicious Use of Ai-Based Deepfake Technology as the New Threat to Psychological Security and Political Stability" (Pantserev 2020), a majority of scholars try to sound the alarm: deepfakes have detrimental effects on international affairs and politics.

It is argued that with the surge in synthetic media and deepfakes, videos are no longer trusted media, and this is especially unsettling for the fields of international relations, diplomacy, and politics (Ajder et al. 2019, 15 ; Kerner and Risse 2021, 99). Although some do admit that existing political deepfakes usually have a humorous purpose, cases appeared online and proved that deepfakes also aim for political deception (Giansiracusa 2021, 52). It is then asserted that deepfake instigators could be "politically-motivated actors and even nation-state attackers" (Kietzmann et al. 2020, 144). Multiple fears are expressed: that political deepfakes could worsen audiences' attitudes toward the depicted politician (Hancock, Bailenson 2021, 150), that they "threaten efficient governance for all democracies if not democracy itself" (Kietzmann et al. 2020, 143), or that they could engender political uprisings and military coups (Giansiracusa 2021, 56). Some authors like Chesney and Citron (2019, 1776) go as far as giving a detailed list of potential political harms such as "[a] fake audio clip [that] might 'reveal' criminal behavior by a candidate on the eve of an election," but these remain *possible*

scenarios with untested (negative) consequences. Ultimately, the overall sentiment is that deepfakes are "just the beginning of a new digital front that we might expect in this and future conflicts," as asserted by Bohácek and Farid who analyzed the 2022 Zelensky deepfake showing the Ukrainian President allegedly admitting defeat during the war (Bohácek and Farid 2022, 1).

1.3.1.2.   *Pornographic Deepfakes*

A second strand in the deepfake literature heavily condemns deepfakes for their pornographic and sexualized content and the subsequent evil consequences. Because pornographic deepfakes are extremely popular online with 96 percent of deepfakes containing pornographic content and targeting almost exclusively women (Kugler and Pace 2021, 613), a subset of scholars lay out the dangers this encompasses. Although research mainly displays the deepfake situation at stake, the argument that everyone can now make AI-generated fake pornography, including the creation of "obscene and child-pornographic videos," scares a lot of researchers (Spivak 2019, 329 and 358). As explained by Hancock and Bailenson (2021, 151), "[g]iven the power of the visual system in altering our beliefs already described, and the influence that such deepfakes can have on self-identity, the impact on a victim's life can be devastating" since "it is not difficult to imagine how deepfakes could be used to extort, humiliate, or harass victims".

Besides the more overall "dangers of automated tools for creating deepfake pornography" (Ajder et al. 2019, 8), more specific risks like pornographic deepfakes' instrumentalization to silence targets or to cause the victim "life-ruining" effects are highlighted (Maddocks 2020, 415). As such, pornographic deepfakes are first and foremost "a new form of sexual privacy invasion" (Meskys et al. 2020, 27).

1.3.1.3.   *Monitoring and Detecting Deepfakes*

Due to this AI recent character and exponential increase, this strand in the literature focuses on how one can monitor the production and the detection of AI-deepfake content (Chesney and Citron 2019; Lussier 2022; Yamaoka-Enkerlin 2020; Bohácek and Farid 2022). This is directly related to the said risks, threats and dangers of synthetic media.

From a legal perspective, the main problem expressed in the literature is that despite deepfakes' dangers and ethical implications, the law does not regulate this new phenomenon or, where it does, is incomplete or inadequate (Perot and Mostert 2020, 38; Meskys et al. 2020, 29). This stems from the fact that "the use of deepfakes does not fulfill the characteristics of any of the offences prohibited under most national laws" (Langer and Wyczil 2020, 207). As a result, this subset of the literature urges government and international organizations to act in order to prevent potential damage that will be generated by deepfakes, whatever its politically or sexually contentious content.

With the current lack of legislation to regulate synthetic media but also the enhanced sophistication and ease of use of deepfake-AI, it is important for scholars to put into light the ways in which audiences can detect fake content. The aim is to teach how one can identify manipulated audio-visual footage so as to avoid online users to fall into the deepfake trap (Akhtar and Dasggupta 2019; Bohácek and Farid 2022; Mirsky and Lee, 2020; Rana et al. 2022). Yet, the overall realization is that the pace at which deepfake technology develops spawns a "cat-and-mouse game between the creator and detector," thus leading to the portrayal of deepfake risks as imminent (Bohácek and Farid 2022, 2).

1.3.2. Gap in the Literature: The Insights of Popular Culture and the Everyday

Similarly to AI, deepfake technology mainly attracts negative academic spotlight as has been showcased in previous subsections. Scholars are especially active in researching the threats and adverse effects, rather than the opportunities offered by deepfake technology by primarily focusing on "recogniz[ing] deepfakes and mitigate their negative effects" (Carjaval and Iliadis, 2020, 3). There is a need to protect the population from deepfakes, an obligation to warn against its dangers and the urgency to prevent its detrimental effects through law and detection. Deepfakes are said to be threatening because they deceive and lie to the public, they obstruct and deconstruct discourses, they deform reality and challenge authenticity. Scholars then understand deepfakes as a rupture of normality in that they do not conform to commonly agreed values or norms, and deeply question our understanding of reality.

Yet, the mainstream academic argument that "deepfakes can be very dangerous" (Giansiracusa 2021, 63) and the tendency to depict them as *risky*, *challenging*, or *threatening* (Pantserev 2020; Boháček and Farid 2022; Chesney and Citron 2018; Diakopoulos and Johnson 2020; Giansiracusa 2021) is a very limiting frame to conceive of deepfakes. The three above-mentioned strands of deepfake literature have shared limits in that scholarly literature has not yet taken the insights of popular culture and the everyday onboard. It is worth noting that in these literatures creativity, resisting, and challenging established knowledge is foregrounded, even as it acknowledges a dialectical relation between resistance and inscription on relations of power. Still, deepfakes have not been recognized as a possible site in which power, ideology, identity and agency are constituted, performed, resisted, and challenged. They have not been looked at as artifacts for knowledge production, where one can creatively be and know otherwise. Hence, the current literature has so far not explored everyday knowledge production about deepfakes by "ordinary" people who regularly engage with them outside any academic or research context.

This thesis aims to contribute to this gap in the literature, more precisely by exploring whether, and if so to what extent, popular engagement with deepfakes exhibits the sorts of other thinking identified as subtle, creative, and/or resisting types of knowledge as unveiled in Chapter 1.

Chapter 2. Everyday Deepfake Knowledge: The Role of Reddit

In demonizing deepfakes, scholars seem to make generalizations without giving due attention to the communities where they bloomed. As explained in the previous chapter, this stems from a gap in the deepfake literature as the latter has not yet adopted the insights of everyday knowledge production on deepfakes by the everyday (wo)man even though knowledge production also takes place within non-academic communities. This thesis is concerned with exactly that: bridging this gap by uncovering alternative, everyday knowledge on deepfakes.

Because turning towards popular culture and the everyday in IR enables a better engagement with audience interpretation and uncovers new sites of knowledge production (Crilley 2021, 172), we have been looking for everyday sites of deepfake discussions. In that aspect, Reddit appears as the most appropriate site for popular knowledge on deepfakes.

2.1. Reddit as Locus for Deepfake Discussions

Reddit is a social media platform on which members can submit videos, images, posts, which are voted up or down by users. This online community is comprised of user-created forums termed "subreddits" (recognizable by the r/ before the title of the subreddit), and submissions with the most upvotes appearing at the top of the discussion thread. Diving into Reddit deepfake discourses is justified, first due to its distinctive traits as a social media platform and, second, because of the deepfake niche it contains.

Its singular architecture allows users a character limitation of 40.000, considerably greater than Meta (63.206 for posts, 8.000 for comments) or Twitter (280), thus encouraging more deliberative conversations (Treen et al. 2022, 682). In addition, Reddit is structured around themed topics (or subreddits) instead of being a follower-based social media platform which enables more visibility to whoever receives the most up-votes, while also highlighting those

comments who are down-voted. This gives the reader a more comprehensive idea of both the supported and unsupported reactions. Furthermore, "Reddit is limited to the user handle only (i.e. unlike Twitter, there is no user profile where details such as location, interests, political views or affiliation may be found)" (Treen et al. 2022, 682). This last point concurs with Reddit ensuring individuals a safe level of anonymity, enabling them to discuss more 'risky' or 'sensitive' topics which may not be discussed openly in other online public spaces where users could be identified (Gamage et al. 2022, 4). Hence, Reddit datasets have historically been the source of knowledge when dealing with more sensitive topics (See Maxwell et al. 2020; Sowles et al. 2018; Wang et al. 2015; Gamage et al. 2022). The platform enables a higher degree of transparency in user exchanges who are more eager to provide more intimate details which is unlikely on less anonymous social media like Meta or Twitter.

This could explain why deepfakes saw the light for the very first time on Reddit with a pornographic deepfake (Mirsky and Lee 2020, 1). The adaptation of the websites' terms and conditions following this incident did not curb deepfake creators' imagination since the variety in deepfake content has spread out beyond the pornographic world. Reddit has thus become the most adequate popular culture site for research since it "provides an incubator for developing deepfake generators/creators" with "an incremental distribution of posts and comments on deepfake each year" (Gamage et al. 2022, 9 and 11). Indeed, Reddit hosts users that create, share and discuss deepfakes in an effort to cultivate their identity and imaginations of the world that surrounds them. Hence, the platform is home to a community that is heavily aware and involved in the deepfake phenomenon which justifies even more the choice of discourse analysis of Reddit comments to shed light onto Redditors' deepfake knowledge production.

2.2. Unveiling Redditors' Role in Deepfake Knowledge

Unfortunately, academic literature on deepfakes, besides neglecting the insights offered by the the everyday and popular culture literatures, also barely targets Reddit forums. To fill in this dual gap, this thesis questions the forms of knowledge that Reddit users, or Redditors, produce about and with deepfakes, and how these relate to dominant scholarly knowledge. It does not only enable both literatures an entry to the world of deepfakes within the lives of the everyday (wo)men, but also targets Reddit as a locus for AI-deepfake debates and knowledge creation.

To do so, this empirical exploration is organized around the dialectics of power and resistance as being inextricably entangled, and any everyday knowledges that may be more or less on the side of established power (or dominant academic knowledge) or on the side of resistance (or knowing otherwise). This will give us a comprehensive overview of the deepfake knowledge spectrum, shedding light onto the dynamic relationship between dominant and alternative knowledge that Redditors (re)produce by identifying how the latter aligns with power structures or resistance types.

CHAPTER 3. METHODOLOGY

3.1.   Congruence Analysis and its Relevance to Deepfake Knowledge

Whereas congruence analysis can be used in both quantitative and qualitative research designs, in this thesis we opt for its qualitative variety. As explained by Mills, Durepos and Wiebe (2012, 1) "congruence analysis has a clear affinity for relativist epistemologies, which results in the conviction that empirical research cannot be used to verify or falsify theories but just to provide evidence for the relative strength of a theory by providing understandings and explanations." Put differently, congruence analysis is the use of "case studies to provide empirical evidence for the explanatory relevance or relative strength of one theoretical approach" (Blatter and Haverland 2012, 144). This is achieved by making hypotheses derived from theory which are subsequently tested for consistency with the empirical record (Mills, Durepos, Wiebe 2012, 2). As explained by Blatter and Haverland, "[t]his approach implies that theories lead to complementary implications in the real world," or in other words that it "provides the basis not only for more comprehensive explanations but also for conceptual and practical innovations" (Blatter and Haverland 2012, 145).

What we are interested in is probing the predilection of IR literatures, which tend to stress the emancipatory, resisting, creative, rupturing… potentialities of the everyday and popular culture. This is done by looking at AI and deepfakes, that is, in elation to technologies that have the power to enable 'other thinking' to be powerfully visualized and materialized (in the form of alternate worlds).

Congruence analysis then appears as the perfect fit for the current research purpose of analyzing the forms of knowledge produced by Redditors about and with deepfakes, and how these relate to dominant academic knowledge. It requires theoretically informed expectations "about the most important actors, their perceptions, and their motivations (traces of micro-foundational causal mechanisms)" (Mills, Durepos, Wiebe 2012, 3), which justifies the turn

towards Redditors. They are the main actors creating, interacting with and discussing deepfakes, and form as such a deepfake niche community. As will be clarified in the upcoming sections, based on the identified themes in Redditors' comments, two subreddit threads have been selected as case studies. Hence, this thesis does not generate a systematic analysis but a probe, which is justified by the dearth of the aforementioned literatures at the intersection of the everyday/popular culture and deepfakes (see Chapter 1).

It must be noted however that there seems to be a power imbalance between the Reddit community and other powerful knowledge regimes. Scholars, for instance, seem to possess superiority when it comes to knowledge: they are researchers or analysts, and are methodologically trained (Carver 2010, 425). In other words, they are "the academic knower[s]," or the naturally perceived experts in the acquisition of knowledge (Carver 2010, 421). And since "[p]ower produces and defines knowledge" (Stahl 2004, 4330) it is easy to understand scholars' upper hand in the deepfake discourse. Cheek explains that a consequence of power relations is that "not all discourses are afforded equal presence or equal authority" (Cheek 2008, 2). However, deepfakes as the object of study are inherent to the life world of Reddit users, whereas located outside of the life world of academics. Fortunately, using the theoretical lenses of the everyday within IR and popular culture in world politics "lays the pathway for more robust engagement with audience interpretation, affect, and emotion" (Crilley 2021, 172), and Redditors, or the audience of "ordinary" people as opposed to academics who hold the power of knowledge, then become the center of this research.

Employing congruence analysis sheds light onto popular culture and the everyday in IR as a framework that gives the relevant explanatory insights on deepfakes that no other theory has so far revealed. In the deepfake literature, scholars have not yet adopted the theoretical vision offered by the everyday and popular culture despite their potential in providing more thorough explanations as well as innovative knowledge not yet uncovered. By embracing those

literatures, we "are able to provide new or neglected explanatory insights" (Blatter and Haverland 2012, 145) and highlight both dominant but also marginalized knowledge within the micro-level deepfake realm. Congruence analysis is then consistent with this research's belief that knowledge is complemented and actively constructed by "ordinary" people and shaped by their experiences, perspectives, and interactions with their environment in their everyday.

As proposed by Blatter and Haverland, we will first deduce predictions from popular culture and the everyday, before "comparing these deduced expectations with empirical observations" (Blatter and Haverland 2012, 146) extracted from discourse analysis, understood as analysis of meaning making, of Reddit deepfake discussions (see Section 3.2). Subsequently, and as mirrored in the following hypotheses, we aim to explore to what extent one can identify innovative, challenging, creative, resisting… knowledge manifestations on deepfakes on Reddit, while staying alert of the dialectical idea that the everyday and popular culture are sites for both the reproduction of dominant knowledge and for creatively being, doing, and knowing otherwise (see Chapter 1).

Hypothesis 1. Redditors will reproduce dominant scholarly knowledge on deepfakes

Since Reddit is an important site of popular culture and the everyday, we hypothesize that this social media platform reproduces dominant knowledge on deepfake. As outlined in previous chapters, the dominant knowledge stems from academia which paints deepfakes as dangerous, risky and threatening. This rationale is expected to be present in Reddit discussions as well: a subset of Redditors are expected to reproduce dominant scholarly knowledge and, hence, vilify deepfakes in subreddit discussions because of its adoption in the everyday and popular culture of "ordinary" people's lives. The fear of AI deepfake is expected to have penetrated the online sphere with users envisioned to express identical worries as the scholarly

literature: the negative impact of political deepfakes, the detrimental effects of pornographic deepfakes and the subsequent need for monitoring and detecting such content.

Hypothesis 2. Redditors will produce deepfake knowledge otherwise, in creative ways, and away from dominant discourses

The everyday and popular culture as theoretical frameworks offer space for creatively being, doing, and knowing otherwise. Accordingly, we hypothesize that Reddit also enables users to creatively be, do, or know differently when it comes to deepfakes, away from established academic knowledge. This stems from the view that we are molded to think according to established standards: media outlets, politicians, academics and others push us to think in a certain way, that is by condemning deepfakes. Going against the flow bears the risk of being rejected, excluded, or simply neglected. As such, deepfakes appear as the perfect solution to shift away from dominant views creatively: the deepfake creator's own imaginary is artificially brought to life in original and innovative content that appears as real as the creator's clear-cut fantasy. Furthermore, AI's technological affordance—the easiness with which everyone can use and access emerging technologies, not to mention deepfakes' technological advances—coupled with Reddit's strong anonymity standards and users limited governmental oversight and discipline, leads to the expectation that Reddit deepfakes will engender an outbreak of creativity.

In the stitching together of different audio-visual possibilities, people are assumed to break free of traditional patterns which engenders a free flow of creativity and novelty, and ultimately knowledge. It follows that deepfakes may enable individuals to give birth to their wildest imaginaries which would typically be condemned under a regime of pre-established narratives. This type of AI then appears as the appropriate alternative to not only allow for more individual agency but also to enable people to display alternative realities in extremely creative

ways. Hence, the combination of AI-deepfake technology and human imagination and artistry are expected to open up new inspiring avenues, thus empowering "ordinary" people to be, do, and know in unexpectedly original ways, away from dominant scholarly knowledge and, as such, potentially reach a peak in creativity.

Hypothesis 3. Redditors will express counter-conducts that resist established deepfake discourses and further progressive ideas and critiques

This third hypothesis has strong ties with Foucault's notion of counter-conduct (Foucault 2007, 200) which presents affinities with both literatures on the everyday and popular culture. Designing and interacting with deepfakes is assumed to comply with the notion of counter-conduct as a subtle form of original, improbable, sometimes shocking practice of dissent that appears online and is created and shared. Hence, Redditors are hypothesized to express counter-conduct by picturing and discussing improbable scenarios that could be categorized as unethical or shocking. Redditors' shared thoughts may take the shape of resistance which appears at the microlevel (Death 2010, 238), that is among individuals who are deeply implicated in deepfake technology on social media, and who are therefore potentially active "in the transgression and contestation of societal norms; in the disruption of metanarratives of humanism; ... in the "re-appearance" of "local popular", "disqualified", and "subjugated knowledges"; and in the aesthetic of self-creation" (Kulynych 1997, 328). Redditors' comments and deepfakes are then a form of counter-conduct in that they strive to escape "involute rationalities and technes of conduct" (Odysseos, Death, and Malmvig 2016, 153) and resist dominant political, social, or ethical discourses. This can ultimately result in a new type of knowledge, what may be termed as counter-knowledge at the intersection of counter-conduct and creative knowledge resistance.

Following the belief that the digital, the political, the ethical and the cultural can be blended into one audio-visual and, ultimately, rally others to the cause (or the offered imaginary), deepfakes could be viewed by Redditors as a potent alternative to violence. Instead of attending rallies, hanging posters, or organizing street protests, citizens could resist official discourses creatively and digitally from home. They may become Redditors' tool to advance progressive ideas. Deepfakes, similarly to other types of visuals as argued by Mamlvig (2016, 262), may function as an original weapon or emancipatory language freely created and appropriated by the everyday (wo)man. Rather than entering the mold of pre-made narratives and thinking, individuals creating and interacting with deepfake content could have adhered to a novel way to resist. In this way, and as part of the assumption that they enable a peak in creativity, deepfake may be a counter-conduct tool in hands of the ordinary individual to originally resist dominant narratives, while furthering new and progressive agendas.

In other words, we hypothesize that some Redditors will apply a digitized and progressive form of Foucauldian counter-conduct to resist the dominant knowledge: amongst the Redditors expected to construct knowledge differently (hypothesis 2), a subset of them is anticipated to do so by adopting creative counter-knowledge.

3.2.   Deepfake Knowledge Production: A Discourse Analysis

Two case studies (see Section 3.2.1) provide the empirical foundation for the congruence analysis and, thus, serve as "evidence for the relative strength of a theory by providing understandings and explanations" (Mills, Durepos, Wiebe 2012, 2). Discourse analysis will be employed to analyze these case studies as discourses are "ways of thinking and speaking about aspects of reality" and structure this reality in specific ways (Cheek 2008, 2). This methodology enables the discovery of both dominant knowledge as well as knowing otherwise, which is compatible with congruence analysis since the latter "provides the basis not only for more

comprehensive explanations," that is dominant knowledge, "but also for conceptual and practical innovations," or new knowledge (Blatter and Haverland 2012, 145).

Dominant knowledge that underpins a discourse may be used to seclude other discursive framings or ways of understanding these same aspects of reality. For scholars, deepfakes are inherently bad and represent a rupture of normality in today's reality (see Section 1.2). This discourse prevails as a result of power and "exclude[s] the production of understandings and knowledge that could offer alternative views of that reality" (Cheek 2008, 3). But other discourses can expose power and render it fragile (Foucault 1976, 133) among others by exposing other ideas and individuals who do not hold the power (Wall, Stahl and Salam 2015, 261).

This thesis is concerned with exactly that. Derived from the assertion that knowledge is not only produced in academia but also outside of it, such as in digital popular culture, the aim is to display both the dominant and alternative views that form complementary knowledge about and with deepfakes. Based on the argument that Reddit, understood as an everyday and popular culture site, is home to the reproduction of dominant knowledge and also for creatively being, doing and knowing otherwise, discourse analysis enables the exploration of the mutually constitutive production of knowledge by displaying established *and* marginalized knowledge on and about deepfakes by the everyday person.

3.2.1. Case Studies

To identify the most relevant and insightful case studies, we first turn towards previous research on deepfake trends. Reddit enjoys a networked constellation of deepfakes, with topical categories of subreddits emerging on the platform. Gamage et al. examined deepfake-related conversations on Reddit between 2018 and 2021, and using a mixed-method approach they concluded that discussions are centered around two main topics: leaders, with Trump being the

favorite deepfake protagonist, and pornographic content (Gamage et al. 2022, 6). Although their research has not been extended to 2022 and 2023, it is assumed that these trends remain stable. Based on these results, this thesis divides deepfakes into two dimensions: leading figures on the one hand and porn on the other hand, although it must be noted that both may overlap.

Due to the overwhelming amount of deepfake content and limited space in this thesis, two subreddit posts are extracted for analysis, each of which identifies as one case study pertaining to one of the above dimensions so as to encompass the deepfake tendencies. Taken together, these case studies translate the AI technology's tendencies based on the deepfake trends analyzed by Gamage et al. (2022, 6). This is not to say that both cases serve as ultimate proofs nor are generalizable, rather they inform our understanding of deepfakes in ordinary people's lives in congruence with the said topical threads.

The first case, referred to as the Pope deepfake (attachment 1), will illustrate the first dimension of deepfakes on leaders. This deepfake picture of Pope Francis wearing a couture jacket gained worldwide attention and made headlines. Initially posted on Reddit on March 24th, 2022, the image of the Pope in a fancy coat was quickly shared across social media (Huang 2023). As the head of the Holy See and the Catholic Church's religious leader, the high-end streetwear outfit drastically departed from Popes' traditional clothes and his vow of poverty (Pell 2013). This dramatic shift in the pontiff's style instigated a wave of views, likes and comments online (Huang 2023). But the viral photo was fake and brought to life using *Midjourney*, an AI-software tool to create deepfakes based on short prompts introduced by users (Dolan 2023). Despite its incongruent character, the image "fooled the world" (Stokel-Walker 2023) with countless individuals believing it was real.

This first case study consists in the Reddit discussion of the said deepfake and was retrieved on a forum dedicated solely to international news discussions called r/news. This forum helped us target the most recent deepfake discussions of prominent leaders in today's

global landscape which acquired international fame. To generate interesting comments on deepfake reality, we were aiming for a post that gathered sufficient reactions. "AI image of Pope in a puffer jacket fooled the internet, and experts fear there's worse to come" published by Redditor KingArthursLance on March26[th], 2023, with its 43.5 thousand upvotes and 2.3 thousand comments by the time of writing, appeared as the most commented deepfake post on r/news (KingArthursLance 2023) that directly related to a (religious) leader. This deepfake discussion is also a very recent one and is therefore of interest since it will reflect the most recent reactions to this emerging AI technology.

The second case serves as a sample for the second dimension on pornographic deepfakes and is hereafter referred to as the porn deepfake (attachment 2). In contrast with the first case study, this Reddit discussion is sparked by a screenshot of an online user's strong opinion on pornographic deepfakes overall. Hence, and contrary to the first case study, it does not display a targeted sexual deepfake audio-visual but did spark a generic discussion on deepfake porn content based on a statement. As such, this case study enables the discussion of pornographic deepfakes *in general* (generic discussion on pornographic deepfakes) and, at the same time, of a *particular* deepfake (specific discussion on QTCinderella porn deepfake) because the post has been published shortly after and in link with a porn deepfake scandal in the online community.[1]

For years, pornographic and sexual deepfakes have been at the center of debates all over social media platforms: Twitter, 4chan, Meta, Reddit… Yet, thanks to the strong anonymity guarantees offered by Reddit and 4chan, one can witness fierce and unusual points of view regarding deepfake porn on these platforms. Because 4chan hosts "a discordant bricolage of humor, geek cultures, fierce debates, pornography, in–jokes, hyperbolic opinions and general

---

[1] A popular female gamer on the platform Twitch, QTCinderella, has been the subject of deepfake pornography in February 2023, a scandal that engendered a wave of comments amongst online communities, among which Twitter and Reddit (Court 2023).

offensiveness" (Knuttila 2011, 1), the subreddit r/4chan serves as the perfect combination of so-called AI content and pornography. It follows that r/4chan is the perfect site to identify interesting and atypical discourses on AI porn. When searching for recent posts using the key word "deepfake" on the r/4chan subreddit, the most commented post was the provocative "EMERGENCY ALERT: Hoes Mad" (Aphix 2023). It was published on February 2nd, 2023, and was a commented repost of a Tweet: "How is deepfake porn not illegal, that shit is basically virtual rape imo" (DreamLeaf5 2023). It comes as no surprise that the Tweet made its way to Reddit since Twitter does not give users extensive room for comment (see Section 2.1). Furthermore, due to the lack of anonymity, Redditors felt more comfortable to speak openly about deepfake pornography on Reddit. In total, the post gathered 1.4 thousand upvotes and 108 comments by the time of writing.

3.2.2. Exploring Discourse Analysis: Unveiling Knowledge (Re)production Processes

3.2.2.1. Reproducing Dominant Scholarly Knowledge: The Dangers of Deepfakes

Our first hypothesis stressed the expectation that dominant knowledge on deepfakes would be reproduced. Indeed, based on the empirical data a panoply of users in both case studies side with academia: the rupture of normality and dangers caused by deepfakes is taken as a fact.

"Well yeah, i think AI is dangerous" (u/jfduval76, March 27, 2023, 00:16 a.m., comment on KingArthursLance 2023), the overall Reddit argument goes. With deepfakes on the rise, "the problem is with the emergence of better ways to distort reality" (u/johnn48, March 27, 2023, 00:38 a.m., comment on KingArthursLance 2023) and "[s]adly it's gonna get a lot worse" (u/HTTRWarrior, February 2, 2023, 01:35 p.m., comment on Aphix 2023). Undeniably, the reproduction of established knowledge creates an online climate of fear among users: "Damn, I'm very worried about the future. I thought the past decade was nuts on disinformation, but we're headed to a whole new level now" (u/Tinfoilhatmaker, March 27, 2023, 04:58 a.m.,

comment on KingArthursLance 2023). As such, "Deepfake technology is already good but when it gets GOOD that's when the real problems will start" (u/yeetskeetrepeat420, February 2, 2023, 10:19 a.m., comment on Aphix 2023). In the case of the Pope deepfake, for instance, the most popular comment was "I admit I fell for it" (u/potatodog247 March 26, 2023, 08:09 p.m., comment on KingArthursLance 2023) which gathered 16.7k upvotes. In other words, almost 17.000 users confess to having been fooled by the AI-generated image and this generates fears among Redditors.

It stems from the potential of AI to make ordinary people believe fake audio-visuals to be real. While prior to deepfakes "[v]ideo was, in a way, the easiest way for people to verify potentially doctored images/events. [...] Now even the matching of voices and facial expressions is getting scary good and they have Obama and even Abraham Lincoln talking. Those already vulnerable to misinformation are going to be eating up all the falsehoods without question" (u/TheReasonsWhy, March 27, 2023, 05:05 a.m., comment on KingArthursLance 2023). As a result, a number of comments express fear and anxiety regarding the dangerous *potential* of deepfakes, thus aligning with scholars on the threats it poses. Although deepfakes are said to not seem harmful at first, because "[t]here's nothing "controversial" about the image, so I think that makes it much more believe" (u/PolicyWonka March 27, 2023, 01:00 a.m., comment on KingArthursLance 2023), an important amount of users do agree that "the slippery slope is what's dangerous here. A gradual increase in this kind of thing is going to be a boiling frog scenario" (u/Reborn_Rhubarb, March 26, 2023, 8:23 p.m., comment on KingArthursLance 2023). As beautifully put by one user, the "danger" then becomes "the truth manipulations on our peripheries" (u/Millionthvisitor, March 27, 2023, 00:09 a.m., comment on KingArthursLance 2023).

In the same vein as academics, Redditors general concerns over deepfake technology rapidly shift to problems specific to the political and pornographic worlds. Deepfakes are

perceived by some users as transcending political boundaries. "I could see this kind of thing becoming a major problem in politics very soon" the overall comment goes (u/Uncreativite, March 27, 2023, 02:19 a.m., comment on KingArthursLance 2023), with specific situations given as examples such as "political organizations using deepfake photos and video that 'they thought were real, whoops'" (u/Blazinnie, March 27, 2023, 01:17 p.m., comment on KingArthursLance 2023). Furthermore, "deepfake could possibly be used to further political agendas" (u/rainyfort1, February 4, 2023, 12:21 p.m., comment on Aphix 2023) or "a malicious actor [could] convince a few million to vote for the next Hitler or to just not vote for the candidate you like" (u/mimocha, March 27, 2023, 05:49 a.m., comment on KingArthursLance 2023).

Regarding the intersection of deepfake politics and porn, one user expresses anxiety about how "[t] his will be used to create fake porn of female politicians to try to get them to resign" (u/SandboxOnRails, March 27, 2023, 05:34 p.m., comment on KingArthursLance 2023), or more generally the creation of "revenge porn against women" (SandboxOnRails, March 27, 2023, 07:31 a.m., comment on KingArthursLance 2023) since deepfakes will enable "to make porn of anyone you want pretty soon" (u/themariokarters, March 27, 2023, 03:52 a.m., comment on KingArthursLance 2023). For Reddit users, sexual deepfakes are seen as a rupture with normality in that it can lead to severe problems and cause harm, especially for women. These users follow the path of academia, in that porn deepfakes are labeled as "[m]aybe not rape but absolutely fucking unethical and vile" (u/yaboy_jesse, February 2, 2023, comment on Aphix 2023). Indeed, a minority of users on the porn deepfake discussion thread do argue that this "should be illegal and wouldn't want it done to me" (u/SibrenTF, February 3, 2023, 04:41 p.m., comment on Aphix 2023).

3.2.2.2. Producing Knowledge Otherwise: Deepfakes are Normal

According to our second hypothesis, Redditors produce different knowledge about and with deepfakes, a subset of which consists of counter-conduct, or what we term counter-knowledge. When comparing this prediction with the data from the two case studies, knowledge otherwise could erupt with deepfakes *in general* and deepfakes *in particular*. This is why the following section will be divided into two parts: the one exposing the general but non-dominant knowledge, and the other one uncovering particular critiques understood as counter-knowledge.

(1) Deepfakes: An Extension of Normality

(1.1) "It's not like this is a new thing"

When analyzing Reddit deepfake discourses, one common, yet unexpected argument resurfaced in both case studies: deepfakes? "[I]t's not like this is a new thing" (u/detahramet, March 27, 2023, 00:15 a.m., comment on KingArthursLance 2023). For a minority of Redditors deepfakes should not be labeled a new phenomenon. Equalled to photoshop, revenge porn, hoaxes, parodies… they are said to simply be a reproduction of what was already possible in the past. "[I]t's not all thaaaaat revolutionary" one user commented when discussing the AI generated image of Pope Francis, "it's not like we haven't had photoshop for the past few decades. I could make this exact same pic in photoshop in under 30 minutes. Less if I had great reference material like a high res[olution] version of the jacket" (u/hparamore, March 27, 2023, 01:35 p.m., comment on KingArthursLance 2023). His point gathered a not insignificant 918 upvotes by the time of writing and was backed up by others. For instance, siding with the "normality" argument, user detahramet explains that "Photoshop has existed for decades, and hoaxes have existed long before that" (u/detahramet, March 27, 2023, 00:15 a.m., comment on KingArthursLance 2023).

When it comes to deepfake pornography, users followed the same path by establishing parallels with other practices: "[i]t's more akin to revenge porn" (u/Dr_barfenstein, February

2nd, 2023, 10:52 a.m., comment on Aphix 2023), and equaled to "parodies" (u/MostWholesomeReditor, February 2nd, 2023, 11:28 a.m., comment on Aphix 2023) or even "cartoon caricatures" and "satire" (u/Background_Cycle7676, February 2, 2023, 01:08 p.m., comment on Aphix 2023). As such, we "don't NEED sophisticated advanced AI algorithms to generate false imagery" (u/SandboxOnRails, March 27, 2023, 07:31 a.m., comment on KingArthursLance 2023) because we already had pre-existing audio-visual alteration systems that long prevailed deepfakes. AI-deepfake porn is understood as something normal in that it does not diverge from previously used tools like revenge porn, Photoshop, parodies, or cartoons, to name just a few.

Accordingly, no surprise or shock is expressed whatsoever: "I have absolutely no idea why is this a hot topic right now. We already figured this out, for decades you could take someones face and put it on the body of someone" (u/Tulee, February 2, 2023, 02:51 p.m., comment on Aphix 2023). Deepfakes are perceived as a prolongation of what society was already used to in terms of practices and technology. In contrast to the established scholarly knowledge, this sample of Redditors expresses indifference around deepfake technology, as if society was undergoing a process of deepfake normalization. First, because "Reddit has been a breeding ground for ai for ages now" and second because "[p]eople are freaking out about stuff that, in a lot of ways, already existed" (u/matlynar, March 27th, 2023, 02:10 a.m., comment on KingArthursLance 2023). For these users who produce different versions of deepfake understandings, it goes without saying that no monitoring or specific detection tools should be created to "protect" ordinary people. Because deepfakes form no danger and are equaled to what we already had at our disposal, there is no one to 'protect'. Thus, and as opposed to dominant knowledge, no monitoring or detection system should be put in place.

It follows that since deepfakes are no different from older technological advances to which societies have continuously adapted, the same will happen with deepfakes. "All that's

left to capture with fakery is the youth and they are growing up with AI and social media filters on the mind. They will grow up knowing the tell tale signs and asking the right questions about videos that we may be too old to catch onto" (u/Deducton, March 27th, 2023, 02:35 a.m., comment on KingArthursLance 2023). If the element of fakery in deepfakes appears as a problem for scholars and the majority of Redditors, this is contrasted with a minority who attests that the population will simply accommodate to the problematics it poses like it has been the case with previous innovations. "[T]his is just us being old people now and not knowing how to adapt to the next big thing. Imagine the first photoshop, the first fake newspaper website, all the scam websites, etc. we saw a lot of it and started filtering through the bullshit. The same will continue to happen and life will move on" ([deleted user], March 27th, 2023, 07:53 a.m., comment on KingArthursLance 2023). Deepfakes are implicitly categorized as the "next big thing" but immediately downgraded. In the end, they are just another version of 'photoshop', 'fake news', or 'scam website', to which we got used to. Pursuant to this trend, deepfakes then form the next step towards a small adaptation in our everyday lives. Dominant narratives about deepfakes' dangerousness must be watered down as deepfakes are not a rupture of normality, but rather a continuation of it.

However, some argue that although this might be true to some extent, "[t]he difference is that you can have a bot spitting these out. Photoshop still requires human artistry" (PickledPlumPlot, March 27, 2023, 02:21 a.m., comment on KingArthursLance 2023). This subset of Redditors brings more nuance in their judgment, explaining that "[t]hese faked images, while they've been around for sometime with photoshops, can now be produced near effortlessly with little effort. And I imagine they will subtly influence many people who see them" (LeanTangerine, March 27, 2023, 02:31 a.m., comment on KingArthursLance 2023). Indeed, "[t]he thing is, now anyone can do this with zero effort or skill" (laika_rocket, March 27, 2023, 02:17 a.m., comment on KingArthursLance 2023). The parallel drawn with

Photoshop is endorsed, but the strength of AI deepfake technology is also highlighted when mentioning the easiness with which everyone can produce fake imagery nowadays. By way of deduction, deepfakes are perceived by this group of users as the upgraded version of existing practices such as Photoshop or parodies where one can combine AI with human creativity. While this technology does show new traits compared to previous tools, deepfakes can't be labeled as an innovative breakthrough that shakes normality. There is no mention of this AI to be dangerous, risky, negative, but the users do concede that it might have some subtle influence on audiences, although not explicitly detrimental.

(1.2) "It's not artistic or creative it's just functional"

The "normality" argument around deepfakes resonates with the next finding: the expected outburst of imagination supposedly rendered possible by AI deepfake technology isn't as ambitious or innovative as one might expect. This is not only confirmed by a subset of Redditors who, as mentioned, believe deepfakes to be part of our normality, but also by the lack of creativity in the Reddit discussion threads. Hence, the assumption that the combination of new AI-deepfake technology and human imagination would take creativity to a whole new level does not correspond with the retrieved data.

Initially, it was argued that with deepfakes, the worlds of politics, pornography, the private and public spheres, could converge with the "ordinary" person's every day, framed by humor, drama or sex, the idea being to capture (or rather produce) an untouched juxtaposition. It followed that individuals are encouraged to step out of their comfort zone creatively speaking, and that by engaging with AI-tools unexpected deepfake fantasies and discussions of it would become the end product. Deepfakes would then appear to transcend boundaries so as to integrate online users' wishes to it, whatever shocking, atypical, sexual, comic, or dangerous those may be. Gradually, the everyday (wo)man would adapt to it, thus generating new deepfake content to satisfy his/her imaginary by progressively pushing the boundaries of the

35

unthinkable to an unknown limit. Based on this, it was predicted that the everyday (wo)man would form new types of knowledge, of being, of doing in extremely creative ways as a result of the AI at hand. Deepfakes would have appeared as a tool, a weapon in the hands of the everyday person to create the extraordinary in terms of knowing, being and doing.

However, empirical data in both case studies failed to back up this hypothesis. As put by one user, AI-deepfake is "not artistic or creative it's just functional" (u/Pabus_Alt, March 27, 2023, 03:46 p.m., comment on KingArthursLance 2023). Instead of creatively being, doing, knowing otherwise, Redditors did not express new ideas. Rather than discovering innovative forms of knowledge through deepfakes, we are confronted with Redditors who, likewise, also wonder about the lack of originality: "I'd be curious if this could be replicated with something relatively unignorable" one user said when commenting on the absence of creative elements in the Pope deepfake (u/WhiteWolf3117, March 26, 2023, 11:59 p.m., comment on KingArthursLance 2023). On the other hand, others express faith in AI-generated utopias, without proposing any themselves, but simply wondering about it like Redditor ecksboy who "can't wait to watch new episodes of Star Trek TNG" created by the AI deepfake machinery (u/ecksboy, March 27, 2023, 04:03 p.m., comment on KingArthursLance 2023). Deepfakes do not seem to put into motion an outburst in creativity, rather it generates a sort of creativity laziness within users. Instead of producing their own creative ideas using deepfakes, Redditors express a preference for consuming AI-generated content: they look forward to fantasies produced by AI in lieu of their own imaginations supported by AI. This suggests that the impact of deepfake technology on creativity may has been overestimated in our hypothesis.

(2) Deepfakes: A New Practice of Counter-Conduct

Because the emergence of deepfake technology has brought about both opportunities and challenges in terms of how individuals can resist or perpetuate dominant knowledge, we predict to find counter-conducts, or what may be labeled counter-knowledge within Reddit deepfake

discussion threads. In other words, deepfake technology makes it possible for the "ordinary" person to resist dominant knowledge in more original ways compared to traditional forms of resistance. Redditors are thought to be enabled to resist knowledge creatively, which is understood as articulating ideas and assumptions that may be socially distant from established ones in order to further progressive ideas.

(2.1) Regressive Counter-Knowledge

Interestingly, the opposite takes place in pornographic deepfake subreddits: instead of furthering progressive views, an important number of users seems to advance regressive ideas and further exploit women's bodies. The most popular comment retrieved from the porn deepfake case already sets the stage: "[women are] finding out their only value is being stolen from them" (u/Digitalispurpureae, February 2, 2023, 08:24 a.m., comment on Aphix 2023). Users typically comment on deepfakes overtaking women's role in pornographic clips, thus stripping women of their agency and reducing them to objects of sexual desire.

By minimizing women to sexual objects, Redditors resist standards of equality and ethics. Indeed, besides condemning the dominant narrative of deepfakes as being dangerous, the counter-opinion is that the fear that derives from dominant knowledge merely stems from "fear that [women's] simps will stop supporting [their] lifestyle once they realize they're less than 3 clicks away from nudifying anyone they fancy" (u/kiraYoahikage, February 2, 2023, 05:55 p.m., comment on Aphix 2023). This point of view is repeated by others like user MTGBruhs who mocks "women thinking onlyfans was gonna be their meal ticket," yet they are "forever getting replaced by an AI that does it faster better and cheaper" (February 2, 2023, 09:17 a.m., comment on Aphix 2023). Others go even further: besides rejecting the common idea that deepfakes are a new and dangerous phenomenon as a result of technological advances, deepfakes are normal and women are to blame for pornographic deepfakes: "[i]f only there weren't 109 trillion hours of high definition naked-lady content self published by women" the

argument goes (u/CodAway6491, February 2, 2023, 08:35 p.m., comment on Aphix 2023). The deepfake "scandal" erupting from it serves them: "Maybe now w*men will finally learn a thing or two about digital privacy and not being a narcissistic self absorbed cunt fishing for attention like some garden tools" (u/hedroks, February 3, 2023, 03:29 a.m., comment on Aphix 2023). Hence, women are not the victims of non-consensual pornographic content, but the instigators of sexual deepfakes.

Additionally, whereas the dominant knowledge assumes deepfakes to be a threat to all, some Redditors creatively resist this trend: women are the only ones feeling threatened by it. They are the ones bearing the (financial) consequences of it since they are the ones "lamenting the loss of pooosyy power," that is they are losing the possibility to sell the objectification of their bodies since sexual deepfakes already do the job (u/Tommy-AllArk-io, February 2, 2023, 08:42 a.m., comment on Aphix 2023).

Those Redditors reject standards of equality and ethics, refuse women's battles against their objectification and sexualization, and ignore their autonomy and agency. These attitudes point toward the will not to be governed in a certain way by those who hold those dominant values and impose politically and ethically correct vocabularies. By expressing counter-knowledge, these Redditors aim to crack sedimented and institutionalized forms of established knowledge. But whereas counter-conducts typically cause fissures in favor of the improvement of our social, political, or ethical lives, the current comments are a type of counter-conduct that alters those. Knowledge can be resisted from very different positions and sexism would be one way to counter dominant understandings and discourses. The displayed discourses are a form of counter-knowledge, but one which promotes the adoption of retrogressive lenses. Instead of aiming for a move forward, Redditors sexist counter-knowledge mirror a move backwards towards values and ideas that are nowadays perceived as ethically wrong and part of the past. The porn case study then shows how (pornographic) deepfakes bring to light ordinary people's

most basic instincts within the Reddit bubble. Sexism, gendered assumptions and women's commodification color the discussions: pornographic deepfakes' female protagonists are not human subjects nor victims, but sexual objects threatened by deepfakes due to the female body's monetary devaluation in the face of AI.

(2.2) Progressive Counter-Knowledge

Besides regressive counter-knowledge, one can also decipher progressive forms of counter-conducts. This is especially true with the Pope deepfake and subsequent Reddit discussions. Although those critiques are not particularly new, Redditors do criticize the Pope and religion using creative means: deepfakes and humor. The Pope deepfake speaks for itself: dressing the religious leader in a high-end couture coat can be perceived as a critique of the very foundations of his position and the religious organization he represents. When clothing him that way, the contradiction between Pope Francis' vow of poverty and his own material wealth is highlighted. Such an image could be interpreted as a commentary on the state of the religious institution itself, and the challenges it faces in reconciling its lofty ideals with the practical realities of the modern world. That is also how a group of Redditors reacted to it in the comments' section: "I just thought he was on his way to Swiss to get his money from some shady bank" (u/ PiSsOUtMYASs-, March 27, 2023, 03:18 a.m., comment on KingArthursLance 2023), or "I did have this nagging feeling of 'well, he is rich as fuck'" (u/IPeedOnTrumpAMA, March 27, 2023, 05:36 a.m., comment on KingArthursLance 2023). This deepfake image and comments are a subtle, yet original ways to critique the papal institution, and to challenge others in considering the complexities of religion, wealth and power.

The well-known criticism on pedophilia within the church has also been brought to the forefront of the deepfake discussions. Pursuant to this, users made use of the deepfaked image to further emphasize these existing controversies, as seen in comments such as "obviously that's

how he smuggles his chattel of kids" (u/flaker111, March 27, 2023, 00:40 a.m., comment on KingArthursLance 2023), "New album called 'F--k the Kids'" (u/Soft-Intern-7608, March 27, 2023, 03:26 a.m., comment on KingArthursLance 2023), or "Well, white does hide the stains that are common in this profession" (u/Thetrav1sty, March 27, 2023, 01:33 a.m., comment on KingArthursLance 2023).

These statements show how the use of imagery and cultural artifacts can be instrumentalized as a form of counter-knowledge to critique a powerful institution. By combining the critique of the Pope's luxurious lifestyle with the ongoing criticism of the church's handling of money and child abuse scandals, these users are engaging in a transparent exposure of the state of the religious institution and the challenges it faces in reconciling its ideals with the realities it faces. Through these comments, we are reminded of the power of creative (yet harsh) language and imagery to resist official knowledge, effect positive change and further progressive ideas that challenge structures of inequality.

3.2.3. Conclusion: Summary of Findings

Because knowledge is also constructed within "ordinary" people's everyday, combining congruence analysis with discourse analysis enabled us to show how popular culture and the everyday in IR lead to complementary and comprehensive knowledge on deepfakes. Through two deepfake case studies retrieved on Reddit we have displayed how online users navigate the deepfake "knowledge spectrum," but more importantly the surprising limits of creativity and the dual agendas of progressive and regressive counter-knowledge produced within Reddit threads.

As hypothesized, whereas some Redditors reproduce dominant scholarly knowledge and attest deepfakes to be a danger, others create knowledge differently. Interestingly, while the prediction of deepfakes engendering an outbreak of creativity was wrong, this "knowing

otherwise" did take an unexpected turn: as opposed to categorizing deepfakes as a danger and a rupture of normality, some users normalized the technology, explaining it was in fact an extension of normality. These dual understandings create a knowledge spectrum within which users move freely, balancing more towards one or the other end of the grid in shades of more or less strong knowledge allegiances to either dominant or alternative knowledge, and in the absence of a creativity peak.

Although counter-conducts were hypothesized to promote forward-looking critiques and ideas within deepfake Reddit communities, the technology has created a complex landscape for resistance by enabling *both* regressive and progressive forms of what we term counter-knowledge. Among those who construct knowledge otherwise, thought-provoking counter-knowledges have been retrieved with some expressing regressive ideas through the commodification of women, and others progressive (although less innovative) ideas regarding the religious establishment and papacy. As such, deepfakes serve the dual goal of being a tool of resistance understood as Foucauldian counter-conducts, by not only constructing ideas, values and worldviews, but also by deconstructing them.

3.3. *Limitations and further research avenues*

3.3.1. Limitations

Congruence and discourse analysis have enabled us to dig deeper into Reddit as an everyday popular site for knowledge production on deepfakes in exploratory ways. However, these approaches present some limitations. First, employing discourse analysis brings with it the complexity in drawing fixed conclusions about the prevalence of certain understandings and attitudes regarding deepfakes. While basing one's reasoning on the analysis of Redditors' language use does provide insights into the production of everyday popular knowledge about this emerging AI, it may conceal other beliefs. Similarly, congruence analysis and the fact that

we have constrained ourselves to two case studies as representative samples of deepfake categories bears the risk of not capturing the variability in deepfake cases, and thus miss important nuances in users' attitudes towards the technology. Accordingly, when researching the everyday person's understanding of deepfakes, there may be differences in the ways that Redditors produce knowledge about and with deepfakes according to the specific subreddit. Unfortunately, the overwhelming amount of deepfake subreddits makes it almost impossible to analyze all discourses around each of them and the limited space capacity of this thesis restricted the possibility of additional case studies. Furthermore, both types of analyses rely on subjective interpretation of data, hence the difficulty of drawing definite conclusions on how Reddit users produce deepfake knowledge, and may also leave room for bias or misinterpretation.

3.3.2. Exploratory Research

It must be noted that this study serves a preliminary purpose, that of discovering new understandings of deepfakes in *exploratory* ways. The results reflect the thoughts and sense-making of Reddit users in their discussions of the pope and the porn deepfakes, which means that the main findings cannot be generalized but can serve as a first probe and set a precedent for further research on deepfake knowledge production in the everyday and popular culture spheres.

Reddit serves a niche community (Kuehl, 2019, 3) that is especially interested in deepfake content. Since this gives us a narrower version of the public's conversations on the topic, there are a myriad of other possible future directions for research. Prominent implications that may not have appeared on Reddit feeds might be found on other social media platforms such as Twitter, Meta, Instagram, and others. Despite lower levels of anonymity and more limited writing space, such platforms are also part of the public's AI-playground and might capture different types of conversations related to topics which are less prominent on Reddit, thus

uncovering other deepfake knowledges and their (un)relatedness to dominant academic discourses.

Chapter 4. Conclusion

Deepfakes have emerged as a prominent technology in the AI landscape, yet the prevailing understanding, produced in academia and echoed in media outlets, solely focuses on their dangerous character. This dominant knowledge has so far remained unchallenged, and popular communities within which deepfake AI bloomed have not yet been scrutinized. To point this out is to highlight the gap in the deepfake literature which tends to overlook the implications of both popular culture and the everyday in IR. By employing theoretical foundations rooted in the everyday and popular culture, this research illuminates their significance for deepfake knowledge production in the lives of ordinary people.

This thesis demonstrates that Reddit provides a unique platform for open and transparent deepfake discussions and contributes to a comprehensive understanding of this emerging AI. Using a qualitative methods approach through the use of congruence and discourse analysis, we do not only observe the reflection of dominant knowledge among Reddit users but also uncover alternative forms of knowledge that emerge outside the confines of academia. the unexpected limited scope of creativity, resisting and counter-conduct knowledge. Most importantly, the empirical contribution of this thesis is the unexpected limited scope of creativity, resisting and counter-conduct knowledge.

The research hypotheses supported by empirical data suggest that Redditors contribute to the reproduction of dominant scholarly knowledge, but also generate novel knowledge on deepfakes. With Redditors navigating a spectrum of knowledge on deepfakes and opening up the path towards knowing *otherwise*, the surprising limits of creativity have been unveiled. While the anticipated originality outbreak was not confirmed in any of the two case studies, a distinct understanding emerged where deepfakes were normalized and considered an extension of normality rather than a departure from it. This new knowledge production also involved the exploration of counter-conduct, what has been coined as counter-knowledge, within deepfake
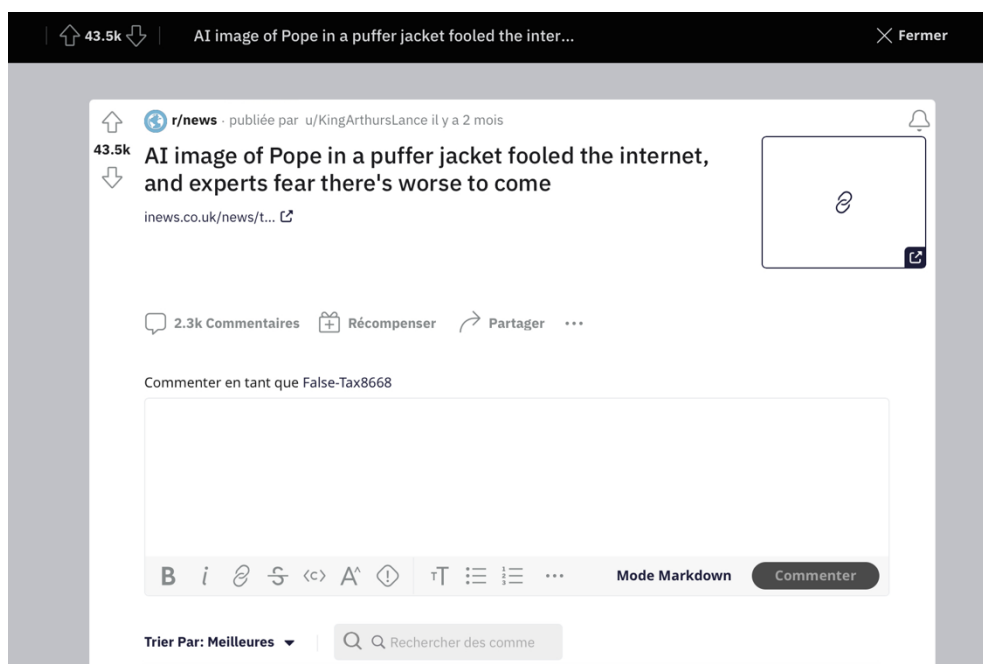
communities that reveal a complex landscape of resistance. The hypothesis that deepfakes would serve as a progressive tool was only partially true, in that Redditors made use of it for *both* constructing and deconstructing new ideas, values, and worldviews, hence furthering not only progressive, but also regressive agendas through the expression of counter-knowledge.

Although the employed theoretical lenses provided some insights into new knowledge, their limited extent calls for further exploration. Rather than aiming for generalized assumptions, this thesis serves as an initial exploration of deepfake knowledge (re)production by ordinary individuals, leveraging the theoretical framework of the everyday and popular culture. This is not to challenge existing knowledge on deepfakes, but to complement it. Consequently, the study's methodology and theoretical lenses lay the foundation for future research to delve deeper into the understanding of deepfakes within everyday contexts and the potential for alternative knowledge production.

While Reddit has emerged as a primary forum for deepfake (re)production, it is essential to explore other popular culture sites to obtain a comprehensive understanding of deepfakes. Social media platforms such as Meta or Twitter could offer new paths into alternative deepfake knowledge production. Scholars are therefore encouraged to engage in critical reflection of deepfake literature, and to bridge the prevailing gap by incorporating insights from popular culture and everyday experiences. This research points towards the need to move beyond the prevailing understanding of deepfakes as solely dangerous, recognizing the significance of popular culture and everyday experiences in contributing to knowledge production. By shedding light on the diverse knowledge spectrum among Redditors and their engagement in resistance, this thesis proposes new avenues for further exploration and encourages future researches to delve into popular culture sites for deepfake knowledge (re)production.

Appendices

Attachment 1. The Pope Deepfake—AI-generated image of Pope Francis wearing a couture jacket and ensuing Reddit discussion thread.

Attachment 2. The Porn Deepfake—Reddit discussion thread on deepfake pornography.

Reference List

Ajder, Henry; Patrini, Giorgio; Cavalli, Francesco and Cullen, Laurence. 2019. *The State of Deepfakes 2019: Landscape, Threats, and Impact*. Amsterdam: Deeptrace Lab.

Akhtar, Zahid and Dasgupta, Dipankar. 2019. "A Comparative Evaluation of Local Feature Descriptors for DeepFakes Detection." *Conference: IEEE Symposium on Technologies for Homeland Security*, 1-5.

Aphix (u/Aphix). 2023. "EMERGENCY ALERT: Hoes Mad." *Reddit*, February 2nd, 2023. https://www.reddit.com/r/4chan/comments/10rjhe7/emergency_alert_hoes_mad/.

Benve, Andrea. 2022. "Angela Merkel Shows Her Boobs. Fake." *Mr.Deepfakes,* May 2023. https://mrdeepfakes.com/video/23459/not-angela-merkel.

Boháček, Matyáš and Farid, Hany. 2022. "Protecting World Leaders against Deep Fakes Using Facial, Gestural, and Vocal Mannerisms." *Proceedings of the National Academy of Sciences of the United States of America* 119 (48): 1-3.

Boháček, Matyáš and Farid, Hany. 2022. "Protecting President Zelensky Against Deep Fakes." *arXiv*: 1-8.

Blatter, Joachim and Haverland, Markus. 2012. *Designing Case Studies. Research Methods Series*. London, UK : Palgrave Macmillan.

Bleiker, Roland. 2000. *Popular Dissent, Human Agency and Global Politics*. Cambridge, UK: Cambridge University Press.

Bos, Daniel. 2018. "Answering the Call of Duty: Everyday Encounters with the Popular Geopolitics of Military-Themed Videogames." *Political Geography* 63: 54–64.

Braunschweig, Bertrand and Ghallab, Malik. 2021. "Reflections on AI for Humanity: Introduction." In *Reflections on Artificial Intelligence for Humanity*, edited by Bertrand Braunschweig and Malik Ghallab, 1-12. Cham, Switzerland: Springer.

Byman, Daniel L.; Gao, Chongyang; Meserole, Chris; Subrahmanian, V.S. 2023. "Deepfakes and International Conflict." *Foreign Policy at Brookings*, 1-22.

Carver, Terrell F. 2010. "Cinematic Ontologies and Viewer Epistemologies: Knowing International Politics as Moving Images." *Global Society* 24 (3): 421–31.

Chatila, Raja *et al.* 2021. "Trustworthy AI." In *Reflections on Artificial Intelligence for Humanity*, edited by Bertrand Braunschweig and Malik Ghallab, 13-39. Cham, Switzerland: Springer.

Cheek, Julianne. 2008. "Foucauldian Discourse Analysis." In *The SAGE Encyclopedia of Qualitative Research Methods*, edited by Lisa M. Given, 1-5. Thousand Oaks, CA: SAGE Publications.

Chesney, Robert and Citron, Danielle. 2019. "Deepfakes and the New Disinformation War The Coming Age of Post-Truth Geopolitics." *Foreign affairs* 1 (98): 147-155.

Chesney, Robert and Citron, Danielle. 2018. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review*, 1735-1820.

Court, Andrew. 2023. "Twitch star QTCinderella's deepfake porn nightmare: 'F–k the internet'." *New York Post*, February 6, 2023. https://nypost.com/2023/02/06/twitch-star-tearfully-reveals-shes-victim-of-deepfake-porn-f-k-the-internet/.

Crilley, Rhys. 2021. "Where We At? New Directions for Research on Popular Culture and World Politics." International Studies Review 23: 164-180.

Da Silva, Raquel, and Crilley, Rhys. 2017. "'Talk About Terror in Our Back Gardens': An Analysis of Online Comments About British Foreign Fighters in Syria." *Critical Studies on Terrorism* 10 (1): 162–86.

Davidson, Arnold. 2011. "In Praise of Counter-Conduct." *History of the Human Sciences* 24 (4): 25–41.

Death, Carl. 2010. "Counter-conducts: A Foucauldian Analytics of Protest." *Social Movement Studies*, 9 (3): 235-251.

Demetriou, Olga. 2016. "Counter-conduct and the everyday: anthropological engagements with philosophy." *Global society*, 30 (2): 218-237.

Diakopoulos, Nicholas and Johnson, Deborah G. 2020. "Anticipating and Addressing the Ethical Implications of Deepfakes in the Context of Elections." *New Media & Society* 23 (7): 2072–2098.

Dixit, Priya. 2012. "Relating to Difference: Aliens and Alienness in Doctor Who and International Relations." *International Studies Perspectives* 13 (3): 289–306.

Djeffal, Christian. 2019. "AI, Democracy and the Law." In *The Democratization of Artificial Intelligence*, edited by Andreas Sudmann, 255-284. Wetzlar: Majuskel Medienproduktion.

Dolan, Leah. 2023. "Look of the Week: What Pope Francis' AI puffer coat says about the future of fashion." *CNN*, March 29, 2023. https://edition.cnn.com/style/article/pope-francis-puffer-coat-ai-fashion-lotw/index.html.

DreamLeaf5 (@DreamLeaf5). 2023. "How is deepfake porn not illegal, that shit is basically virtual rape imo." Twitter, January 30, 2023. https://twitter.com/DreamLeaf5/status/1620187104457408512.

Easthope, Antony. 1991. *Literary into Cultural Studies*. London, UK: Routledge.

Fiske, John. 2011. *Understanding Popular Culture.* London, UK: Routledge.

Foucault, Michel. 2007. *Security, Territory, Population 1977-1978*. Translated by Graham Burchell. Basingstoke: Palgrave Macmillan.

Foucault, Michel. 1994. *La Volonté de Savoir*. Paris: Gallimard.

Foucault, Michel. 1990. *The history of sexuality, vol. 1 : An introduction.* Translated by Robert Hurley. New York, USA : Vintage.

Foucault, Michel. 1976. *Histoire de la sexualité I: La volonté de savoir*. Paris: Gallimard.

Gamage, Dilrukshi et al., 2022. "Are Deepfakes Concerning? Analyzing Conversations of Deepfakes on Reddit and Exploring Societal Implications." *ArXiv*, 1-19.

Giansiracusa, Noah. 2021. "Deepfake Deception." In *How Algorithms Create and Prevent Fake News*, edited by Noah Giansiracusa, 41-66. Berkeley, CA: Apress.

Grayson, Kyle, Matt Davies, And Simon Philpott. 2009. "Pop Goes IR? Researching the Popular Culture : World Politics Continuum." *Politics* 29 (3): 155–63.

Hancock, Jeffrey T. and Bailenson, Jeremy N. 2021. "The Social Impact of Deepfakes." *Cyberpsychology, Behavior, and Social Networking*, Vol. 24 (3): 149-152.

Hasselstein, Ulla ; Ostendorf, Berndt ; Schneck, Peter. 2001. "American Studies." *Popular Culture* 46 (3): 331-338.

Huang, Kalley. 2023. "Why Pope Francis Is the Star of A.I.-Generated Photos." *New York Times*, April 8, 2023. https://www.nytimes.com/2023/04/08/technology/ai-photos-pope-francis.html.

Iacobucci, Serena; De Cicco, Roberta; Michetti, Francesca; Palumbo, Riccardo and Pagliaro, Stefano. 2021. "Deepfakes Unmasked: The Effects of Information Priming and Bullshit Receptivity on Deepfake Recognition and Sharing Intention." *Cyberpsychology, Behavior, and Social Networking* 3 (24): 194-202.

Innes, Alexandria J., and Topinka, Robert J. 2016. "The Politics of a 'Poncy Pillowcase': Migration and Borders in Coronation Street." *Politics* 37 (3): 273–87.

Katz, Yarden. 2020. *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. New York, USA: Columbia University Press.

Kerner, Catherine and Risse, Mathias. 2021. "Beyond Porn and Discreditation: Epistemic Promises and Perils of Deepfake Technology in Digital Lifeworlds." *Moral Philosophy and Politics*, Vol. 8 (1): 81- 108.

Kietzmann, Jan; Lee, Linda W.; McCarthy, Ian P. and Kietzmann, Tim C. 2020. "Deepfakes: Trick or treat?," *Business Horizons* 63, 135-146.

Kiersey, J. Nicholas and Neumann, Iver B. 2015. "Worlds of Our Making in Science Fiction And International Relations." In *Popular Culture and World Politics : Theories,*

Methods, Pedagogies edited by Federica Caso and Caitlin Hamilton, 74-82. Bristol, UK : E-International Relations.

KingArthursLance (u/KingArthursLance). 2023. "AI image of Pope in a puffer jacket fooled the internet, and experts fear there's worse to come." *Reddit*, March 26th, 2023, 8:03 p.m. https://www.reddit.com/r/news/comments/122ua4t/ai_image_of_pope_in_a_puffer_jacket_fooled_the/.

Knuttila, Lee. 2011. "User unknown: 4chan, anonymity and contingency." *First Monday* 16 (10): pp. n/a. https://firstmonday.org/ojs/index.php/fm/article/view/3665/3055.

Kugler, Matthew B. and Pace, Carly. 2021. "Deepfake Privacy: Attitudes and Regulation." *Northwestern University Law Review* 116 (3): 611-680.

Kulynych, Jessica J. 1997. "Performing Politics: Foucault, Habermas, and Postmodern Participation." *Polity* 30 (2): 315-346.

Langer, Maciej and Wyczik, Jakub. 2020. "Legal Aspects of Deepfakes." *Annuals of The Administration and Law* 20 (4): 205-212.

Lefebvre, Henri. 1991. *Critique of Everyday Life*. Vol. 1. London: Verso.

Lefebvre, Henri. 1988. "Toward a Leftist Cultural Politics: Remarks Occasioned by the Centenary of Marx's Death." In *Marxism and the Interpretation of Culture*, edited by Cary Nelson and Lawrence Grossberg, 75-88. London: Macmillan.

Lorenzini, Dianiele. 2016. "From Counter-Conduct to Critical Attitude : Michel Foucault and the Art of Not Being Governed Quite So Much." *Foucault Studies* 21, 7-21.

Lussier, Natalie. 2022. "Nonconsensual deepfakes: detecting and regulating nonconsensual deepfakes: detecting and regulating the rising threat to privacy the rising threat to privacy." *Idaho Law Review* 58 (2): 352-381.

Maddocks, Sophie. 2020. " 'A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political' deep fakes." *Porn Studies* Vol. 7 (4): 415-423.

Malmvig, Helle. 2016. "Eyes wide shut: Power and creative visual counter-conducts in the battle for Syria, 2011–2014." *Global Society* 30 (2): 258–278.

Maxwell, December; Robinson, Sarah R.; Williams, Jessica R. and Keaton, Craig. 2020. "A Short Story of a Lonely Guy: A Qualitative Thematic Analysis of Involuntary Celibacy Using Reddit." *Sexuality & Culture* 24 (6): 1852–1874.

Meskys, Edvinas ; Liaudanskas, Aidas ; Kalpokiene, Julija ; Jurcys, Paulius. 2020. "Regulating Deep Fakes: Legal and Ethical Considerations." *Journal of Intellectual Property Law & Practice* 15 (1): 24–31.

Mills, Albert J.; Durepos, Gabrielle and Wiebe, Elden. 2012. "Congruence Analysis." In *Encyclopedia of Case Study Research* edited by Albert J. Mills, Gabrielle Durepos and Elden Wiebe, 1-7. Thousand Oaks, CA: Sage.

Mirsky, Yisroel and Lee, Wenke. 2020. "The Creation and Detection of Deepfakes: A Survey." *ACM Computer Surveys* 1 (January): 1-38.

Odysseos, Louiza, Death, Carl and Malmvig, Helle. 2016. "Interrogating Michel Foucault's Counter-Conduct:Theorising the Subjects and Practices of Resistance InGlobal Politics." *Global Society* 30 (2): 151–56.

Pantserev, Konstantin A. 2020. "The Malicious Use of Ai-Based Deepfake Technology as the New Threat to Psychological Security and Political Stability." *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity*, 37–55.

Pears, Louise. 2016. "Ask the Audience: Television, Security and Homeland." *Critical Studies on Terrorism* 9 (1): 76–96.

Pell, Carolina. 2013. "Pope Francis takes poverty vow to heart as a true Christian." *The Daily Telegraph*, June 30, 2013. https://www.dailytelegraph.com.au/news/opinion/pope-takes-poverty-vow-to-heart-as-a-true-christian/news-story/42a939531c873d50df7b1f38844b12e9.

Pusca, Anca. 2015. "Representing Romani Gypsies and Travelers: Performing Identity from Early Photography to Reality Television." *International Studies Perspectives* 16 (3): 327–44.

Rana, Sohel; Nur Robi, Mohammad; Murali, Beddhu and Sung, Andrew H. 2022. "Deepfake Detection: A Systematic Literature Review." *IEEE* 10: 25494-25513.

Rowley, Christina. 2015. "Popular culture and the politics of the visual." In *Gender Matters in Global Politics: A Feminist Introduction to International Relations*, edited by Laura J. Shepherd, 361–374. London and New York: Routledge.

Sachleben, Mark. 2014. *World Politics on Screen: Understanding International Relations through Popular Culture*. Kentucky, USA: The University Press of Kentucky.

Schmid, Julian. 2020. "(Captain) America in crisis: popular digital culture and the negotiation of Americanness." *Cambridge Review of International Affairs* 33 (5): 690-712.

Sellman, Mark. "Emma Watson reads Mein Kampf on 4Chan in deepfake audio trick." *The Times*, January 31, 2023, https://www.thetimes.co.uk/article/ai-4chan-emma-watson-mein-kampf-elevenlabs-9wghsmt9c.

Sensity Team. 2021. "How to Detect a Deepfake Online: Image Forensics and Analysis of Deepfake Videos." *Sensity AI* (August). https://sensity.ai/blog/deepfake-detection/how-to-detect-a-deepfake/.

Sowles, Shaina J.; McLeary, Monique;, Optican, Allison; Cahn, Elizabeth; Krauss, Melissa J.; Fitzsimmons-Craft, Ellen E.; Willey, Denise E. and Cavazos-Rehg, Patricia A. 2018.

"A content analysis of an online pro-eating disorder community on Reddit." *Body image* 24: 137–144.

Spivak, Russel. 2019. ""Deepfakes": The Newest Way To Commit One of The Oldest Crimes." *Georgetown Law Technology Review* 3 (2): 339-400.

Stahl, Bernd C. 2004. "Whose Discourse? A Comparison of the Foucauldian and Habermasian Concepts of Discourse in Critical IS Research." *Conference paper for the Proceedings of the Tenth Americas Conference on Information Systems*, New York (August): 4329-4336.

Stivale, Carles J. 2002. "Mythologies Revisited: Roland Barthes and the Left." *Cultural Studies* 16 (3): 457-484.

Stokel-Walker, Chris. 2023. "We Spoke To The Guy Who Created The Viral AI Image Of The Pope That Fooled The World." *BuzzFeed News*, March 27, 2023. https://www.buzzfeednews.com/article/chrisstokelwalker/pope-puffy-jacket-ai-midjourney-image-creator-interview.

Storey, John. 1993. *An Introductory Guide to Cultural Theory and Popular Culture*. London, UK: Harvester Wheatsheaf.

Studley, John. 1998. "Dominant knowledge systems and local knowledge." *The Mountain Institute* (April-May): 1-25.

Treen, Kathie; Williams, Hywel; O'Neill Saffron and Coan, Travis G. 2022. "Discussion of Climate Change on Reddit: Polarized Discourse or Deliberative Debate?" *Environmental Communication*, Vol. 16 (5): 680-698.

Wall, Jeffrey D.; Stahl, Bernd Carsten and Salam, A.F. 2015. "Critical Discourse Analysis as a Review Methodology: An Empirical Example." *Communications of the Association for Information Systems*, vol. 37: 257-285.

Wang, Lei; Zhan, Yongcheng; Li, Qiudan; Zeng Daniel D.; Leischow, Scott J. and Okamoto, Janet. 2015. "An examination of electronic cigarette content on social media: analysis of e-cigarette flavor content on Reddit." *International journal of environmental research and public health* 12 (11): 14916–14935.

Warner, William B. 1990. "The Resistance to Popular Culture," *American Literary History* 2 (4): 726–42.

Wedderburn, Alister. 2019. "Cartooning the Camp: Aesthetic Interruption and the Limits of Political Possibility." *Millennium* 47 (2): 169–89.

Weldes, Julia and Rowley, Christina. 2015. "So, How Does Popular Culture Relate to World Politics?" *E-IR Popular Culture and World Politics*: 1-16. https://www.e-ir.info/pdf/55781.

Yamaoka-Enkerlin, Anna. 2020. "Disrupting Disinformation: Deepfakes and The Law." *New York University Journal of Legislation & Public Policy* 22 (3): 725-749.