Distribution Agreement

Alexa Mohsenzadeh March 17, 2023

Social Media as Neurotechnology: A Case Study of Censorship and Propaganda on Twitter

By

Alexa Mohsenzadeh

Gillian Hue, Ph.D.

Adviser

Neuroscience and Behavioral Biology

Gillian Hue, Ph.D.

Adviser

Hossein Samei, Ph.D.

Committee Member

Mark Risjord, Ph.D.

Committee Member

2023

Social Media as Neurotechnology: A Case Study of Censorship and Propaganda on Twitter

By

Alexa Mohsenzadeh

Gillian Hue, Ph.D.

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Neuroscience and Behavioral Biology

2023

Abstract

Social Media as Neurotechnology: A Case Study of Censorship and Propaganda on Twitter
By Alexa Mohsenzadeh

Neurotechnology is an emerging subset of technology that involves the direct manipulation and/or recording of neural activity. This study presents evidence that social media platforms can access and influence neural activity. Therefore, social media platforms ought to be classified as neurotechnology and subject to the same ethical considerations, especially in cases where social media are manipulated for political purposes to control people's beliefs and restrict the free flow of information. These principles are applied to the case of the usage of Twitter in Iran by government affiliated accounts since the start of the protests in September 2022. The ethical, legal, and societal implications of the government's control of social media and spread of digital propaganda are assessed. The use of social media as the primary outlet for communication and information sharing through the protests in Iran is an example of how people's conceptions, belief systems, and behaviors can be tied to social media. Thus, censoring and filtering these platforms to influence people's thoughts is both a human rights issue and a neuroethical concern that must be addressed using existing guidelines for regulating neurotechnology.

Social Media as Neurotechnology: A Case Study of Censorship and Propaganda on Twitter

By

Alexa Mohsenzadeh

Gillian Hue, Ph.D.

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Neuroscience and Behavioral Biology

2023

Table of Contents

Acknowledgements

I would like to extend a special thank you to my adviser, Dr. Gillian Hue, for her continued mentorship, wisdom, and humor throughout this process. She instilled in me a deep appreciation for neuroethics and she gave me the space to push the bounds of a neuroscience honors thesis project. I am immensely grateful for her support.

I would also like to thank my committee members, Dr. Hossein Samei and Dr. Mark Risjord, for their support and advice over the course of this project. I appreciate Dr. Samei for his willingness to meet with me so often this year and for his unwavering kindness as a mentor since my first year at Emory. Thank you to Dr. Risjord for providing such valuable feedback on the rigor of this work through each phase of the project and for guiding me in the early stages of ideation.

Lastly, I would like to thank my family and friends for their continued support this year. I am lucky and thankful to be surrounded by such inspiring people.

Social Media as Neurotechnology

Introduction and Significance

What is Neurotechnology?

According to the Institute of Electrical and Electronics Engineers (IEEE) Brain Initiative, neurotechnology is defined as any technology that provides greater insight into brain or nervous system activity, or affects brain or nervous system function. In the 28th session of the UN Human Rights Council Advisory Committee called to assess the human rights implications of neurotechnology, neurotechnologies were defined as “any electronic device, method or process conceived to access the human brain’s neuronal activity”, the use of which poses a threat to “the ability of individuals to govern freely their own behaviour.” In their draft report on the Ethical Issues of Neurotechnology, the International Bioethics Committee (IBC) cites the Organisation for Economic Co-operation and Development’s (OECD) definition of neurotechnology as "the field of devices and procedures used to access, monitor, investigate, assess, manipulate, and/or emulate the structure and function of the neural system of animals or human beings”. In addition to citing OECD’s standardized definition of neurotechnology, the IBC also classifies neurotechnology as any device/application that fundamentally “influences how people understand the brain and various aspects of consciousness, thought, and higher order activities in the brain,” (UNESCO IBC, 2020).

Under these definitions, technologies that involve direct manipulation and/or recording of neural activity are clear qualifiers for neurotechnology. Deep brain stimulation, Brain Computer Interfaces (BCIs), artificial intelligence in clinical neuroscience, and the use of neuroscience for market research are typically at the forefront of neuroethical conversations regarding their use and regulation (Müller and Rotter, 2017). Nearly all of these forms of technology that we recognize as "neurotechnology" are specifically developed and used to interact with the brain. For technologies that are developed and used for other purposes, their ability to access, influence, and/or manipulate neural activity is less clear but no less important.

Social media is a prime example of a technology that has not yet been identified as a neurotechnology in previous literature but *should* be. The purpose of this paper is to fill this gap in the literature on the classification of neurotechnology, as social media can access and influence neural activity. Therefore, social media platforms ought to be viewed as neurotechnology and subject to the same ethical and legal considerations.

Social Media Defined

The term "social media" refers to online platforms that allow for interactions between users and the creation of virtual communities by sharing information (Tufts University Relations, 2022). Instagram, Facebook, Youtube, Twitter, LinkedIn, and Snapchat are all social media platforms. Carr and Hayes (2015) define social media as "Internet-based channels that allow users to opportunistically interact and selectively self-present, either in real-time or asynchronously, with both broad and narrow audiences who derive value from user-generated content and the perception of interaction with others." Importantly, this definition establishes a distinction between social media and other media platforms, like online news services, streaming platforms, and videoconferencing applications.

Media vs. Social Media

In their definition of social media, Carr and Hayes (2015) identify 4 characteristics that make social media distinct from other forms of media, including its perceived interactivity, persistent channels, user-generated value, and capacity for mass-personal communication. In the paragraphs that follow, I will describe and build upon Carr and Hayes framework. I also propose that the additional characteristics of **addictivity** and **algorithmic personalization** make social media distinct from other forms of media and social manipulation.

First, perceived interactivity refers to the user's perception of interpersonal interaction via their engagement with the social platform. This form of interactivity on social media applications is not necessarily equivalent to interpersonal interaction in real life, but the "social" nature of these platforms, e.g. liking posts, retweeting, watching stories, and sending direct messages (DMs), keeps users engaged (Carr and Hayes, 2015). Unlike traditional media platforms that are typically limited to passive viewing, social media allows people to engage with individuals far beyond the scope of their social network, including public figures, celebrities, and government officials. This can lead to the formation of parasocial relationships between the user and public figures (Carr and Hayes, 2015; Lueck, 2015; Hoffner and Bond, 2022; Chung and Cho, 2017). Chung and Cho (2017) found that parasocial relationships formed via social media increase the user's sense of intimacy and connectedness. Most importantly, users report higher levels of source trustworthiness and credibility for public figures with whom they have formed a parasocial relationship. The latter finding will be revisited in the next section in relation to propaganda on social media, specific to the case of Iranian government officials.

Persistent channels refers to the continuous availability and activity of social media networks, regardless of whether the user is engaged with the application or not. Unlike platforms like Zoom, Netflix, or Gmail, the activity of social media platforms is not dependent on the engagement of an individual user at a single point in time. The ubiquity of these platforms on a mass-scale allows for the perpetual stream of new content that a single user can interact with at any point. The ability to have interpersonal interactions on a platform without temporal restrictions is known as “channel disentrainment” (Carr, 2017; Walther, 1996). Social media platforms are disentrained channels because they allow for asynchronous communication between users. This asynchronicity also grants users more time to interpret and respond to messages in a way that is in accordance with the identity they have curated online (Carr, 2017).

The third distinction identified by Carr and Hayes between social media and other forms of media is its user-generated value. On traditional media platforms, the value is typically produced by the host(s) of the platform. For example, when a user opens Apple News, they consume content that is deemed significant by journalists and news agencies. By contrast, social media platforms offer various mechanisms for users to flag valuable content and promote its recirculation. Posts with millions of likes and retweets appear to be more relevant than a post with minimal engagement, and level of engagement is a signifier of value on social media. Therefore, it is the users of social media platforms, not the curators, that derive and assign value.

Lastly, social media users have the flexibility to interact with each other on an individual level or on a mass scale, a.k.a. “masspersonal communication” (O'Sullivan and Carr, 2018). On a platform like Facebook or Twitter, users have the option of broadcasting information to the public through posts, likes, and comments or engage more privately via direct messages. Carr and Hayes identify 4 modes of information transmission that can occur on social media, including user to user, user to audience, audience to user, or audience to audience. The user’s ability to create content for an audience just as much as they consume is a distinct capability that is not possible on other media platforms.

In addition to the characteristics proposed by Carr and Hayes, this study proposes the addition of addictivity and algorithmic personalization which makes social media distinct from other media sources (Ricci, 2018). As explored in the next section, social media platforms are intentionally designed to hold the user's attention for as long as possible and hijack the brain's reward processes through a rapid supply of new and engaging content. The randomization of reward and the release of dopamine that occurs with each notification makes social media more addictive than other forms of media and social influences. There is also a highly personal element to the user's experience of social media, as the self-learning algorithm responds to the preferences of the user and can personalize the user's experience over time (Kozyreva et al., 2021). Like other acknowledged forms of neurotechnology, the use of social media can threaten the privacy and the personal agency of the user, altering the structure of the nervous system over time via addictive usage (Goering, 2021).

Social Media and the Brain

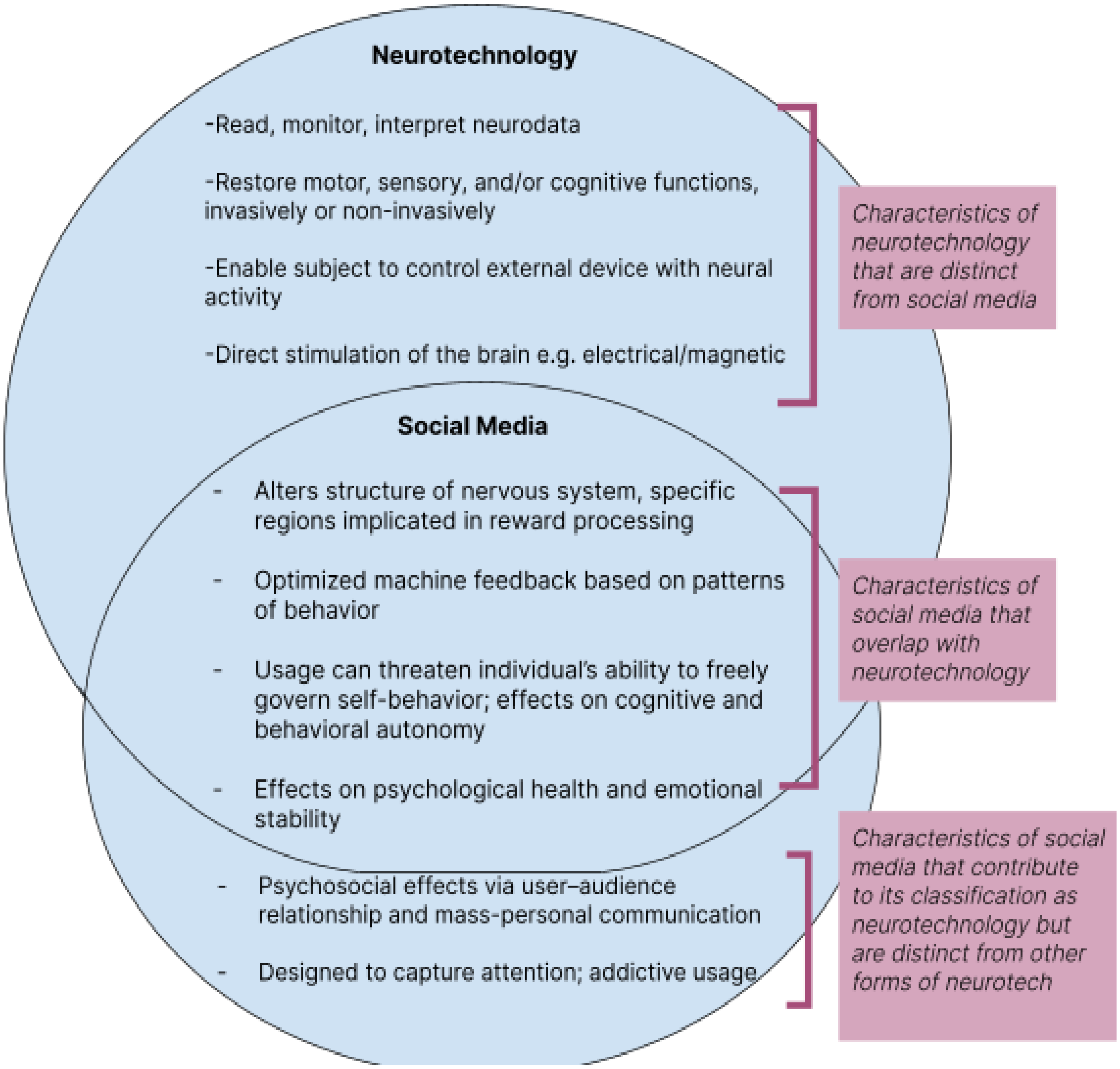

Neurotechnology
-Read, monitor, interpret neurodata
-Restore motor, sensory, and/or cognitive functions, invasively or non-invasively
-Enable subject to control external device with neural activity
-Direct stimulation of the brain e.g. electrical/magnetic
Social Media
- Alters structure of nervous system, specific regions implicated in reward processing
- Optimized machine feedback based on patterns of behavior
- Usage can threaten individual's ability to freely govern self-behavior; effects on cognitive and behavioral autonomy
- Effects on psychological health and emotional stability
- Psychosocial effects via user–audience relationship and mass-personal communication
- Designed to capture attention; addictive usage
Characteristics of neurotechnology that are distinct from social media
Characteristics of social media that overlap with neurotechnology
Characteristics of social media that contribute to its classification as neurotechnology but are distinct from other forms of neurotech

Fig. 1. Features of social media that contribute to its classification as neurotechnology and distinguish it from other forms of neurotechnology.

Attention and Addiction

Social media platforms like Twitter and Instagram are designed to hold users attention for significant periods of time by suggesting stimulating, personalized content. Whereas before information was a scarce resource, the advent of the internet and the emergence of social media has allowed for an abundance of information at the expense of our attention. As Herbert Simon states, “What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it,” (Variam, 1996). Williams (2018) argues that the risk of attention scarcity is not that our attention is completely occupied by information, but that our attentional processes are fundamentally altered such that our capacities for self-regulation and filtration are diminished. Social media platforms strategically provide us with endless informational rewards, resulting in repetitive usage and, in many cases, addiction.

The growing concerns regarding social media usage, addiction, and mental health are not unfounded. On average, people spend 2 hours and 27 minutes on social media (Kemp, 2022). In 2018, the World Health Organization (WHO) coined social media addiction as a growing issue that warrants serious consideration. In 2022, Tristan Harris, co-founder and president of the Center for Humane Technology, argued that we ought to treat social media addiction as a public health emergency (Center for Humane Technology). Increased time on social media has been correlated with lower psychological well-being, including “lower self-control, more distractibility, more difficulty making friends, less emotional stability, being more difficult to care for, and inability to finish tasks,” (Twenge and Campbell, 2018).

In a UK-wide survey of 1,479 people between the ages of 14–24, the Royal Society for Public Health found a correlation between social media usage and mental health issues, including increased anxiety and depression, poor sleep quality, and feelings of loneliness (RSPH, 2017). The COVID-19 pandemic has exacerbated these issues, as Facebook, Twitter, Instagram, and TikTok became the primary form of communication for many people under quarantine (Fullerton, 2021). Statistics on social media show a significant uptick in social media usage in

2022, with 4.62 billion social media users around the world (Kemp, 2022). Social media users account for 58.4% of the total global population, which is 10.1% more than in 2021 (Kemp, 2022).

Platforms like Twitter, Facebook, and Youtube can fundamentally shape our psychology and identity by encouraging habit formation and eventual user addiction, and this is intentional by design. Former Facebook president, Sean Parker, described this strategy in a 2017 interview, stating, “The thought process that went into building these applications, Facebook being the first of them, ... was all about: 'How do we consume as much of your time and conscious attention as possible?'” (Allen, 2017; Fisher, 2022). Social media’s command of our attention works hand-in-hand with the “social-validation feedback loop” and the randomized release of dopamine that promotes addiction, Parker explains. The randomization of reward, aka “intermittent variable reinforcement”, is a key feature of social media platforms that makes them particularly addictive (Fisher, 2022). Intermittent variable reinforcement is what makes slot machines so addictive, and it is not a coincidence that platforms like Twitter engage similar strategies to keep the user on the platform for as long as possible (Fisher, 2022). The simple act of “refreshing” your newsfeed on Twitter is akin to pulling the lever on a slot machine with hopes that you’ll eventually be rewarded with money, or in Twitter’s case, a salient notification.

Cognitive Salience and Reward

Previous research by Meshi et al. (2015) shows that the use of social media engages the reward network in the brain, activating brain regions including the ventral striatum, ventral tegmental area, and the ventromedial prefrontal cortex. Platforms such as Twitter offer consistent rewards via notifications, likes, messages, retweets, and mentions. Meshi et al. liken these reward-triggers on social media to forms of prosocial behavior that elicit positive feedback from others and activate the reward system, such as giving a compliment or engaging in acts of service. These social stimuli on social media positively reinforce our behavior on social media can elicit a surge of dopamine by activating our mesolimbic dopaminergic system, leading to a cycle of endless wanting and dissatisfaction that contributes to social media addiction (Haynes, 2018).

Prior research has also shown that social media usage can impact gray matter volume in the brain, specifically in regions of the brain associated with reward processing. Montag et al. found that individuals with higher daily frequencies of checking Facebook had decreased gray matter volumes in the left and right nucleus accumbens (Montag et al., 2017). The nucleus accumbens of the ventral striatum is implicated in motivational and reward processing (Salgado and Kaplitt, 2015). The rewarding nature of social media notifications increases an individual's daily usage and thereby affects neural reactivity of the nucleus accumbens.

Using social media can evoke an affective state that is marked by high positive valence and arousal, contributing to our extended engagement on these platforms (Mauri et al., 2011). Valence is the pleasantness or unpleasantness of an emotional stimulus and arousal refers to the intensity of the emotion as measured by the level of autonomic activation that occurs in response to a stimulus (Kauschke et al., 2019; Bestelmeyer et al., 2017).

High valence and arousal has an effect on time perception (Van Volkinburg and Balsam, 2014). The frequent and varied notifications on social media also activate the "salience network", including the anterior insula and dorsal anterior cingulate cortex (Center for Humane Tech). The salience network has been identified as a series of brain regions that are implicated in the process of identifying and responding to relevant stimuli (Seeley, 2019). The constant input of stimuli from social media often leads to multitasking and repetitive behavior that impacts our cognitive control, weakening levels of activation in the prefrontal cognitive control network which is implicated in impulse-control, attention, and working memory (Center for Humane Tech).

These effects are amplified in cases of media multitasking i.e. engaging with multiple media platforms simultaneously (Uncapher et al., 2017). Individuals with high levels of media multitasking have been shown to have differences in cognition and poorer memory, increased impulsive behavior, and decreased gray matter volume in the anterior cingulate cortex which is associated with decision-making processes and social-emotional control (Uncapher et al., 2017; Lavin et al., 2013). Although the relationship of causality for media multitasking has not been fully determined, these findings further support the claim that social media platforms access and influence higher activities of the brain.

Emotion and Radicalization

In addition to influencing activities of the brain associated with salience and reward, social media usage can significantly influence our emotional behavior. Platforms like Twitter, Instagram, and Facebook expose us to an endless stream of content that is emotionally salient and varied in reward (Lewis, 2017). The algorithms on these platforms are specially designed to promote content that catches a user's attention via emotionally arousing posts (Munn, 2020). Controversial content, specifically content that elicits emotionally volatile and negative reactions online, has the highest sharing ratio on social media platforms (Oliveira and Azevedo, 2023). As a result, users of social media can experience a range of positive or negative emotions with varying intensities when engaging with these platforms, and these emotions can persist long after the user has exited the software (Steinert and Dennis, 2022).

Twitter introduced machine learning in 2016 to personalize the user experience through personalization algorithms for the Home timeline (Huszár et al., 2022). Studies have shown that the recommendation systems on Youtube, Facebook, and Twitter tend to promote incendiary content because anger drives engagement (Huszár et al., 2022; Munn, 2020). In fact, designers of these platforms have admitted to the exploitation of negative content to drive user engagement (Lewis, 2017). Therefore, engaging with these platforms can significantly impact a user's emotional stability and wellbeing. Further, the circulation of content that prompts these negative emotions can contribute to hate speech and toxic communication that are prevalent on social media (Oliveira and Azevedo, 2023).

The constant consumption of negative forms of communication and extreme disinformation can devolve into long-term feelings of distress and prompt harmful behaviors offline. This phenomenon is demonstrated through cases of internet radicalization where social media plays a prominent role in proliferating hateful, extremist views (Odag et al., 2019). Research on online jihadism shows that the proliferation of Jihadist propaganda on social media preceded Islamist terrorist attacks between 2014-2016 (Enomoto and Douglas, 2019).

Of course, given the influence of other external factors, the direction of causality in the relationship between terrorist propaganda on social platforms and terrorist attacks cannot be fully determined. This data, however, lends support to the idea that social media can play a significant role in fueling an echo chamber for angry voices and influencing behavior offline. Thompson (2011) describes social media as the "perfect platform for the radical voice" that simultaneously deconstructs social-norm behaviors, feeds into addictive user behavior, and encourages private information sharing on a public platform. As a result, Thompson argues that social media applications, like Twitter and Facebook, make the individual user feel as though they are connected to an event as it is unfolding in real time, increasing their emotional reaction and the probability of their radical support. Victimization narratives, i.e. narratives that make the user feel intentionally targeted, are often circulated on social media by radical groups seeking recruits (Decety et al., 2018).

From a social neuroscience perspective, the success of these narratives in radicalizing individuals can be explained by our human sensitivity to injustice, otherwise known as the justice motivation phenomenon (Decety and Yoder, 2017). When individuals feel as though their in-group is threatened via the consumption of targeted algorithmic information, they often experience an emotional reaction that can influence their perceptions and behaviors (Decety and Yoder, 2017).

Psychosocial Behavior and The Effects of Globalization

Social media platforms have played a significant role in accelerating globalization over the past decade. According to the Pew Research Center, in 2021 approximately 7 out of every 10 Americans used social media and visited the platforms at least once a day (Pew, 2021). The expansion of people's social networks through these online applications has been linked to changes in brain structures linked to social cognition (Kanai et al., 2012). Specifically, it was found that the gray matter densities of the amygdala, left middle temporal gyrus, right superior temporal sulcus, and right entorhinal cortex are correlated with online social network size, which is the same effect observed with real-world social networks (Kanai et al., 2012). These findings were based on experiments conducted on a group of primarily college students, as younger people are more likely to use social media and integrate their real-world networks with online networks.

In many cases, the connections that are formed via social media are viewed as an equivalent and/or a replacement for in-person interactions. Individuals belonging to younger generations, including Generation Z and Alpha[1], tend to have smaller social networks offline than their adult counterparts (Dunbar, 2016). As a result, greater emphasis is placed on online social networks, and parasocial interactions become prevalent among younger audiences (Bond, 2016). In a survey of 316 adolescents, Bond (2016) found that teens are more likely to learn from public figures with whom they have formed parasocial relationships. Thus, not only do parasocial relationships contribute to people's feelings of social connectedness on social media, but they can be a useful strategy for public figures to exploit in their messaging and advertising.

The high density of connections and interactions circulating on social media platforms also promotes the illusion of informational accuracy, because the same information can be consumed through multiple sources (Thompson, 2011). As social creatures, human learning is a highly social endeavor and we acquire our knowledge through the testimony of other people (O'Connor and Weatherall, 2019). The more times a piece of information is circulated on social media, the more inclined we are to believe it as fact (Hassan and Barber, 2021). As O'Connor and Weatherall warn, however, this pattern of circulation and amplification through peer-to-peer transmission on social media can promote false beliefs. In a study conducted on ISIS supporters on Twitter it was found that ISIS's success on social media was largely due to the work of a small group of users, 500 to 2,000 accounts, that posted a high volume of tweets (Berger and Morgan, 2015). The work of a small group of users can have impacts on public perception on a mass-scale.

[1] According to Pew Research Center (2019), people belonging to Generation Z are defined as being born between 1997-2012. Generation Alpha typically refers to individuals born between 2012-2025 (Library of Congress).

Propaganda and Censorship Defined

The term "propaganda" broadly refers to the deliberate attempt to shape the opinions, personal beliefs, and behavior of a mass target audience through the spread of strategically devised messages (Parry-Giles, 2002; Jowett and O'Donell, 1986). Through a review of existing definitions and requisites for propaganda, Huckin (2016) proposes a composite definition of propaganda that includes 5 features that are both necessary and sufficient for content to be considered propaganda:

> "Propaganda is ***false or misleading*** information or ideas addressed to a ***mass audience*** by parties who thereby gain ***advantage***. Propaganda is created and disseminated ***systematically*** and ***does not invite critical analysis or response***."

In the context of social media, the features of digital propaganda vary slightly from offline propaganda, as individuals/groups posting propaganda often do not hide their identities. Additionally, the level of engagement with propaganda content plays a more significant role compared to other media outlets in promoting its circulation. Lock and Ludolph use the term "digital organizational propaganda" to describe the use of propaganda on online channels, describing digital propaganda as "direct persuasive communicative acts by organizations with an unethical (i.e. untruthful, inauthentic, disrespectful, or unequal) intent through digital channels" (Lock and Ludolph, 2020).

Censorship refers to the "suppression of words, images, or ideas" from public access that are considered offensive or objectionable (ACLU, 2019; ALA, 2008). In the United States, under the First Amendment, censorship by the government is unconstitutional (ACLU). As social media usage becomes widespread and the free flow of information poses a threat to political powers seeking control, digital censorship and the use of closed intranet networks to blockade against foreign websites have been implemented in countries like China and Iran to control the spread of information (Ding, 2020).

Propaganda and the Brain

Previous studies show that exposure to propaganda can contribute to the formation of false memories, as propaganda employs specific strategies in content presentation to improve learning and memory formation. The novelty and emotional salience of propaganda can have an influence on our learning and memory processes, resulting in greater uptake of fake news (Barr, 2019). This study identifies four distinct characteristics of social media propaganda that shape learning and memory formation: repetition, linguistic extremity, logical fallacies, and mis/disinformation.

Repetition

Repeated propaganda messaging on social media presents a distinct harm to individual autonomy, as the propagation of extreme narratives can become subliminal primers that shape people's actions and beliefs (Farahany, 2023). Humans have better memory of things that are overrepresented, as repetition helps our brains encode memories (Hassan and Barber, 2021). In neuroscience, this effect is known as "repetition priming" and refers to the phenomenon where repeated exposure to a stimulus elicits an improved response to the stimulus over time (Lee et al., 2020). Repeated information is more likely to be perceived as truthful than new information, otherwise known as the "illusory truth effect" (Hassan and Barber, 2021). Specifically, Hassan and Barber (2021) found that when people were exposed to a statement repeatedly, there was a significant increase in the perceived truth, with the largest increase occurring after the second encounter.

Linguistic Extremity

The encoding bias from repetition is amplified when the language is more extreme. People are more likely to attend to extreme language compared to neutral language, and research suggests that linguistic extremity increases message processing and attitude strength (Craig and Blankenship, 2011). Attitude strength refers to the extent to which an attitude is resistant to change and shapes cognition and behavior (Krosnick and Smith, 1994).

More extreme language also yields significantly higher response rates (Andersen and Blackburn, 2004). The study conducted by Andersen and Blackburn (2004) sent email surveys to nearly 12,000 participants with either high or low language intensity, e.g. “We are sure that you heard..." for high intensity vs. "As you may be aware ..." for low intensity. The results show that there was greater compliance for the higher language-intensity message compared to lower. This indicates a strategic application for using more extreme language to increase people’s engagement amidst a saturation of information.

Not only does linguistic extremity affect responsiveness, but it can also impact people’s perceptions of source credibility, such that people who use more powerful, extreme language are viewed as more reliable (Sparks and Areni, 2007; Hosman and Wright, 1987; O’Barr, 1982). This effect may be amplified on Twitter where the 280-character limit conveniently enhances people’s ability to make sweeping, powerful statements without space to provide evidence. Sparks and Areni (2007) found that language power acts as a peripheral cue when people are limited in their ability to process the details of the message. A peripheral cue denotes a factor that is external to the merits of an argument but plays a role in a person’s evaluation of the argument (APA Dictionary of Psychology).

Logical Fallacies

Logical fallacies are faulty arguments with fundamental errors in reasoning (Svedholm-Hakkinen and Kiikeri, 2022). Often used as rhetorical devices, fallacies can be psychologically compelling, especially on social media platforms where message brevity is commonplace and rewarded (Svedholm-Hakkinen and Kiikeri, 2022). Logical fallacies are an attempt to convince people to accept an argument as fact and can be especially appealing in cases when it strongly appeals to people's emotions at the expense of reason. With logical fallacies on Twitter, there are additional social factors, including replies, retweets, views, and likes, that may influence one's perception of the truth of an argument. The social quality of argumentation is amplified on Twitter, and this often plays to the advantage of fallacious arguments seeking to appeal to emotion over logic.

The cognitive processes involved in social cognition are directly tied to those implicated in argumentative reasoning, as our ability to perceive other people's opinions can influence our own perceptions (Prado et al., 2020). Similarly, social cognition allows for more effective argumentation, as people spreading propaganda messaging engage metacognitive processes to understand how specific groups of people will respond to targeted information. In cases where moral reasoning is involved–which is often the case when consuming content on the protests in Iran–regions of the brain implicated in emotional processing are activated, including the medial orbitofrontal cortex, the temporal pole and the superior temporal sulcus of the left hemisphere (Moll et al., 2002). Emotional responses that are elicited by fallacies can impact one's cognitive control and decision-making process.

The medial prefrontal cortex (mPFC) is a primary brain region implicated in argumentative reasoning and discourse processing (Prado et al., 2020). By tapping into emotional processes and/or diverting one’s attention to an alternate topic, logical fallacies may hijack the process of argumentative reasoning. Ad hominem and red herring fallacies are two prominent forms of faulty argumentation identified on Twitter. An ad hominem fallacy attacks the person/group making the argument, rather than addressing the merits of the argument itself. A red herring fallacy redirects to another topic unrelated to the argument at hand (Butte). Both types of fallacies are effective because they serve as strategic distractions that can elicit an emotional response and lead to false conclusions (Rivera, 2018).

Misinformation and Disinformation

With the expansion of social media usage in the past decade, the threat of disinformation is becoming more prevalent. Social media platforms allow anyone to share information at a high speed without checks for factual accuracy. The saturation of information on social media and group polarization has contributed to the rapid spread of fake news on platforms like Facebook and Twitter, and the habitual use of social media can be a significant driver in the spread of misinformation. The reward-based learning systems on social media can cause users to get into habits of sharing information without considering the veracity of the information (Ceylan, 2023). People who engage in the habitual behavior of liking, sharing, and reposting content on social media are likely responding to autonomic platform cues without considering the response outcomes of sharing false information (Ceylan, 2023).

Misinformation is defined as false or inaccurate information that is shared without deceitful purpose (O’Connor and Weatherall, 2019). By contrast, disinformation refers to content that is intentionally shaped to mislead the population (O’Connor and Weatherall, 2019; UNHCR, 2021). In recent years, disinformation has become harder to identify as the saturation of content on Twitter allows for false information to become widespread in short bursts of time. The difficulty to determine whether false information is intentionally being spread to deceive has become an even greater concern with the acquisition of Twitter by Elon Musk in 2022. Musk's "free speech" agenda has resulted in layoffs of nearly 5,000 of Twitter's 7,500 employees, including Twitter employees involved in content moderation, misinformation policy, and state media (Alba and Wagner, 2023; Das, 2023).

With diminished checks in place for informational accuracy on Twitter, the spread of disinformation is becoming a subversive threat. State-affiliated media sources can take advantage of decreased regulation on Twitter through curated messaging that spreads rapidly via peer-to-peer transmission (O'Connor and Weatherall, 2019). The constant re-sharing of information that occurs between Twitter users allows for disinformation through propaganda to appear as misinformation spread by ill-informed individuals. Consequently, propaganda can take on new, subtle forms as targeted messaging spreads through individual users rather than already discredited accounts.

Why Neuroethics?

Neuroethics is an interdisciplinary field that examines ethical issues within neuroscience and questions how our values and moral behavior may be shaped by underlying neural mechanisms (Roskies, 2002). Roskies outlines two primary divisions of neuroethics, namely the ethics of neuroscience and the neuroscience of ethics. As the categories imply, the former explores the ethics of neuroscientific developments and the implications of using neuroscience to explain or enhance behavior. By contrast, the neuroscience of ethics investigates the influence of brain function on ethical behavior and the biological basis of moral cognition, raising questions on how free will, autonomy, and/or personal biases are impacted by brain function (Roskies, 2002). With the proposed classification of social media as neurotechnology, this study addresses both divisions of neuroethics, including the effects of social media propaganda and censorship on the brain and the ethical implications of digital propaganda from a neuroscientific lens.

Neuroethics is not the only field of study to rely on feedback between ethics and science. Feminist neuroscience and molecular biology scholar, Dr. Deboleena Roy states, "All ethics-amalgamated disciplines, including bioethics and the newly fashioned field of neuroethics, have always reacted back upon themselves—neuroethics is not unique in this sense," (Roy, 2011). Despite its overlap with other frameworks of ethical analysis, neuroethics is a distinctly useful framework to examine ethical issues that arise from our ability to access, monitor, and/or influence the brain with novel technologies (Roskies, 2016). Neurotechnologies are typically the center of neuroethical review, as their ability to monitor, influence, and/or manipulate the brain present relatively novel ethical challenges (Farah, 2010). Although there is substantial overlap between the fields of bioethics and neuroethics, the latter is not merely a subset of the former. As Roskies (2002) states, "...our ever-increasing understanding of the brain mechanisms underlying diverse behaviors has unique and potentially dramatic implications for our perspective on ethics and for social justice. These are the issues that warrant the introduction of a new area of intellectual and social discourse."

The field of neuroethics has faced skepticism as "neuroethical" issues are not novel issues and, therefore, do not appear to warrant a distinct ethical framework for analysis. In response, several neuroethics proponents argue for a more instrumental framing of neuroethics. Racine and Sample (2019) argue that if we view neuroethics as a tool for analysis rather than as an end in itself, then we can gain important knowledge about ourselves and our moral values in novel contexts related to the brain. For example, by exploring the neurological foundation of moral behavior, the neuroscience of ethics offers a new lens of analysis that bioethics does not address (Racine and Sample, 2019). William Safire, Chairman of The Dana Foundation, stated, “The brain is the organ of individuality…when we examine and manipulate the brain—unlike the liver or… the pancreas–we change people’s lives in the most personal and powerful way. The misuse or abuse of this power, or the failure to make the most of it, raises ethical challenges unique to neuroscience,” (The Dana Foundation, 2004). Ethical issues within neuroscience research and technology give rise to prominent values in neuroethics that are distinct from bioethics framings, such as cognitive liberty, mental privacy, and the relationship between personal identity and consciousness.

For the purposes of this study, Martha J. Farah’s framework of neuroethical analysis is followed, which involves the examination of the ethical, legal, and societal implications of a neuroscientific issue at hand (Farah, 2012). In previous work, this framework of analysis has been applied to several issues in neuroscience, including brain enhancement and equity, artificial intelligence ethics, and the impact of emerging neurotechnologies on agency (Farisco et al., 2022; The Dana Foundation, 2004). Previous work in neurotechnology ethics specifically identifies agency as being a central phenomenon for framing the ethical implications of novel neurotechnologies (Goering et al., 2021). Agency in this context refers to an individual’s “ability to enact their intentions on the world through authoring their actions” and encompasses autonomy, identity, and authenticity (Goering et al., 2021). In cases where neural manipulation compromises our agency, either via direct or indirect influence, neuroethics provides a comprehensive framework for developing agency-competencies. By becoming aware of how we are influenced by neurotechnology, we may take measures to preserve our capacity to resist influences we deem to be negative (Brown, 2020).

In the context of this study, using a neuroethical framework allows for a deeper understanding of how the neurological effects of social media manipulation and digital propaganda can affect the user's agency and autonomy, privacy, and freedom of thought. The neurological effects of social media as it pertains to addiction, shifts in cognitive salience and reward processes, and patterns of emotional arousal can make digital propaganda and censorship insidious forms of neural manipulation that necessitate a tailored framework for ethical analysis.

Manipulation of Social Media: A Neuroethical Issue

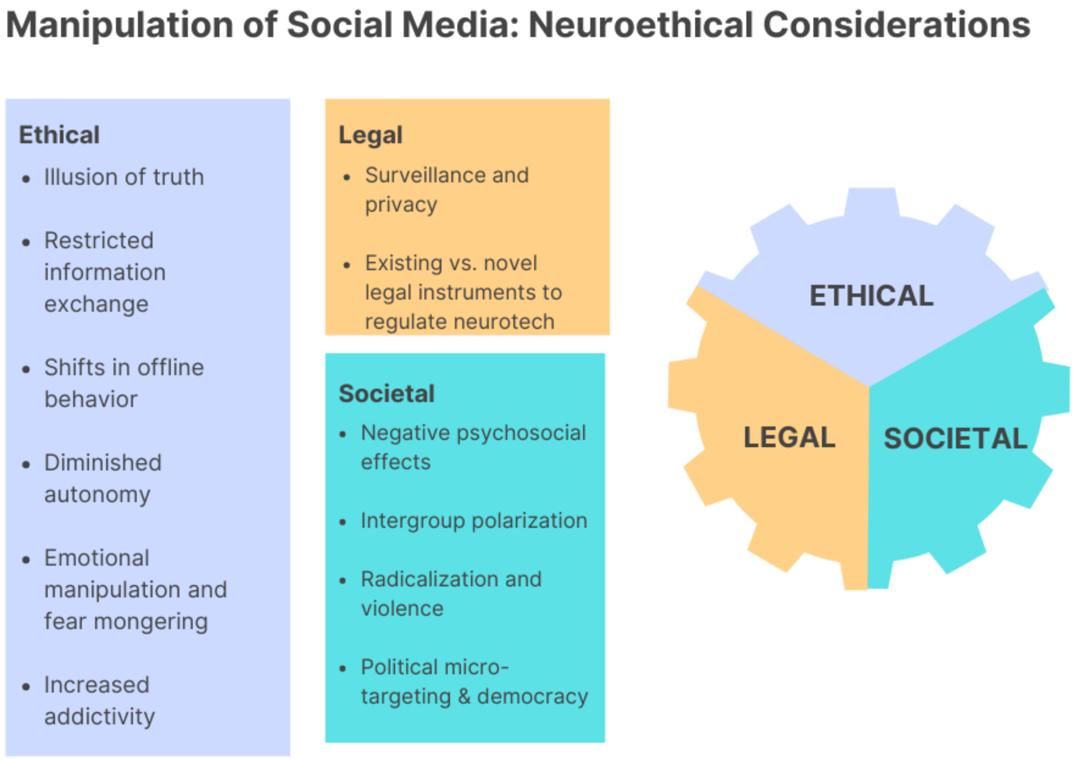
The COVID-19 pandemic has highlighted the power of social media as a platform for information exchange. At the same time, however, it has illuminated the power of the government to censor and alter information to maintain favorable public opinion and shape international perceptions (Simon and Mahoney, 2022). In 2020, the governments in China, Iran, Russia, Egypt, Brazil, and the US all engaged in forms of COVID censorship to maintain the social order (Simon and Mahoney, 2022). The most extreme cases of censorship and propaganda regarding COVID-19 infection rates were prevalent under totalitarian regimes, such as Iran and China, which also had the worst COVID outbreaks in the Middle East and Asia (Turak, 2020. In Iran, authorities conducted waves of arrests in February 2020 of people who shared information about the virus that went against the word of the government. On February 26, it was reported that the Iranian Cyber Police arrested 24 people for "online rumor-mongering about the spread of the coronavirus in the country" and another 118 people received warnings (The Times of Israel, 2020; Simon and Mahoney, 2022).

As part of the punishment for people who spoke out against their government's response to the COVID-19 pandemic, several individuals were banned from using social media. In Iran, Mostafa Nili, a human rights lawyer, was charged with “propaganda against the state” for speaking out against the government’s mismanagement of the virus (Human Rights Watch, 2022). For his sentence, Nili faced 4 years of imprisonment, along with a 2-year ban from practicing law and a 2-year ban from using social media (Human Rights Watch, 2022).

With the ubiquity of social media usage, banning Nili from social media and punishing several others for their posts was an attempt to censor any information that contradicted the government’s messaging about the state of the pandemic in Iran. Since the protests in December 2017, the government’s surveillance and filtration of the internet has posed a significant threat to freedom of expression and the free flow of information (Article 19, 2019). In cases of unrest, including the current protests, Iranians have experienced several nationwide internet shutdowns. Although it is unclear who ordered these shutdowns, evidence points to the involvement of the Communication Regulatory Authority (CRA) under the Ministry of Information and Communications Technology (Article 19, 2022). These blackouts are an attempt to restrict the flow of information within and outside of Iran and also strategically prevent protesters from communicating with one another via social media and messaging apps.

In cases when people do have access to social media, they may undergo a process of self-censorship for fear of being charged with vaguely outlined offenses, including “spreading propaganda against the system”, “insulting Islamic sanctities”, and “insulting the Supreme Leader” (Article 19, 2021). Iranians already have a complex history with self-censorship through other forms of expression, such as art, poetry, cinema, and literature, but social media surveillance introduces a new level of self-censoring interpersonal interactions online (Guardians of Thought, 1993).

These instances of censorship, propaganda, and punishment illuminate the existing threat of social media manipulation and restrictions on information exchange. Given the underlying effects of social media on the brain and its function as neurotechnology, any manipulation or restriction of social media via internet blackouts, filtration, bans, and/or mass propaganda ought to be considered a neuroethical issue, especially in cases of political unrest. Fig. 2 outlines the neuroethical considerations that arise with social media manipulation and the discussion of possible solutions for harm-reduction.



Manipulation of Social Media: Neuroethical Considerations

Ethical

- Illusion of truth
- Restricted information exchange
- Shifts in offline behavior
- Diminished autonomy
- Emotional manipulation and fear mongering
- Increased addictivity

Legal

- Surveillance and privacy
- Existing vs. novel legal instruments to regulate neurotech

Societal

- Negative psychosocial effects
- Intergroup polarization
- Radicalization and violence
- Political micro-targeting & democracy

Fig. 2. Ethical, legal, societal implications of the manipulation of social media.

The Case of Twitter Usage in Iran, Sept. 2022-March 2023

Overview

Protests in Iran have been ongoing since the death of Mahsa Zhina Amini on September 16th, 2022. Mahsa was a young Kurdish Iranian woman who was killed by the Iranian morality police for wearing her hijab improperly. Since news of Mahsa Amini's death became public, thousands of Iranians have taken to the streets in protest of the oppressive regime under Supreme Leader Khamenei. Mahsa Amini quickly became a symbol of the public's desire to regain freedom and relinquish the oppressive grip of Iran's authoritarian theocracy (Ioanes, 2022).

Since September, protests have spread to over 50 cities in Iran. Students of all ages are protesting against the regime through acts of civil disobedience. Videos of girls ripping off their mandatory veils while chanting the slogan "Women, Life, Freedom" have spread across the globe through social media. Several other minority groups in Iran have also gotten involved, resulting in a diverse population of protestors, each with their own grievances. As Firoozeh Kashani-Sabet recently stated to Vox journalists, "For Iranian dissenters, gender issues are not their only grievances, but this fight has enabled them to connect gender violence and inequality to the regime's other authoritarian behaviors" (Ioanes, 2022). What began as a fight against gender violence has expanded to a full-scale call for the death of the regime.

The protests against the Islamic Republic have prompted government shutdowns of the internet and prosecutions of individuals who express sentiments against the government. Knowing that access to social media has significant global influence, the Iranian government has censored all key internet services in Iran and the internet infrastructure has become centralized under the Telecommunications Infrastructure Company of Iran (CSIS, 2022). As a result, people in Iran have restricted and/or skewed access to information due to the government's attempt to control people's thoughts and beliefs about the Islamic Republic.

Internet and social media censorship in Iran is nothing new. From December 2017-January 2018, the Iranian government blocked access to Facebook messenger, Instagram, and Telegram due to anti-government protests (Basso et al., 2022). In November 2019, amidst another set of anti-government protests coined "Bloody November", Iranians experienced an internet blackout. People could only access the national "intranet" for news and updates, a platform that only hosts Iranian websites and promotes Islamic values (OONI, 2022).

In October 2021, there were significant disruptions in Iran's internet. The international bandwidth, controlled by the Telecommunication Infrastructure Company of Iran (TIC), significantly declined and people continue to face issues accessing international internet services (Article 19, 2022). In the same year, the Protection Bill was proposed in Iran's Parliament, aiming to restrict access to all foreign-based social media platforms and criminalize the use of VPNs to bypass website restrictions (Article 19, 2022; Ziabari, 2022). The bill faced considerable public backlash and dispute, and although it has not officially been passed, many Iranians believe it is silently being implemented. This includes the possibility of hierarchical access to the internet and social media platforms in Iran, such that government officials and affiliated institutions, pro-Islamic Republic journalists, and pro-Islamic Republic public figures have broader access than most citizens (Esfandiari, 2022).

Following the death of Mahsa Amini in September, there were severe mobile network shortages in Iran and Instagram, Twitter, TikTok, Facebook, WhatsApp, Google Play Store, Skype, LinkedIn, and the Apple App Store were all blocked (Reuters, 2022). It is clear that the values enforced by the Islamic Republic are part of a strategic effort to maintain the power of the government. FactNameh, an independently-run fact-checking platform for Iranian news and social media has run several reports on recent disinformation campaigns run by the Iranian government that have worked to deface any opponent of the regime (FactNameh, 2022). These social media campaigns have included ایران_عفیف# (Pure_Iran, promoting mandatory hijab), سلبریتی_های_دوزاری# (Cheap_Celebrities, campaign to deface cinematographers), and دولت_مردم# (Peoples_Government, anniversary campaign celebrating the achievements of the 13th government). The #People's_Government campaign, which circulated in the last week of August 2022, was a prime example of the Iranian government's use of social media to organize sweeping media campaigns and shape public opinion. As FactNameh stated in their report, the spokesperson of the 13th government, Ali Bahadri Jahromi, stated in a meeting that "media activism" was needed to "counter the enemy's media attacks" (FactNameh, 2022). Analysis of the tweets posted with the #People's_Government hashtag shows that many of the top twitter accounts spreading the hashtag were bots and over a quarter of accounts using the hashtag were newly created.

Iran's government has a deep-rooted history with organized propaganda and public defacement of opponents. Messages spread by the government contribute to an "image of truth" that is maintained by regime propaganda. Following the 1979 revolution, the Ministry of Information and Tourism became the Ministry of Culture and Islamic Guidance. The Ministry of Culture is responsible for reviewing and approving all books, films, music, and press in Iran to ensure that everything is in the service of the mission of the Islamic Republic (CHRI, 2013). Now, with the space for anonymity and brevity that platforms like Twitter provide, the spread of disinformation can be more insidious.

Social media and globalization have motivated Iranians to fight against their own oppression, as they can see on social networks that their way of life is not universal, nor is it moral. Children as young as 10 years old are protesting, being imprisoned, and receiving death threats from the regime (Ghobadi, 2022). The situation in Iran is distinct from previous protests in Iranian history as these protests are primarily led by Generation Z individuals who are very active on social media. These young Iranians are motivated to achieve the type of freedom they see online.

Hosein Ghazian, an Iranian sociologist, describes the impact social media has had on young Iranians, stating, “This generation is more up to date and aware of the world they live in. They've realised life can be lived differently,” (Ghobadi, 2022). Social media has become a tool for rapid globalization, and unlike older generations, young Iranians are now exposed to the liberal realities of other countries through platforms like Twitter, Instagram, and TikTok. Consequently, Gen Z Iranians have become skeptical of the Islamic Republic’s construction of oppressive values.

In light of the rapid expansion in social media as a primary outlet for communication, government censorship and propaganda via social media is becoming increasingly prevalent as a control tactic to put forth semi-truths that influence public knowledge and opinion. Given Iran’s significant history of censorship, propaganda, and persuasion through various forms of media, literature, art, and history, this study offers crucial insight into the evolution of censorship and propaganda in the age of social media. As Blitz and Bublitz state, “Even when our thoughts remain wholly unexpressed, modern neuroscience and psychology may now give the government a way to extract information about them—and perhaps, to coerce them,” (Blitz and Bublitz,). Our thoughts are the substrates of all other forms of expression and freedoms. Thus, the case of the Iranian government's control of social media messaging to intentionally and substantially influence people's thoughts about the regime is both a human rights issue and a neuroethical concern.

Methods

For the purposes of this study, a case study qualitative design approach was used to assess social media coverage of the uprisings in Iran. The primary source for data collection was Twitter. All unfiltered coverage of the protests that make it onto Twitter by individuals unaffiliated with the IR typically contrast the narrative maintained by Iranian government officials. Here, I conducted an analysis of tweets posted by individuals affiliated with the Islamic Republic government and press to examine how the current protests and state of life in Iran were represented.

Data Collection

The data set was narrowed to content posted from September 15th, 2022 through March 9th, 2023 by state officials and/or people in Iran who have legally active accounts. Given the Twitter ban in Iran, typically the only legally active Twitter accounts from inside Iran are run by state officials, news agencies, journalists, or public leaders who support the regime. A convenience sample of 41 Twitter accounts was analyzed, including 18 state official accounts, 10 affiliated media accounts, 8 journalist accounts, 4 public figure accounts, and 1 pro-Islamic Republic content account (See Appendix B).

Based on a review of existing research on the cognition of propaganda, posts were categorized as propaganda based on the following elements: repetition, linguistic extremity, the use of logical fallacies, and disinformation.

Data Analysis

Data analysis involved the process of translation, in-vivo coding, categorization, and thematic analysis, as outlined by J. Saldana in *The Oxford Handbook of Qualitative Research*. The Google Translate function on Twitter was used to provide an English translation for tweets in Farsi. The accuracy of the English translation for each tweet was checked and corrected with the help of Dr. Hossein Samei, a native Farsi speaker and Persian language professor.

To assess the theme of linguistic extremity, the relative frequency of the terms “enemy” and “devil” in English and Farsi was calculated for each account holder. The terms “enemy” and “devil” were frequently used interchangeably, especially in Farsi, to describe opponents of the Islamic Republic. Thus, both keywords were included in Twitter Advanced Search to account for all tweets expressing bitter sentiments against the “enemy”. Twitter Advanced Search was used to identify absolute frequency of keyword appearances using input “(enemy OR enemies OR devil OR devils OR دشمن OR دشمنان OR شیطان OR شیاطین) (from:[account]) until:2023-03-09 since:2022-09-15”. For accounts outside of Iran included in the comparative analysis (See Table 1) the absolute frequency of keyword appearances was searched with input “(enemy OR enemies OR devil OR devils OR [aforementioned keywords in account holder’s primary language]) (from:[account]) until:2023-03-09 since:2022-09-15” (See Appendix B).

Relative frequency was calculated by [(Total Keyword Count) / (Avg. Tweet Count * 176)) * 100] where Avg. Tweet Count refers to the average tweets per day for the last 1000 tweets on and before March 9, 2023 and 176 is the total number of days between the start and end date. The average tweet count was calculated by a free Twitter analytics software hosted by accountanalysis.app. More than 1000 tweets per account could not be accessed due to the software's paywall. For accounts that exceeded 1000 tweets before reaching September 15, 2022, the average tweet count was calculated by [(Total number of tweets) / (Number of days between adjusted start and end date)].

Previous studies that have analyzed content patterns on Twitter employed similar equations to calculate relative frequency, i.e. [(Total number of tweets posted within a specified time frame) / (Total number of tweets including the selected keywords)] (Santos and Matos, 2014; Sang and Van den Bosch, 2013). In line with previous studies, retweets were excluded from the sample to ensure popular tweets were not counted more than once. Replies to tweets and quote tweets were included in the sample as several accounts used the reply feature to contribute new content or to respond to individuals.

For the purposes of the qualitative comparison, the relative frequencies of "enemy" and "devil" for the accounts of state officials in other countries were compared to the maximum and minimum estimated relative frequencies for the Iranian state official accounts studied. The terms "enemy" and "devil" were translated into the language of the account holder as needed and included in the Twitter Advanced Search query e.g. (دشمن OR دشمنون OR شيطان OR شياطين OR enemy OR enemies OR devil OR devils) (from:ImranKhanPTI) for keywords in Arabic. State official accounts from the US, Ukraine, Russia, Ethiopia, Yemen, and Pakistan were included for qualitative comparative analysis as these countries were engaged in civil unrest and political conflict over the studied time frame (Crisis Group, 2023).

Martha J. Farah's structure of neuroethical analysis was used to assess the ethical, legal, and societal implications of the Iranian government's control of social media as a tool for censorship and propaganda. The use of social media as the primary outlet for communication and information sharing through the protests in Iran is a clear example of how people's conceptions, belief systems, and worldviews are inextricably tied to social media. Therefore, analyzing the ways in which the government interferes with these platforms of knowledge creation is critical to understand how freedom of thought is impacted by censorship and propaganda.

Results

Propaganda and Persuasion on Social Media

Repetition and Linguistic Extremity

In the case of social media, posts with extreme language often receive higher rates of engagement and greater circulation by algorithms to encourage users to spend more time on the platform. By its design, Twitter minimizes people's ability to critically examine all aspects of a position by restricting people to a short paragraph and allowing individuals to control the replies to their Tweets. The latter feature was enabled on one of the analyzed accounts. Restricting the "Reply" function to only people the user follows or mentions can play to the user's advantage, as those people are likely to agree with the user's Tweets. As referenced earlier, the more times a piece of information is supported and recirculated on a platform, the more valid it appears to be (Lee et al., 2020). As a result, peripheral cues, including language power, can play a strong role in people's assessments of the reliability of a Tweet's message.

Of the 41 Twitter accounts analyzed, 37 accounts had a relative frequency of the keywords "enemy"/"دشمن" , "enemies"/ "دشمنان", "devil"/ "شیطان", "devils"/"شیاطین" greater than zero (See Appendix A, Table A1). Seven accounts had an estimated relative frequency greater than 10%, two of which were accounts affiliated with the Supreme Leader of Iran (See Appendix

A, Table A1). The maximum relative frequency was 18.79%. The term “enemy” and all associated keywords were often used without reference to a specific target, although it can be inferred that these extreme terms are referring to foreign powers, specifically Western powers (see Table 1).

The qualitative comparative analysis between Iran state officials and state official accounts from the US, Ukraine, Russia, Ethiopia, Yemen, and Pakistan count show a considerable difference in the relative frequency for the Supreme Leader compared to all other sampled accounts. The lowest RF of the Iranian state official accounts studied (1.42%) was still greater than the highest RF for the non-Iranian accounts studied (0.60%). Six accounts, including the accounts for the President and VP of the US, had an absolute and relative frequency of 0.

Name of Individual/Organization	Absolute Frequency of “Enemy” / “Devil” in Tweets and Replies (September 2022 to March 2023)	Average Number of Tweets / Day	Relative Frequency (%)
Supreme Leader of Iran	43	1.3	**18.79**
Spokesman of the Ministry of Foreign Affairs of Iran	2	0.8	**1.42**
Current Prime Minister of Pakistan	2	1.9	**0.60**
President of Ukraine	3	3.1	**0.55**
Former Prime Minister of Pakistan	2	2.5	**0.45**
Ministry of Foreign Affairs of Russia	4	6.3	**0.36**
President of USA	0	11.8	**0.00**
Vice President of USA	0	1.6	**0.00**
President of Yemen	0	0.1	**0.00**
President of Ethiopia	0	0.1	**0.00**
Office of the President, Ethiopia	0	0.8	**0.00**
President of Pakistan	0	0.3	**0.00**

Table 1. Absolute and relative frequencies of “enemy”/“devil” for Iranian state official accounts and state official accounts from other countries in similar states of political unrest. State official accounts from the US, Ukraine, Russia, Ethiopia, Yemen, and Pakistan were included for qualitative comparative analysis as these countries were engaged in civil unrest and political conflict over the studied time frame. Of the Iranian state official accounts studied, the highest and lowest RFs were selected to be included in this comparative analysis.

Since the start of the protests in Iran in September 2022, many appearances of the keyword “enemy” have been in reference to enemy “interference” and “plotting” as the explanation for the unrest. The Supreme Leader of Iran, other state officials, affiliated news agencies, and journalists maintain the narrative that the protests are not a backlash to the state of oppression under the Islamic Republic but an effect of “the enemy’s” attacks. Ebrahim Raisi, the president of Iran, employed this tactic in several of his statements to the public in the first few months of the protests. His quotes were then shared and re-shared via Twitter through his personal account as well as government-affiliated news agencies on Twitter, including the following tweeted statements:

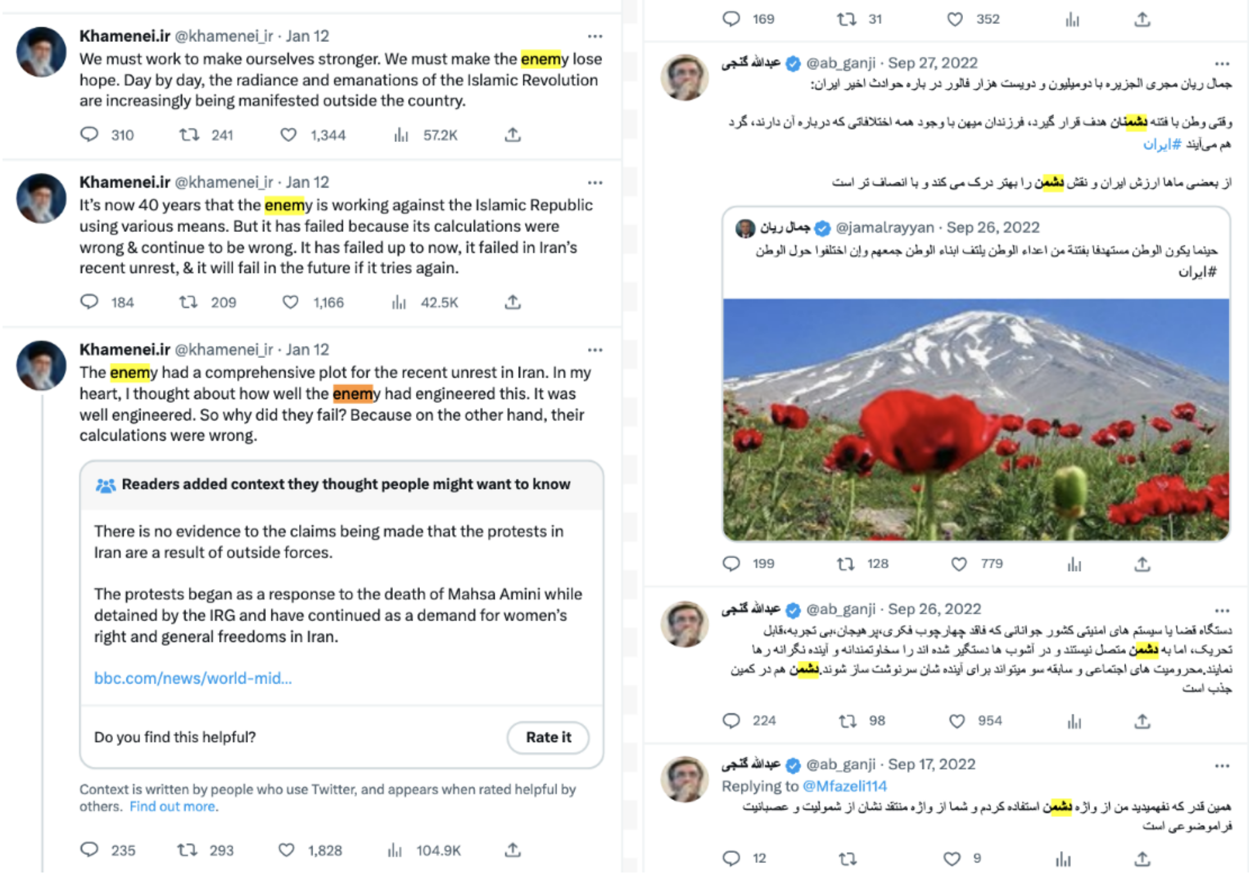
Twitter Account Source: @raisi_com

- ”The enemy's goal is to weaken national unity and bring people, artist to artist, and athlete to athlete.”
- “The enemy seeks to create fear among students and parents. Regarding the issue of student poisoning, I instructed the Minister of Information and the Minister of the Interior to follow up on the issue as soon as possible and provide the people with the reports in full.”

Twitter Account Source: @Iran_GOV

- “President Raisi: Enemies of #Iran take revenge for desperation of creating rifts in the united lines of nation with violence and terror. #ShirazAttack”
- “President Raisi terms false news as part of enemy’s psychological operations”
- “President Raisi: ‘Today, if the situation in our cities is calm and we are safe, it is thanks to the blood of dear young people who stood against the rioters. Although the grief of their loss is very difficult for us, but their achievement was the despair of the enemy’. #iran”
- “President Raisi: ‘The enemy seeks to induce despair and hopelessness by the recent events and we must take effective measures to advance affairs and solve people’s problems by confronting this conspiracy by the enemies’”
- “President Raisi: Islamic Republic has always been subject to the wrath of enemies due to its lofty goals”
- “#Iran’s President Raisi: ‘The enemy thought that they could follow their desires inside the university, unaware that our students were awake and would not allow the enemy’s false dreams to come true”

The tweets above incorporate brief but targeted messaging against an unnamed enemy. In most cases, the enemy is described as a foreign entity with a plot to undermine the progress of the Islamic Republic. The term "enemy" takes on a broad meaning in messaging by the Islamic Republic, as it is not solely a military concept but a word to describe any political, ideological, and/or economic opponent of the regime. There is no evidence that the current protests are the result of foreign interference but demonizing a common enemy is a propaganda technique to promote national unity and discredit the unrest (Çakı et al., 2018).



Khamenei.ir @khamenei_ir · Jan 12 ···
We must work to make ourselves stronger. We must make the enemy lose hope. Day by day, the radiance and emanations of the Islamic Revolution are increasingly being manifested outside the country.
310 241 1,344 57.2K

Khamenei.ir @khamenei_ir · Jan 12 ···
It's now 40 years that the enemy is working against the Islamic Republic using various means. But it has failed because its calculations were wrong & continue to be wrong. It has failed up to now, it failed in Iran's recent unrest, & it will fail in the future if it tries again.
184 209 1,166 42.5K

Khamenei.ir @khamenei_ir · Jan 12 ···
The enemy had a comprehensive plot for the recent unrest in Iran. In my heart, I thought about how well the enemy had engineered this. It was well engineered. So why did they fail? Because on the other hand, their calculations were wrong.

Readers added context they thought people might want to know

There is no evidence to the claims being made that the protests in Iran are a result of outside forces.

The protests began as a response to the death of Mahsa Amini while detained by the IRG and have continued as a demand for women's right and general freedoms in Iran.

bbc.com/news/world-mid...

Do you find this helpful? Rate it

Context is written by people who use Twitter, and appears when rated helpful by others. Find out more.

235 293 1,828 104.9K

169 31 352

عبدالله گنجی @ab_ganji · Sep 27, 2022 ···
جمال ریان مجری الجزیره با دومیلیون و دویست هزار فالور در باره حوادث اخیر ایران:

وقتی وطن با فتنه دشمنان هدف قرار گیرد، فرزندان میهن با وجود همه اختلافاتی که درباره آن دارند، گرد هم می‌آیند #ایران

از بعضی ماها ارزش ایران و نقش دشمن را بهتر درک می کند و با انصاف تر است

جمال ريان @jamalrayyan · Sep 26, 2022
حينما يكون الوطن مستهدفا بفتنة من اعداء الوطن يلتف ابناء الوطن جمعهم وإن اختلفوا حول الوطن
#ايران

199 128 779

عبدالله گنجی @ab_ganji · Sep 26, 2022 ···
دستگاه قضا یا سیستم های امنیتی کشور جوانانی که فاقد چهارچوب فکری،پر هیجان،بی تجربه،قابل تحریک، اما به دشمن متصل نیستند و در آشوب ها دستگیر شده اند را سخاوتمندانه و آینده نگرانه رها نمایند.محرومیت های اجتماعی و سابقه سو میتواند برای آینده شان سرنوشت ساز شوند.دشمن هم در کمین جذب است

224 98 954

عبدالله گنجی @ab_ganji · Sep 17, 2022 ···
Replying to @Mfazeli114
همین قدر که نفهمیدید من از واژه دشمن استفاده کردم و شما از واژه منتقد نشان از شمولیت و عصبانیت فراموضوعی است

12 9

Fig. 3. Sample tweets demonstrating the repetition of the keyword "Enemy" in English and Farsi by @khamenei_ir and @ab_ganji. Keyword "Enemy" was repeated 8 times on January 12, 2023 (4 shown in image). Keyword " دشمن" was repeated 5 times in the right image between September 17–September 27, 2022.

The appearance of the Community Notes feature in Fig. 3 is noted in response to the claim about the "enemy's" involvement in the Iran protests. The Community Notes feature aims to combat misinformation on Twitter and improve content accuracy, according to the company (Sankaran, 2022).

Logical Fallacies

Ad-Hominem Attacks and False Equivalence Fallacies

An analysis of the Supreme Leader's primary account (@khameini_ir) revealed a pattern of several tweets denouncing the state of life in "the West". By denouncing existing living conditions in western countries, these tweets aim to invalidate the human rights concerns of foreign powers regarding the protests via ad hominem attacks. Tweets by the Supreme Leader's affiliated accounts also establish a false equivalence between a distorted western reality and Islamic idealism. For example, several accounts posted about the prevalence of police violence in the United States but made no mention of the violence of the riot police in Iran. This false equivalence is a tactical means of diminishing the voices of any opposition to the Islamic Republic.

Ad hominem attacks were also made by state officials in response to the removal of Iran from the UN Commission on the Status of Women on December 14, 2022. The vote for Iran's removal was made in response to Iran's pattern of human rights violations against women and girls and the government's severe crackdown on the protests since Mahsa Amini's death (United Nations, 2022). On the day of the resolution's passage, @EnsiyehKhazali, the account for the VP for Women and Family Affairs, tweeted: "The claim of supporting the rights of Iranian women becomes ridiculous when America, the proposer of the resolution, has the largest share of the world statistics of imprisoned women. Western countries, especially America, as the biggest violator of women's rights, do not have the authority to decide to remove Iran from @UN_CSW." This was one of several tweets observed that included an ad-hominem attack against the actions and/or statements of the US government by referencing the oppressive state of life in America.

Red Herring Fallacy

Approximately one-third of the accounts posted at least one tweet that included a red herring fallacy. A review of the selected Twitter accounts revealed a pattern of red herring fallacies being tweeted in October and November 2022. Around the same time as demonstrations were occurring in Iran, thousands of people in Paris marched in the streets to protest the inflated cost of living and demand higher wages (Associated Press, 2022). In response, several accounts belonging to Iranian state officials, media agencies, and journalists posted about the unrest in Paris, sharing videos of violence and police brutality against the French protesters. Given the violent suppression of the protests in Iran by the IRGC, the tweets criticizing the French government and police force are logically hypocritical. Nevertheless, on an emotional level, they can redirect people's attention to other emotionally significant stimuli, thereby dispersing some of the attention off of the protests in Iran.

In addition to highlighting the Paris protests, several accounts posted tweets sharing the crimes of other governments and the poor quality of life in western countries. In many cases, these users offered remorse for western countries and shared broad statistics that focused on other governments' shortcomings:

> **"Western capitalist system is a patriarchal system…Every person who can invest more is worth more. Macroeconomic and business management was done by men. Therefore, in the capitalist system, men have priority over women."**
> [Source: @khameineireyhane, 01.04.2023, English translation]

> **"I feel sorry for the people of Europe and America, For thousands of people in the bone-burning cold, They sleep in tents, And with the advancement of technology,**

> **They still burn wood to heat themselves, I'm sad! Poverty is rampant in foreign countries…Then we are here in comfort and peace and prosperity."**
> [Source: @hosseini_eco, 01.09.2023, English translation]

As part of their pro-government propaganda, these accounts aim to highlight suffering and oppression in other parts of the world, especially the West. These tweets serve to redirect attention from the unrest in Iran to the prevalence of poverty, patriarchy, violence, and illness in the United States and Europe.

Data analysis also revealed a pattern of tweets promoting Iranian nationalism and Iranian women's pride since the start of the protests. The hashtag #ایران_عزیز appeared 29 times in tweets by the selected accounts, often accompanying tweets celebrating athletic achievements by Iranians as a source of pride. A review of the accounts showed an increase in the frequency of the hashtag after September 15th, 2022 compared to the months before.

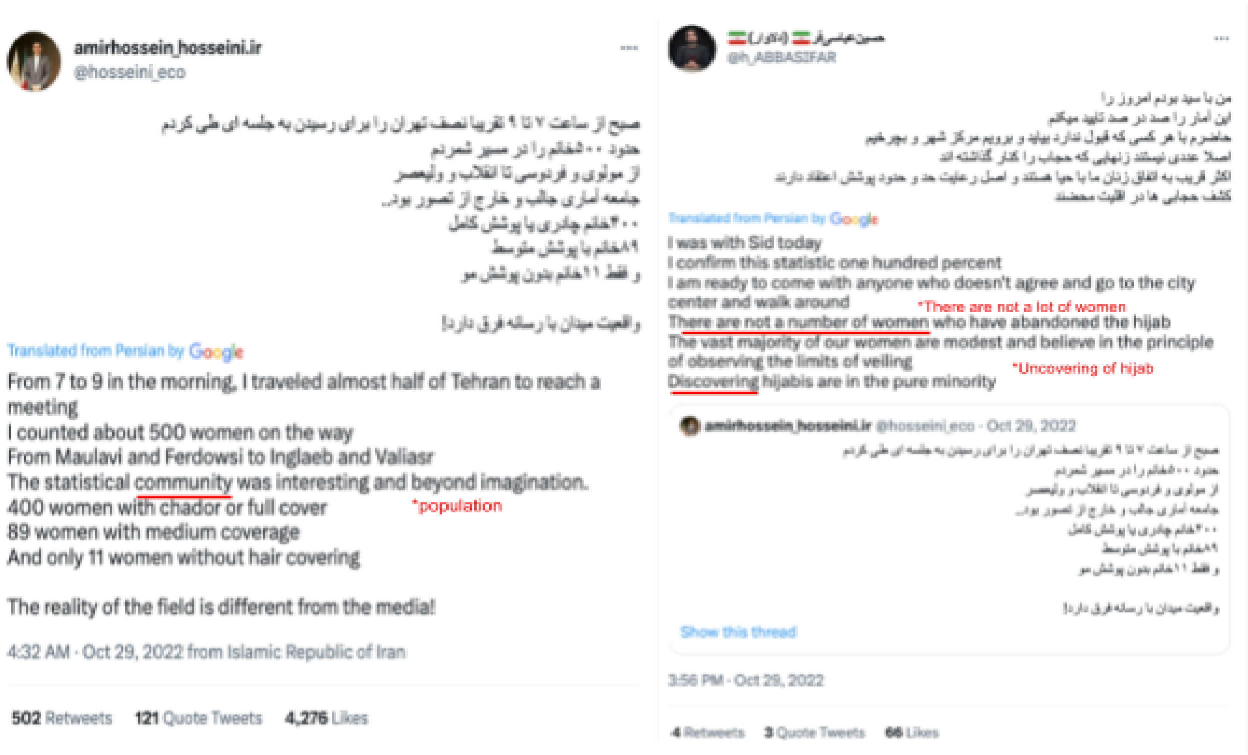
Disinformation

Analysis of potential disinformation on Twitter revealed two methods of intentionally spreading false information: proof by anecdote and baseless claims. Proof by anecdote refers to a broad generalization that is made based on an individual experience (PropWatch). A baseless claim is a statement that is presented as fact without evidence to support it (PropWatch). Although these methods of argumentation are also logical fallacies, they are types of logical fallacies that make broad claims from limited evidence, and this lends well to the spread of disinformation through brief and frequent tweets.

Proof by Anecdote

As part of the protests, thousands of Iranians have protested the compulsory hijab law that requires women to wear a headscarf in public from the age of 9 (Bazoobandi and Khorrami, 2022). According to reports from within Iran, many women have appeared in the streets without wearing a headscarf and videos of women burning their headscarves have gone viral on Facebook, Twitter, and Instagram (Esfandiari and Zarghami, 2023). This is not the first time Iranian women have protested the compulsory hijab law, but these protests have been the most widespread and have garnered a wide range of support from people in Iran and internationally (Bazoobandi and Khorrami, 2022). Reports from inside Iran reveal that a significant number of Iranian women, especially younger, college-aged women, have stopped wearing hijab in protest of the state-enforced dress code (Far, 2023).

In response to the mandatory hijab protests, Iranian state official and journalist accounts posted tweets solely containing anecdotal evidence to diminish the prevalence of the protests. In several cases, the tweet only included evidence from personal experience and singular instances of observation (see Fig. 4).



amirhossein_hosseini.ir
@hosseini_eco

صبح از ساعت ۷ تا ۹ تقریبا نصف تهران را برای رسیدن به جلسه ای طی کردم
حدود ۵۰۰خانم را در مسیر شمردم
از مولوی و فردوسی تا انقلاب و ولیعصر
جامعه آماری جالب و خارج از تصور بود..
۴۰۰خانم چادری یا پوشش کامل
۸۹خانم با پوشش متوسط
و فقط ۱۱خانم بدون پوشش مو

واقعیت میدان با رسانه فرق دارد!

Translated from Persian by Google

From 7 to 9 in the morning, I traveled almost half of Tehran to reach a meeting
I counted about 500 women on the way
From Maulavi and Ferdowsi to Inglaeb and Valiasr
The statistical community was interesting and beyond imagination.
400 women with chador or full cover
89 women with medium coverage
And only 11 women without hair covering

*population

The reality of the field is different from the media!

4:32 AM · Oct 29, 2022 from Islamic Republic of Iran

502 Retweets 121 Quote Tweets 4,276 Likes

حسین عباسی‌فر 🇮🇷 ([illegible]) 🇮🇷
@h_ABBASIFAR

من با سید بودم امروز را
این آمار را صد در صد تایید میکنم
حاضرم با هر کسی که قبول ندارد بیاید و برویم مرکز شهر و بچرخیم
اصلا عددی نیستند زنهایی که حجاب را کنار گذاشته اند
اکثر قریب به اتفاق زنان ما با حیا هستند و اصل رعایت حد و حدود پوشش اعتقاد دارند
کشف حجابی ها در اقلیت محضند

Translated from Persian by Google

I was with Sid today
I confirm this statistic one hundred percent
I am ready to come with anyone who doesn't agree and go to the city center and walk around
*There are not a lot of women
There are not a number of women who have abandoned the hijab
The vast majority of our women are modest and believe in the principle of observing the limits of veiling
*Uncovering of hijab
Discovering hijabis are in the pure minority

amirhossein_hosseini.ir @hosseini_eco · Oct 29, 2022

صبح از ساعت ۷ تا ۹ تقریبا نصف تهران را برای رسیدن به جلسه ای طی کردم
حدود ۵۰۰خانم را در مسیر شمردم
از مولوی و فردوسی تا انقلاب و ولیعصر
جامعه آماری جالب و خارج از تصور بود..
۴۰۰خانم چادری یا پوشش کامل
۸۹خانم با پوشش متوسط
و فقط ۱۱خانم بدون پوشش مو

واقعیت میدان با رسانه فرق دارد!

Show this thread

3:56 PM · Oct 29, 2022

4 Retweets 3 Quote Tweets 66 Likes

Fig. 4. Sample tweets by a state official and an affiliated journalist, @hossein_eco and @h_ABBASIFAR, using anecdotal evidence to make generalized claims about Iranian women's compliance with the mandatory hijab law.

Fig. 4 shows a tweet by a government official, Amirhossein Hosseini, recounting his observation of ~500 Iranian women in the streets of Tehran from 7 to 9 AM. In the tweet on the right, state-affiliated journalist, Hossein Abbasifar, replies to Hosseini's tweet to attest to his field observations. Since these tweets refer to a personal experience, it cannot be confirmed nor denied whether the users above accurately counted the attire of 500 women in the streets of Iran. Thus, for propaganda purposes, proof by anecdote is highly effective as there is no method of fact-checking the information. The language of both tweets downplays the presence of the protests with the terms "only" and "minority", but regardless of the accuracy of such evidence, observing 11 women without hijab is still a reflection of unrest in Iran.

Baseless Claims

As referenced earlier in the analysis of linguistic extremity, several state officials, journalists, and affiliated media agencies posted tweets with claims that the unrest in Iran was unnatural and likely the result of Western interference. Historically, the Islamic Republic has maintained a narrative demonizing the Western world as part of their efforts to consolidate power through isolation and indoctrinate the youth through fear mongering. With the protests posing one of the most significant threats to the stability of the government since the revolution, the Supreme Leader of Iran and other hard-liners posted several tweets with baseless claims that the protests are the result of foreign interference led by the United States (Nasr, 2023).

Notable tweets by state officials–@Khameini_fa, @IRIMA_SPOX, @nezammousavi–claiming foreign interference as explanation for unrest:
September 2022 to March 2023

- **“England’s action in attacking the national security of the Islamic Republic of Iran has faced Iran’s decisive intelligence and judicial response. The [disruption] of the British regime and the support of some human rights [claimants] in Europe show their lawlessness…”**

- **“Imam Khomeini: Rest assured! America does not send military. America sends those it has trained to disrupt the situation…”**

- **“This incident that happened, a young girl died. It was a bitter incident. Our hearts were also burned. But the reaction to this incident…these movements were not natural. This disturbance was planned.**

- **“The opponents of the regime know well that the people are standing by the Islamic Republic despite all the criticisms. The goal of the [plotters] is to weaken the system in strategic confrontations of the region”**

- **“European Parliament’s action against #IRGC another part of combined war against #Iranian nation”**

These tweets make broad-sweeping claims about plotting by the US and European countries to stir unrest in Iran. Due to the restricted amount of non-government affiliated information coming out of Iran, it is difficult to confirm whether these claims are true. There is no existing evidence, however, supporting the claim that the unrest is due to planned foreign disturbance. In November 2022, a spokesperson for the judiciary in Iran claimed that 40 foreigners were arrested for their involvement in the protests, but the identities of these individuals were never revealed and no evidence of foreign involvement has been uncovered since those claims (Shan, 2022). Therefore, with the current state of evidence, these tweets contain baseless claims and can propagate the spread of disinformation given their frequency and the similarity in messaging across connected accounts.

An additional pattern of disinformation observed by the analyzed accounts was the use of isolated statistics as evidence of women's freedom in Iran. Several users attested to the high quality of life for women in Iran by providing cherry-picked statistics that demonstrate higher rates of literacy and employment of women following the revolution. Independent of the accuracy of these statistics, increased literacy and employment rates are not wholly representative of the quality of life for women in Iran, as Iranian women still face considerable oppression in systems of marriage, divorce, self-expression, employment, and education (US Institute of Peace, 2022).

Discussion

A Neuroethical Analysis: Ethical, Legal, Societal Implications

Ethical Implications	Legal Implications	Societal Implications
● Online manipulation of vulnerable populations within/outside of Iran ● Diminished user autonomy due to censorship and filtration ● Self-censorship; restricted freedom of thought and expression ● Spread of emotionally salient disinformation	● Applications of existing legal instruments for regulating neurotechnology ● Policy action for greater algorithmic sovereignty on Twitter	● Distorted public perception / limited awareness of protests ● Collective impact of manipulative technologies on decision-making ● Precedent for other totalitarian regimes and political microtargeting on social media ● International tensions; animosity towards the US and European countries

Table 2. Ethical, legal, societal implications of social media manipulation and the spread of digital propaganda during the Iran protests.

As the current study has demonstrated, the use of social media can alter people's cognitive patterns and shape behaviors and beliefs. The case of Iran demonstrates how social media can be manipulated to shape people's behaviors and beliefs, restrict people's freedom of thought, and advance government control. Therefore, social media ought to be classified as neurotechnology, and propaganda and censorship on social media is a neuroethical issue that warrants serious consideration in the process of developing guidelines for neurotechnology. The identified patterns of repetition, linguistic extremity, logical fallacies, and disinformation through the Twitter analysis demonstrate the existence of systematic digital propaganda by the Islamic Republic and its distortion of the current state of life in Iran. Without necessary checks and guidelines in place to minimize the spread of disinformation on social media, we face the threat of increased totalitarian control over social networking sites and information suppression.

This study also built upon Carr and Hayes's (2015) definition of social media by proposing the additional characteristics of addictivity and algorithmic personalization that make social media distinct from media. These two characteristics are directly correlated with the effects of social media on the brain and contribute to a neuroscientific understanding of why social media manipulation is ethically distinct from other forms of media manipulation. This expanded definition of social media is therefore a useful framework for future related neuroscientific studies.

Ethical Implications

There are several ethical considerations that arise with the Iranian government's censorship of social media and observed propaganda on Twitter. First, although many Iranians exposed to social media propaganda campaigns have the capacity and historical rationale to reject targeted messaging, we must consider the impact of social media propaganda on more vulnerable groups. Adolescents, people suffering from mental illness, elderly individuals, and even Twitter users abroad unaware of the Islamic Republic's history of information manipulation are more susceptible to having their beliefs and actions shaped by the messaging of hardliners. Although several of the studied accounts are held in the name of individual state officials, it is likely that an administration dedicated to social media management is responsible for running these accounts. The Supreme Leader of Iran, for example, has at least 13 separate accounts linked to his name and image, nine of which are run in different foreign languages. The perceived intimacy and directness of his messaging on social media is an effective means of persuasion and encourages users to form a parasocial relationship with their Supreme Leader.

Second, the combination of Iran's filtration of social media platforms and Twitter's advanced personalization algorithm results in minimal user autonomy on Twitter. With the added pressure of punishment and/or arrest for any content deemed unlawful by the government, individuals endure a process of self-censorship, refraining from sharing certain information or expressing themselves freely for fear of persecution (Shan, 2022). Constant self-censorship can lead to heightened levels of stress, shifts in personal identity, and cognitive exhaustion (Sansone and Sansone, 2012; Rimé, 2009).

Third, the spread of emotionally salient disinformation by Iranian authorities, news agencies, journalists, and public figures can shape people's actions and beliefs on a mass scale. Coverage of the protests is already filtered and restricted. Thus, the saturation of social media with pro-government information can distort the rest of the world's perception of the unrest.

As revealed in the process of data collection, Twitter already has some checks in place for social media to help regulate disinformation. The Community Notes feature, for example, appeared on a Tweet by the Supreme Leader's English account to address the lack of evidence to support the claim of enemy interference in Iran's unrest. This feature, however, only appeared on ***one*** of the tweets analyzed in this study, and progress must still be made to attack the spread of propaganda and disinformation campaigns that may not be as obvious for immediate detection. Twitter's Help Center page explicitly states, "To help enable free expression and conversations, we only intervene if content breaks our rules…Otherwise, we lean on providing you with additional context." Their rules include the crisis misinformation policy, the synthetic and manipulated media policy, and the civic integrity policy. According to Twitter, any manipulated form of media or misleading information that could bring harm to crisis-affected populations or lead to harm through deceit goes against Twitter's policies and the user will face consequences.

With these guidelines in place, it is unclear whether the government-affiliated accounts in Iran that are engaged in propaganda campaigns have faced any type of consequence. It does not appear that these accounts have faced any limitation from Twitter moderators, aside from certain accounts having their affiliation listed in their Twitter biography, e.g. “Iranian state-affiliated media". Previous studies have shown that there is not a statistically significant difference in engagement for labeled vs. unlabeled tweets, therefore, more progress must be made to improve the moderation of disinformation (Papakyriakopoulos and Goodman, 2022). Future studies may investigate the efficacy of alternate soft moderation practices on the spread of disinformation in Iran and other countries with recorded histories of systematic propaganda and censorship.

Societal Implications

On a societal level, the actions of the Iranian government may set a dangerous precedent for other totalitarian regimes to take advantage of digital propaganda and political micro-targeting. Political micro-targeting refers to the process of influencing voters through targeted stimuli, often taking place on social media (Papakyriakopoulos et al., 2017). The widespread use of social media makes it a convenient platform to influence/manipulate public opinion on a global scale. The Iranian government’s response to the protests in Iran demonstrates their systematic control of social media to propagate disinformation. This study demonstrated existing patterns of digital propaganda in Iran, with several accounts claiming Western interference, repeatedly using the term “enemy” at high frequencies, making baseless claims about the quality of life in Iran, and directing attention to conflicts in other countries.

The saturation of digital propaganda from pro-government accounts also helps to deprioritize content that contrasts the government’s curated narrative. Not only does this type of content distort the rest of the world’s perception of the protests taking place, but the observed fear mongering on Twitter against the “enemy” promotes a culture of animosity and international tensions. Instilling fear in the population through digital propaganda works to reinstate power for those in control by manipulating people’s emotions and ideologically isolating them from the rest of the world.

Legal Implications and Steps for Regulation

Notably, there are ethical and legal challenges with attempting to regulate censorship and propaganda on social media. Based on the results of this case study, it is clear that protecting against the possibility of neural manipulation via social media requires the cooperation of social media companies, international governments, and intergovernmental organizations dedicated to promoting ethical practices with neurotechnology. Existing guidelines by the Organisation for Economic Co-operation and Development (OECD) address the ethical, legal, and social challenges of neurotechnology and outline nine principles for protection against harm (OECD, 2019). In their description of the development of these guidelines, the OECD Council notes that the “vulnerability of cognitive patterns for commercial or political manipulation” is an ethical challenge that must be addressed through regulation (OECD, 2019). Five of the nine principles are directly applicable to the neuroethical issue of manipulating social media for political purposes.

1. Promote responsible innovation in neurotechnology to address health challenges.
2. **Prioritise assessing safety in the development and use of neurotechnology.**
3. Promote the inclusivity of neurotechnology for health.
4. Foster scientific collaboration in neurotechnology innovation across countries, sectors, and disciplines.
5. **Enable societal deliberation on neurotechnology.**

6. **Enable the capacity of oversight and advisory bodies to address novel issues in neurotechnology.**

7. Safeguard personal brain data and other information gained through neurotechnology.

8. **Promote cultures of stewardship and trust in neurotechnology across the public and private sector.**

9. **Anticipate and monitor the potential unintended use and/or misuse of neurotechnology.**

Principles 2 and 9 are applicable to the regulation of disinformation and political manipulation on social media platforms. The use of social media can become unsafe when disinformation manipulates people's cognitive control, motivates individuals to behave differently offline, and/or target groups of people based on the messaging they consume online. One function of social media as perceived by the user is to provide a platform for social connection and free information sharing. Thus, when this function is restricted via censorship and/or manipulated through digital propaganda, this ought to be categorized as the misuse of neurotechnology by political actors.

Principles 5, 6, and 8 recommend the creation of spaces to address issues and build trust with neurotechnology. In the context of social media regulation, the responsibility for content moderation and propaganda detection primarily rests on the social media company, with the support of its users and partners to flag and check content for accuracy. In this era of "post-truth", a term described by Higgins (2016) to describe the "blatant lies being routine across society", the regulation of disinformation on social media must be prioritized, especially in cases of political unrest. Part of re-establishing trust with social media requires greater transparency on how these social media algorithms operate. Building trust also entails granting users more agency over the personal data they share and more knowledge on how these platforms curate the information they are exposed to (Kozyreva et al., 2021). Granting users the capacity to have autonomy over personalization algorithms, aka “algorithmic sovereignty”, is a key step toward aligning with the principles outlined above and protecting social media users from manipulation (Reviglio and Agosti, 2020).

Coupled with the impacts of propaganda on the brain, the Twitter propaganda being spread by state officials, journalists, media agencies, and public figures affiliated with the Islamic Republic is a clear example of this ethical challenge. The classification of social media as neurotechnology is therefore a critical first step toward addressing the neuroethical issue of propaganda and censorship on social media. With detailed guidelines for neurotechnology already in place, it is clear that regulating the manipulation of social media is not a matter of creating novel guidelines for social media. Rather, the inclusion of social media as neurotechnology will help to motivate an urgent and tailored response to ethical violations taking place on these platforms.

Limitations and Procedures for Credibility

The findings were validated by Dr. Gillian Hue for robustness of neuroethical analysis, Dr. Hossein Samei for historical and language accuracy, and Dr. Mark Risjord for credibility of the qualitative study design. The study poses no serious ethical problems as all data is publicly available on social media platforms.

The accuracy of the average tweet count for the “enemy/devil” frequency analysis was limited by the software paywall over 1000 tweets. For accounts that exceeded 1000 tweets before September 15th, 2022, the average tweet count was calculated using the available data. This limitation in data access may have impacted the accuracy of the relative frequency. Future studies with access to larger datasets and Twitter analytics may conduct a similar analysis to provide insight on the prevalence of keywords in social media propaganda by the Islamic Republic.

Additionally, this qualitative analysis was conducted on a convenience sample of government affiliated Twitter accounts in Iran. Although steps were taken to ensure the sample was representative of various types of pro-government affiliated users, including state officials, journalists, news agencies, and public figures, it is not guaranteed that the sample is fully representative of the identified population. For the purposes of this study, the examples of digital propaganda presented fulfill the burden of proof that digital propaganda exists, but future studies can incorporate statistical analysis and big data analytics to build upon the findings of this study on a mass-scale.

Conclusion

This study presented a novel analysis of the neuroethical implications of censorship and propaganda of social media and its pertinent applications to the current unrest in Iran. By expanding upon existing definitions of social media to include addictivity and algorithmic personalization, this study provides a definition of social media that is relevant for future neuroscientific work. The results from the case study of Iran demonstrate how repetition, linguistic extremity, logical fallacies, and disinformation are utilized as propaganda strategies to shape people’s beliefs and behaviors on social media. By advocating for the classification of social media as neurotechnology, we may protect against the manipulation of social media platforms through the application of existing legal instruments for regulating neurotechnology.

References

ACLU (2019) What Is Censorship? American Civil Liberties Union Available at: https://www.aclu.org/other/what-censorship [Accessed March 18, 2023].

AFP (2020) Iranian Cyber Police Arrest 24 Over Coronavirus Rumors. wwwtimesofisraelcom Available at: https://www.timesofisrael.com/iranian-cyber-police-arrest-24-over-coronavirus-rumors/ [Accessed March 18, 2023].

Alba W (2023) Twitter Cuts More Staff Overseeing Global Content Moderation. Available at: https://www.bloomberg.com/news/articles/2023-01-07/elon-musk-cuts-more-twitter-staff-overseeing-content-moderation?leadSource=uverify%20wall [Accessed March 18, 2023].

Allen M (2017) Sean Parker Unloads on Facebook: "God Only Knows What It's Doing to Our Children's Brains." Axios Available at: https://www.axios.com/2017/12/15/sean-parker-unloads-on-facebook-god-only-knows-what-its-doing-to-our-childrens-brains-1513306792 [Accessed March 18, 2023].

Al Talei R, Bazoobandi S, Khorrami N (2022) Hijab in Iran: From Religious to Political Symbol. Carnegie Endowment for International Peace Available at: https://carnegieendowment.org/sada/88152.

American Library Association (2008) First Amendment and Censorship. Advocacy, Legislation & Issues Available at: https://www.ala.org/advocacy/intfreedom/censorship.

Andersen PA, Blackburn T (2004) An Experimental Study of Language Intensity and Response Rate in Email Surveys. Communication Reports, 17, 73 - 84.

Anon (n.d.) accountanalysis | Analysis of Twitter Accounts. accountanalysis Available at: https://accountanalysis.app/ [Accessed March 18, 2023].

APA (n.d.) APA Dictionary of Psychology. Available at: https://dictionary.apa.org/peripheral-cue [Accessed March 18, 2023].

Article 19 (2019) Iran: Worsening Situation for Free Expression Must Be Addressed in Upcoming UPR. ARTICLE 19 Available at: https://www.article19.org/resources/iranupr2019/ [Accessed March 18, 2023].

Article 19 (2020) Iran: Tightening the Net 2020. ARTICLE 19 Available at: https://www.article19.org/ttn-iran-november-shutdown/ [Accessed March 18, 2023].

Article 19 (2022) Iran: Parliament's "Protection Bill" will hand over complete control of the Internet to authorities. ARTICLE 19 Available at: https://www.article19.org/resources/iran-parliaments-protection-bill-will-hand-over-complete-control-of-the-internet-to-authorities/ [Accessed March 18, 2023].

Associated Press (2022) Thousands of French people — including a Nobel laureate — protest over inflation. NPR Available at: https://www.npr.org/2022/10/16/1129378667/paris-france-protest-inflation-climate-wages [Accessed March 18, 2023].

Barr RA (2019) Galaxy Brain: The Neuroscience of How Fake News Grabs Our Attention, Produces False Memories, and Appeals to Our Emotions. Nieman Journalism Lab Available at: https://www.niemanlab.org/2019/11/galaxy-brain-the-neuroscience-of-how-fake-news-grabs-our-attention-produces-false-memories-and-appeals-to-our-emotions/ [Accessed March 18, 2023].

Basso S, Xynou M, Filastò A (2022) Iran blocks social media, app stores and encrypted DNS amid Mahsa Amini protests. ooniorg Available at: https://ooni.org/post/2022-iran-blocks-social-media-mahsa-amini-protests/ [Accessed March 18, 2023].

Berger JM, Morgan J (2015) The ISIS Twitter Census: Defining and describing the population of ISIS supporters on Twitter.

Bestelmeyer PEG, Kotz SA, Belin P (2017) Effects of Emotional Valence and Arousal on the Voice Perception Network. Social Cognitive and Affective Neuroscience 12:1351–1358.

Blitz MJ, Bublitz JC (2022) The Law and Ethics of Freedom of Thought, Volume 1 Neuroscience, Autonomy, and Individual Rights. Cham: Springer International Publishing AG.

Bond BJ (2016) Following Your “Friend”: Social Media and the Strength of Adolescents’ Parasocial Relationships with Media Personae. Cyberpsychology, Behavior, and Social Networking 19:656–660.

Botes M (2022) Autonomy and the Social Dilemma of Online Manipulative Behavior. AI and Ethics.

Brown T (2020) Building Intricate Partnerships with Neurotechnology: Deep Brain Stimulation and Relational Agency. International Journal of Feminist Approaches to Bioethics, 13(1), 134-154.

Butte College (n.d.) Fallacies and Propaganda - TIP Sheets - Butte College. Butte Available at: http://www.butte.edu/departments/cas/tipsheets/thinking/fallacies.html [Accessed March 18, 2023].

Çakı C, Çetin M, Gazi MA (2018) The Examination of The Anti-USA Propaganda Posters in The Iran Revolution. International Journal of Social Sciences.

Carr CT (2017) Social Media and Intergroup Communication. Oxford Research Encyclopedia of Communication.

Carr CT, Hayes RA (2015) Social Media: Defining, Developing, and Divining. Atlantic Journal of Communication 23:46–65.

Center for Humane Tech (n.d.) How Social Media Hacks Our Brains. Available at: https://www.humanetech.com/brain-science [Accessed March 18, 2023].

Center for Humane Technology (n.d.) Press - Center for Humane Technology. Humane Tech Available at: https://www.humanetech.com/press.

Ceylan G, Anderson IA, Wood W (2023) Sharing of Misinformation is Habitual, Not Just Lazy or Biased. Proceedings of the National Academy of Sciences 120.

CHRI (2013) Ministry of Culture and Islamic Guidance. Center for Human Rights in Iran Available at: https://iranhumanrights.org/2013/08/ministry-culture/ [Accessed March 16, 2023].

Chung S, Cho H (2017) Fostering Parasocial Relationships with Celebrities on Social Media: Implications for Celebrity Endorsement. Psychology & Marketing 34:481–495.

Craig TY, Blankenship KL (2011) Language and Persuasion: Linguistic Extremity Influences Message Processing and Behavioral Intentions. Journal of Language and Social Psychology 30:290–310.

CSIS (2022) Protest, Social Media, and Censorship in Iran. Available at: https://www.csis.org/analysis/protest-social-media-and-censorship-iran [Accessed March 18, 2023].

The Dana Foundation (2004) Neuroethics: Mapping the Field (Marcus SJ, ed). New York: Dana Press.

Das (2023) Twitter in Freefall: After Firing 50 Top Executives On the Weekend, Musk Sacks 200 More Workers. Firstpost Available at: https://www.firstpost.com/world/after-firing-50-top-executives-from-twitter-over-the-weekend-elon-musk-sacks-200-more-workers-12211322.html [Accessed March 18, 2023].

Decety J, Pape R, Workman CI (2018) A Multilevel Social Neuroscience Perspective on Radicalization and Terrorism. Social Neuroscience 13:511–529.

Decety J, Yoder KJ (2017) The Emerging Social Neuroscience of Justice Motivation. Trends in Cognitive Sciences, 21(1), 6-14.

Ding J (2020) Social Media: Threat to or Tool of Authoritarianism? Harvard International Review Available at:

https://hir.harvard.edu/social-media-threat-to-or-tool-of-authoritarianism/ [Accessed March 18, 2023].

Dunbar RIM (2016) Do Online Social Media Cut Through the Constraints that Limit the Size of Offline Social Networks?. Royal Society Open Science 3:150292 Available at: https://dx.doi.org/10.1098/rsos.150292 [Accessed March 18, 2023].

Enomoto CE, Douglas K (2019) Do Internet Searches for Islamist Propaganda Precede or Follow Islamist Terrorist Attacks? Economics & Sociology 12:233–247.

Esfandiari G (2022) Iran Accused Of Secretly Implementing Controversial Draft Internet Bill. RadioFreeEurope/RadioLiberty Available at: https://www.rferl.org/a/iran-internet-bill-controversy-secretly-implementing/32026313.html [Accessed March 18, 2023].

Esfandiari G, Zarghami M (2023) “This Revolution Is Still Alive”: A Growing Number Of Iranian Women Defy The Hijab Law After Months Of Protests. RadioFreeEurope/RadioLiberty Available at: https://www.rferl.org/a/iran-women-defy-hijab-law-revolution-rights/32308891.html [Accessed March 18, 2023].

FactNameh (2022) نگاهی بر کارزار چندرسانه‌ای #دولت_مردم در حمایت از دولت سیزدهم. factnamehcom Available at: https://factnameh.com/fa/fact-checks/2022-09-27-iran-progovernment-misinformation-social-media-campaign-dolatemardom [Accessed March 18, 2023].

Farah M (2004) Neuroethics: A Guide for the Perplexed. Dana Foundation Available at: https://dana.org/article/neuroethics-guide-for-the-perplexed/ [Accessed March 31, 2023].

Farah M (2012) Neuroethics: The Ethical, Legal, and Societal Impact of Neuroscience. *Annual Review of Psychology*, *63*(1), 571–591. https://doi.org/10.1146/annurev.psych.093008.100438

Farahany N (2023) The Battle for Your Brain. St. Martin’s Press.

Farisco M, Evers K, Salles A (2022) On the Contribution of Neuroethics to the Ethics and Regulation of Artificial Intelligence. Neuroethics 15.

Far TS (2023) Interview: Taking Walks Without Wearing Hijab in Iran. Human Rights Watch Available at: https://www.hrw.org/news/2023/03/07/interview-taking-walks-without-wearing-hijab-iran [Accessed March 18, 2023].

Fisher M (2022) The Chaos Machine: The Inside Story of How Social Media Rewired Our Minds and Our World. New York: Little Brown & Company.

Fullerton N (2022) Instagram vs. Reality: The Pandemic's Impact on Social Media and Mental Health. Penn Medicine Available at: https://www.pennmedicine.org/news/news-blog/2021/april/instagram-vs-reality-the-pandemics-impact-on-social-media-and-mental-health#:~:text=April%2029%2C%202021&text=COVID%2D19%20has%20limited%20in [Accessed March 18, 2023].

Ghobadi P (2022) Iran Protests: Iran's Gen Z "realise life can be lived differently." BBC News Available at: https://www.bbc.com/news/world-middle-east-63213745 [Accessed March 18, 2023].

Goering S et al. (2021) Recommendations for Responsible Development and Application of Neurotechnologies. Neuroethics 14:365–386.

Hampton KN, Goulet LS, Rainie L, Purcell K (2011) Social Networking Sites and Our Lives: How People's Trust, Personal Relationships, and Civic and Political Involvement Are Connected to Their Use of Social Networking Sites and Other Technologies. Pew Internet & American Life Project. Accessed on: December, 11, 2012.

Hassan A, Barber SJ (2021) The Effects of Repetition Frequency on the Illusory Truth Effect. Cognitive Research: Principles and Implications 6 Available at: https://cognitiveresearchjournal.springeropen.com/articles/10.1186/s41235-021-00301-5 [Accessed March 18, 2023].

Haynes T (2018) Dopamine, Smartphones & You: A Battle for Your Time. Science in the News Available at: https://sitn.hms.harvard.edu/flash/2018/dopamine-smartphones-battle-time/ [Accessed March 18, 2023].

Hill RA, Dunbar RIM (2003) Social Network Size in Humans. Human Nature 14:53–72 Available at: https://dx.doi.org/10.1007/s12110-003-1016-y.

Hoffner CA & Bond BJ (2022) Parasocial Relationships, Social Media, & Well-being. Current Opinion in Psychology, 101306.

Hosman LA, Wright JW (1987) The Effects of Hedges and Hesitations on Impression Formation in a Simulated Courtroom Context. Western Journal of Speech Communication 51:173–188.

Huckin T (2016) Propaganda Defined. Propaganda and Rhetoric in Democracy: History, Theory, Analysis, 118-136.

Human Rights Watch (2022) Iranian Society under Crackdown | Human Rights Watch. Available at: https://www.hrw.org/blog-feed/iranian-society-under-crackdown.

Huszár F, Ktena SI, O'Brien C, Belli L, Schlaikjer A, Hardt M (2022) Algorithmic amplification of Politics on Twitter. Proceedings of the National Academy of Sciences 119:e2025334119.

IEEE Brain (n.d.) Neurotechnologies: The Next Technology Frontier. IEEE Brain Available at: https://brain.ieee.org/topics/neurotechnologies-the-next-technology-frontier/.

International Crisis Group (2022) 10 Conflicts to Watch in 2023. International Crisis Group Available at: https://www.crisisgroup.org/visual-explainers/10-conflicts-2023/ [Accessed March 18, 2023].

Ioanes E (2022) Iran's Months-Long Protest Movement, explained. Vox Available at: https://www.vox.com/2022/12/10/23499535/iran-protest-movement-explained [Accessed March 18, 2023].

Jowett GS, O’Donnell V (1986) Propaganda and Persuasion. Sage.

Kanai R, Bahrami B, Roylance R, Rees G (2012) Online Social Network Size is Reflected in Human Brain Structure. Proceedings of the Royal Society B: Biological Sciences 279:1327–1334 Available at: https://dx.doi.org/10.1098/rspb.2011.1959.

Kauschke C, Bahn D, Vesker M, Schwarzer G (2019) The Role of Emotional Valence for the Processing of Facial and Verbal Stimuli—Positivity or Negativity Bias? Frontiers in Psychology 10.

Kemp S (2022) Digital 2022: Global Overview Report. DataReportal Available at: https://datareportal.com/reports/digital-2022-global-overview-report [Accessed March 18, 2023].

Kozyreva A, Lorenz-Spreen P, Hertwig R, Lewandowsky S, Herzog SM (2021) Public attitudes towards Algorithmic Personalization and Use of Personal Data Online: Evidence from Germany, Great Britain, and the United States. Humanities and Social Sciences Communications 8.

Krosnick J, Smith W (1994) Attitude Strength.

Lavin C, Melis C, Mikulan E, Gelormini C, Huepe D, Ibañez A. The Anterior Cingulate Cortex: An Integrative Hub for Human Socially-Driven Interactions. Front Neurosci. 2013 May 8;7:64. doi: 10.3389/fnins.2013.00064. PMID: 23658536; PMCID: PMC3647221.

Leavy P (2020) Oxford Handbook Of Qualitative Research. S.L.: Oxford Univ Press Us.

Lee SM, Henson RN, Lin CY (2020) Neural Correlates of Repetition Priming: A Coordinate-Based Meta-Analysis of FMRI Studies. Frontiers in Human Neuroscience, 14, 565114.

Lewis P (2017) ‘Our Minds Can Be Hijacked’: The Tech Insiders Who Fear a Smartphone Dystopia. The Guardian, October 6, 2017, sec. Technology. https://www.theguardian.com/technology/2017/oct/05/smartphone-addiction-silicon-valley-dystopia [Accessed March 18, 2023].

Lueck JA (2015) Friend-Zone with Benefits: The Parasocial Advertising of Kim Kardashian. Journal of Marketing Communications 21:91–109.

Mauri M, Cipresso P, Balgera A, Villamira M, Riva G (2011) Why Is Facebook So Successful? Psychophysiological Measures Describe a Core Flow State While Using Facebook. Cyberpsychology, Behavior, and Social Networking 14:723–731.

Meshi D, Tamir DI, Heekeren HR (2015) The Emerging Neuroscience of Social Media. Trends in Cognitive Sciences 19:771–782 Available at: http://smnlab.msu.edu/wp-content/uploads/2017/08/Meshi_2015_TICS.pdf [Accessed November 15, 2019].

Moll J, de Oliveira-Souza R, Bramati IE, Grafman J (2002) Functional Networks in Emotional Moral and Nonmoral Social Judgments. NeuroImage 16:696–703.

Müller O, Rotter S (2017) Neurotechnology: Current Developments and Ethical Issues. Frontiers in Systems Neuroscience 11 Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5733340/.

Munn L (2020) Angry by Design: Toxic Communication and Technical Architectures. Humanities and Social Sciences Communications 7.

Nasr V (2023) Iran's Hard-Liners Are Winning. Foreign Affairs Available at: https://www.foreignaffairs.com/iran/irans-hard-liners-are-winning?check_logged_in=1&utm_medium=promo_email&utm_source=lo_flows&utm_campaign=registered_user_welcome&utm_term=email_1&utm_content=20230314 [Accessed March 12, 2023].

O'Connor C, Weatherall J (2019) The Social Media Propaganda Problem Is Worse Than You Think. Issues in Science and Technology 36:30–32 Available at: https://www.jstor.org/stable/26949075 [Accessed March 18, 2023].

Odag Ö, Leiser A, Boehnke K (2019) Reviewing the Role of the Internet in Radicalization Processes. Journal for Deradicalization, (21), 261-300.

OECD (2019) OECD Legal Instruments. OECD Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0457#adherents [Accessed March 18, 2023].

Oliveira L, Azevedo J (2022) Using Social Media Categorical Reactions as a Gateway to Identify Hate Speech in COVID-19 News. SN Computer Science 4.

O'Sullivan PB, Carr CT (2017) Masspersonal Communication: A Model Bridging the Mass-Interpersonal Divide. New Media & Society 20:1161–1180.

Papakyriakopoulos O, Goodman E (2022) The Impact of Twitter Labels on Misinformation Spread and User Engagement: Lessons from Trump's Election Tweets. Proceedings of the ACM Web Conference 2022.

Parry-Giles SJ (2002) The Rhetorical Presidency, Propaganda, and the Cold War: 1945-1955. Westport, CT: Praeger.

Pew Research Center (2021) Social Media Fact Sheet. Pew Research Center.

Philips SU (1985) William M. O'Barr, Linguistic Evidence: Language Power, and Strategy in the Courtroom. (Studies on Law and Social Control.) New York: Academic, 1982. Pp. xv + 192. Language in Society 14:113–117.

Prado J, Léone J, Epinat-Duclos J, Trouche E, Mercier H (2020) The Neural Bases of Argumentative Reasoning. Brain and Language, 208, 104827.

PropWatch (n.d.) Propaganda Techniques. PropWatch Available at: https://www.propwatch.org/propaganda.php [Accessed March 18, 2023].

Racine E, Sample M (2019). Do We Need Neuroethics?. AJOB Neuroscience, 10(3), 101-103.

Reuters (2022) As Unrest grows, Iran Restricts Access to Instagram, WhatsApp. Reuters Available at: https://www.reuters.com/world/middle-east/iran-restricts-access-instagram-netblocks-2022-09-21/ [Accessed March 18, 2023].

Reviglio U, Agosti C (2020) Thinking Outside the Black-Box: The Case for "Algorithmic Sovereignty" in Social Media. Social Media + Society 6:205630512091561.

Ricci J (2018) The Growing Case for Social Media Addiction | CSU. wwwcalstateedu Available at: https://www.calstate.edu/csu-system/news/Pages/Social-Media-Addiction.aspx [Accessed March 18, 2023].

Rimé B (2009) Emotion Elicits the Social Sharing of Emotion: Theory and Empirical Review. Emotion Review 1:60–85.

Rivera H (2018) Red Herring. In, pp 208–211.

Roskies A (2016) Neuroethics Zalta EN, ed. Stanford Encyclopedia of Philosophy Available at: https://plato.stanford.edu/entries/neuroethics/ [Accessed March 31, 2023].

Roy D (2011) Neuroethics, Gender and the Response to Difference. Neuroethics 5:217–230.

Salgado S, Kaplitt MG (2015) The Nucleus Accumbens: A Comprehensive Review. Stereotactic and Functional Neurosurgery 93:75–93 Available at: https://www.karger.com/Article/FullText/368279#ref251 [Accessed October 30, 2022].

Sang ETK & Van den Bosch A (2013). Dealing with Big Data: The Case of Twitter. Computational Linguistics in the Netherlands Journal, 3, 121-134.

Sankaran V (2022) Twitter's New Community Notes Feature Lets People Add Context to Tweets. The Independent Available at: https://www.independent.co.uk/tech/twitter-community-notes-tweet-context-b2243311.html [Accessed March 08, 2023].

Sansone R, Sansone L (2012) Rumination: Relationships with Physical Health. Innovations in Clinical Neuroscience.

Santos JC, Matos S (2014) Analysing Twitter and Web Queries for Flu Trend Prediction. Theoretical Biology and Medical Modelling 11.

Schwarcz J (2020) McGill University. Office for Science and Society Available at: https://www.mcgill.ca/oss/article/covid-19-general-science/cherry-picking-era-covid-19 [Accessed March 02, 2023].

Shan LY (2022) Iran Says 40 foreigners arrested for Taking Part in Anti-Government Protests. CNBC Available at: https://www.cnbc.com/2022/11/22/iran-says-40-foreigners-arrested-for-taking-part-in-antigovernment-protests.html [Accessed February 14, 2023].

Simon J (2022) The Infodemic : How Censorship and Lies Made the World Sicker and Less Free. New York: Columbia Global Reports.

Sparks JR, Areni CS (2007) Style Versus Substance: Multiple Roles of Language Power in Persuasion. Journal of Applied Social Psychology 38:37–60.

Steinert S, Dennis MJ (2022) Emotions and Digital Well-Being: On Social Media's Emotional Affordances. Philosophy & Technology 35.

Svedholm-Häkkinen AM, Kiikeri M (2022) Cognitive Miserliness in Argument Literacy? Effects of Intuitive and Analytic Thinking on Recognizing Fallacies. Judgment and Decision Making 17:331–361.

Thompson R (2011) Radicalization and the Use of Social Media. Journal of Strategic Security 4:167–190.

Tufts University Relations (2022) Social Media Overview - Communications. Communications Available at: https://communications.tufts.edu/marketing-and-branding/social-media-overview/ [Accessed December 14, 2022].

Turak N (2020) Iran now has the highest coronavirus death toll outside of China, threatening the wider Middle East. CNBC Available at: https://www.cnbc.com/2020/02/25/iran-highest-coronavirus-death-toll-outside-china-accused-of-cover-up.html [Accessed March 18, 2023].

Twitter (n.d.) How We Address Misinformation on Twitter. Twitter Help Available at: https://help.twitter.com/en/resources/addressing-misleading-info [Accessed March 18, 2023].

Uncapher MR, Lin L, Rosen LD, Kirkorian HL, Baron NS, Bailey K, Cantor J, Strayer DL, Parsons TD, Wagner AD (2017) Media Multitasking and Cognitive, Psychological, Neural, and Learning Differences. Pediatrics 140:S62–S66.

UNESCO IBC (2020) Preliminary Draft Report of the IBC on Ethical Issues of Neurotechnology.

UNHCR (n.d.) Using Social Media in Community-Based Protection A Guide: January 2021. Available at: https://www.unhcr.org/innovation/wp-content/uploads/2021/01/Using-Social-Media-in-CBP.pdf .

United Nations (2022) Iran removed from UN Commission on the Status of Women. UN News Available at: https://news.un.org/en/story/2022/12/1131722 [Accessed March 18, 2023].

US Institute of Peace (2022) Protest Context: Statistics on Iran’s Women. The Iran Primer Available at: https://iranprimer.usip.org/blog/2022/oct/07/statistics-women-iran [Accessed March 18, 2023].

Van Volkinburg H, Balsam P (2014) Effects of Emotional Valence and Arousal on Time Perception. Timing & Time Perception 2:360–378.

Variam (1995) The Information Economy. Available at: https://people.ischool.berkeley.edu/~hal/pages/sciam.html [Accessed March 18, 2023].

Walther JB (1996) Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction. Communication Research 23:3–43.

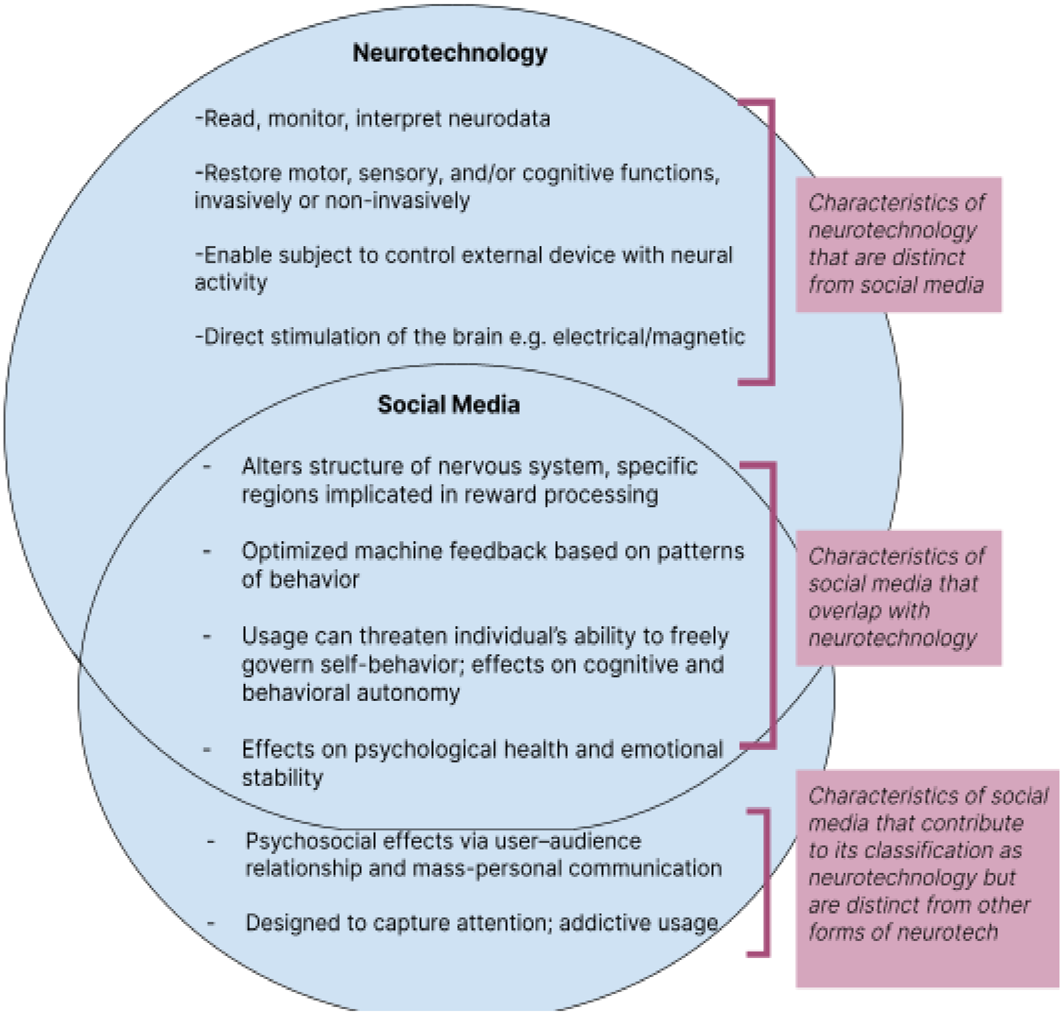
West BJ, Massari GF, Culbreth G, Failla R, Bologna M, Dunbar RIM, Grigolini P (2020) Relating Size and Functionality in Human Social Networks Through Complexity. Proceedings of the National Academy of Sciences 117:18355–18358.

Williams J (2018) Stand Out of Our Light: Freedom and Resistance in the Attention Economy. Cambridge, United Kingdom Cambridge University Press.

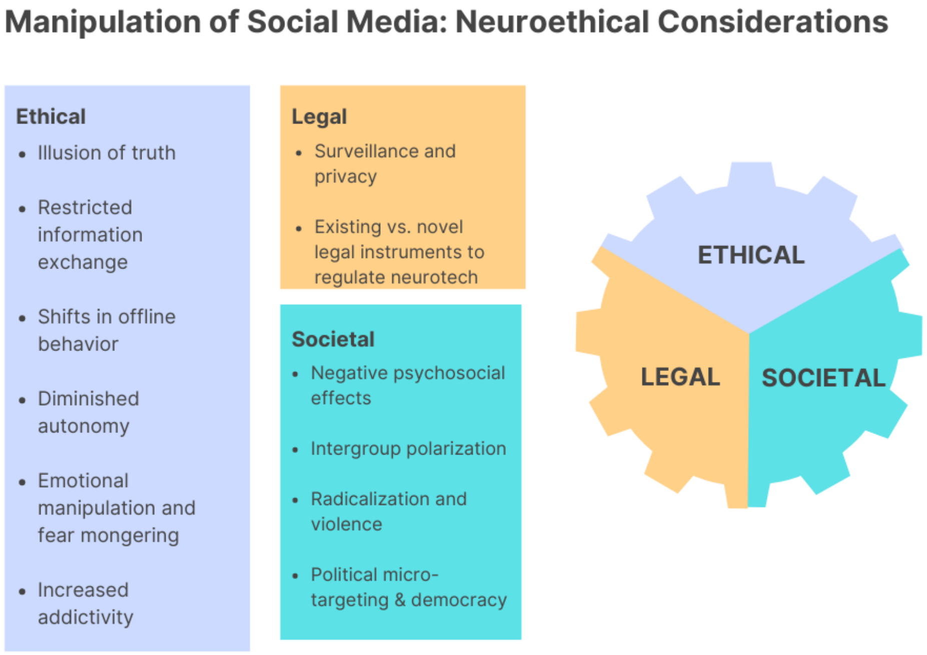
Ziabari K (2022) Iran's Leaders Are Scared of the Internet. Foreign Policy Available at: https://foreignpolicy.com/2022/06/06/iran-internet-protection-bill-curbs-restrictions-unrest/ [Accessed January 7, 2023].

Figures Listed

1. Features of social media that contribute to its classification as neurotechnology and distinguish it from other forms of neurotechnology. (Pg. 6)


Neurotechnology
-Read, monitor, interpret neurodata
-Restore motor, sensory, and/or cognitive functions, invasively or non-invasively
-Enable subject to control external device with neural activity
-Direct stimulation of the brain e.g. electrical/magnetic
Characteristics of neurotechnology that are distinct from social media
Social Media
- Alters structure of nervous system, specific regions implicated in reward processing
- Optimized machine feedback based on patterns of behavior
- Usage can threaten individual's ability to freely govern self-behavior; effects on cognitive and behavioral autonomy
- Effects on psychological health and emotional stability
Characteristics of social media that overlap with neurotechnology
- Psychosocial effects via user–audience relationship and mass-personal communication
- Designed to capture attention; addictive usage
Characteristics of social media that contribute to its classification as neurotechnology but are distinct from other forms of neurotech

2. Ethical, legal, societal implications of the manipulation of social media. (Pg. 24)



Manipulation of Social Media: Neuroethical Considerations

Ethical

- Illusion of truth
- Restricted information exchange
- Shifts in offline behavior
- Diminished autonomy
- Emotional manipulation and fear mongering
- Increased addictivity

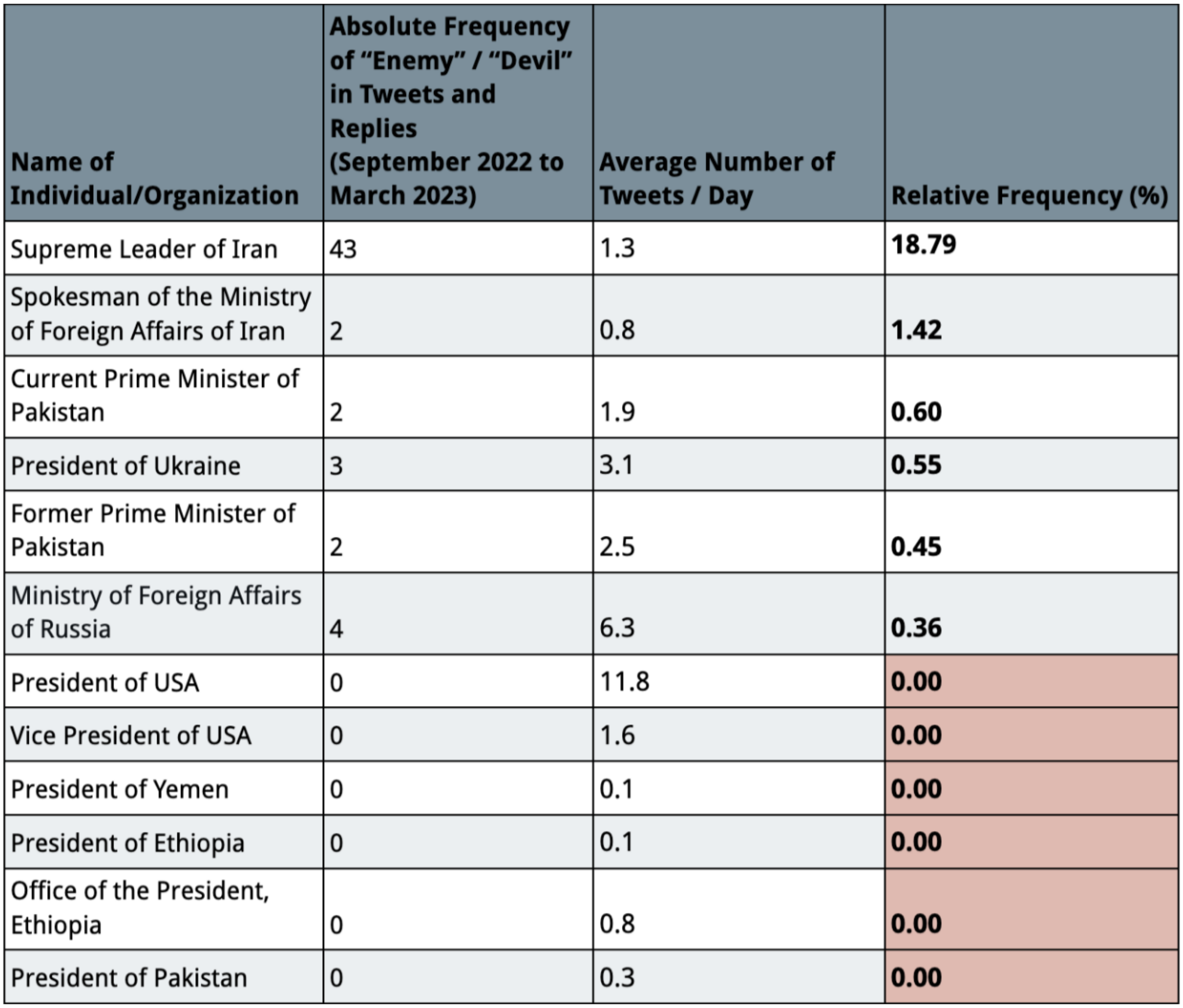
Legal

- Surveillance and privacy
- Existing vs. novel legal instruments to regulate neurotech

Societal

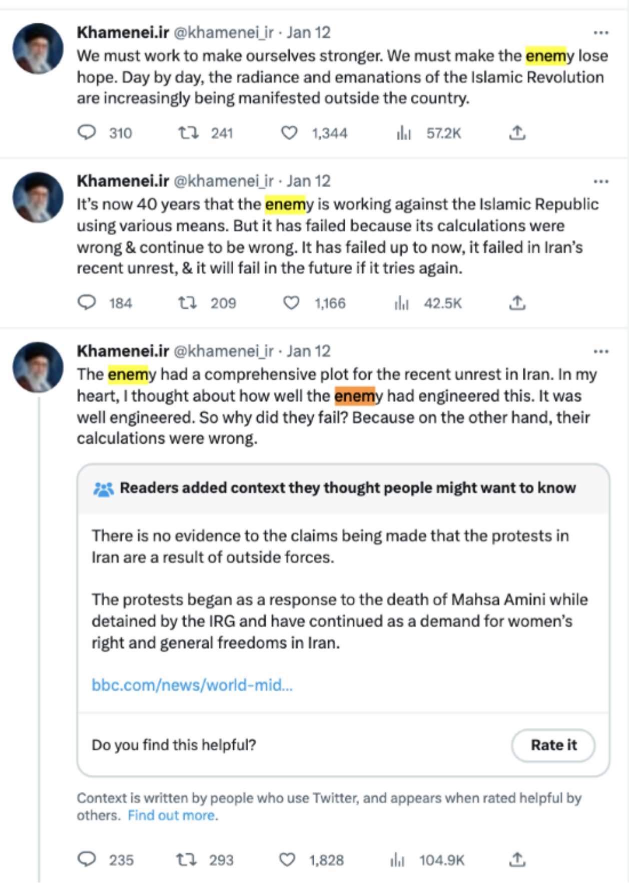
- Negative psychosocial effects
- Intergroup polarization
- Radicalization and violence
- Political micro-targeting & democracy

3. Absolute and relative frequencies of “enemy”/“devil” for Iranian state official accounts and state official accounts from other countries in similar states of political unrest. (Pg. 34)



Name of Individual/Organization	Absolute Frequency of “Enemy” / “Devil” in Tweets and Replies (September 2022 to March 2023)	Average Number of Tweets / Day	Relative Frequency (%)
Supreme Leader of Iran	43	1.3	**18.79**
Spokesman of the Ministry of Foreign Affairs of Iran	2	0.8	**1.42**
Current Prime Minister of Pakistan	2	1.9	**0.60**
President of Ukraine	3	3.1	**0.55**
Former Prime Minister of Pakistan	2	2.5	**0.45**
Ministry of Foreign Affairs of Russia	4	6.3	**0.36**
President of USA	0	11.8	**0.00**
Vice President of USA	0	1.6	**0.00**
President of Yemen	0	0.1	**0.00**
President of Ethiopia	0	0.1	**0.00**
Office of the President, Ethiopia	0	0.8	**0.00**
President of Pakistan	0	0.3	**0.00**

4. Sample tweets demonstrating the repetition of the keyword "Enemy" in English and Farsi. (Pg. 36)



Khamenei.ir @khamenei_ir · Jan 12

We must work to make ourselves stronger. We must make the enemy lose hope. Day by day, the radiance and emanations of the Islamic Revolution are increasingly being manifested outside the country.

Khamenei.ir @khamenei_ir · Jan 12

It's now 40 years that the enemy is working against the Islamic Republic using various means. But it has failed because its calculations were wrong & continue to be wrong. It has failed up to now, it failed in Iran's recent unrest, & it will fail in the future if it tries again.

Khamenei.ir @khamenei_ir · Jan 12

The enemy had a comprehensive plot for the recent unrest in Iran. In my heart, I thought about how well the enemy had engineered this. It was well engineered. So why did they fail? Because on the other hand, their calculations were wrong.

Readers added context they thought people might want to know

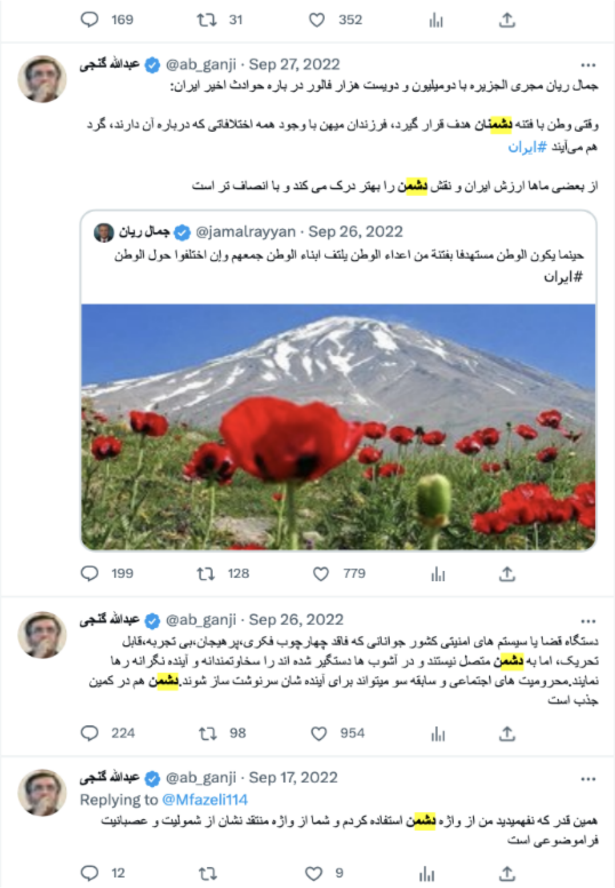
There is no evidence to the claims being made that the protests in Iran are a result of outside forces.

The protests began as a response to the death of Mahsa Amini while detained by the IRG and have continued as a demand for women's right and general freedoms in Iran.

bbc.com/news/world-mid...

Do you find this helpful?

Context is written by people who use Twitter, and appears when rated helpful by others. Find out more.



عبدالله گنجی @ab_ganji · Sep 27, 2022

جمال ریان مجری الجزیره با دومیلیون و دویست هزار فالور در باره حوادث اخیر ایران:

وقتی وطن با فتنه دشمنان هدف قرار گیرد، فرزندان میهن با وجود همه اختلافاتی که درباره آن دارند، گرد هم می‌آیند #ایران

از بعضی ماها ارزش ایران و نقش دشمن را بهتر درک می کند و با انصاف تر است

جمال ريان @jamalrayyan · Sep 26, 2022

حينما يكون الوطن مستهدفا بفتنة من اعداء الوطن يلتف ابناء الوطن جمعهم وإن اختلفوا حول الوطن #ايران

عبدالله گنجی @ab_ganji · Sep 26, 2022

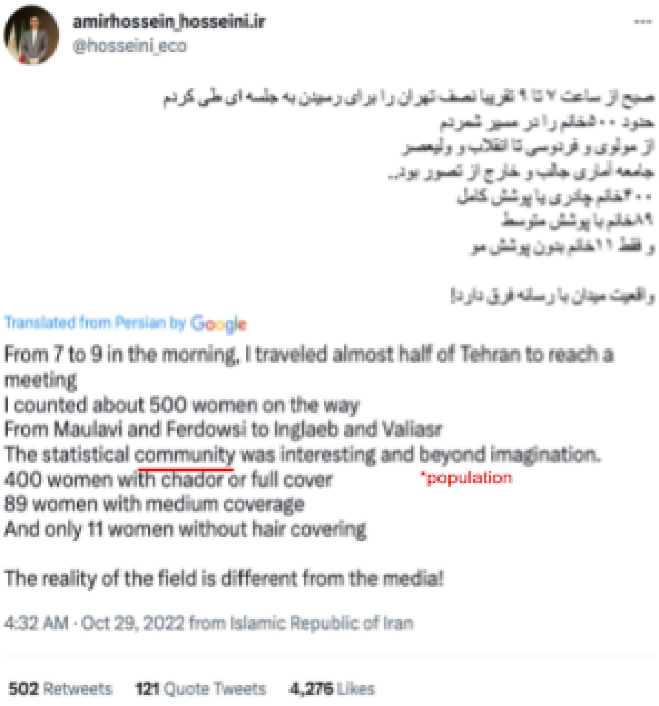
دستگاه قضا یا سیستم های امنیتی کشور جوانانی که فاقد چهارچوب فکری،پر هیجان،بی تجربه،قابل تحریک، اما به دشمن متصل نیستند و در آشوب ها دستگیر شده اند را سخاوتمندانه و آینده نگرانه رها نمایند.محرومیت های اجتماعی و سابقه سو میتواند برای آینده شان سرنوشت ساز شوند.دشمن هم در کمین جذب است

عبدالله گنجی @ab_ganji · Sep 17, 2022

Replying to @Mfazeli114

همین قدر که نفهمیدید من از واژه دشمن استفاده کردم و شما از واژه منتقد نشان از شمولیت و عصبانیت فراموضوعی است

5. Sample tweets by a state official and an affiliated journalist, @hossein_eco and @h_ABBASIFAR, using anecdotal evidence to make generalized claims about Iranian women's compliance with the mandatory hijab law. (Pg. 41)



amirhossein_hosseini.ir
@hosseini_eco

صبح از ساعت ۷ تا ۹ تقریبا نصف تهران را برای رسیدن به جلسه ای طی کردم
حدود ۵۰۰ خانم را در مسیر شمردم
از مولوی و فردوسی تا انقلاب و ولیعصر
جامعه آماری جالب و خارج از تصور بود..
۴۰۰ خانم چادری یا پوشش کامل
۸۹ خانم با پوشش متوسط
و فقط ۱۱ خانم بدون پوشش مو

واقعیت میدان با رسانه فرق دارد!

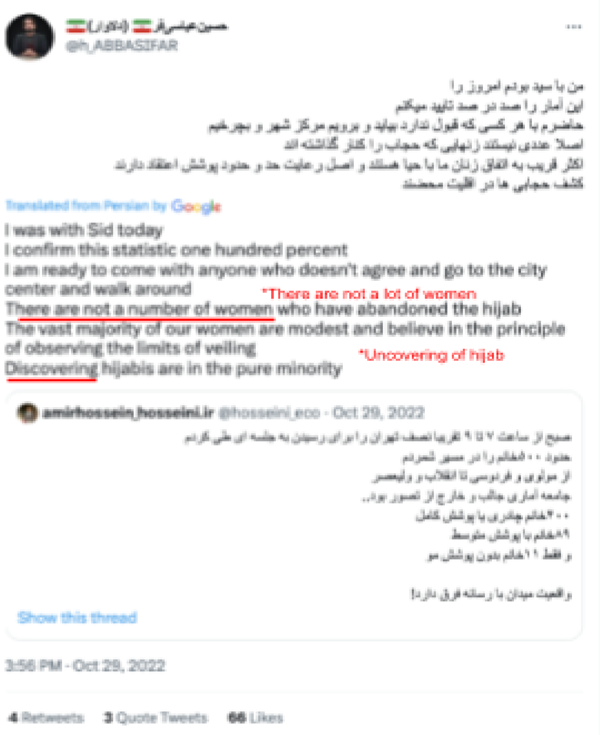
Translated from Persian by Google

From 7 to 9 in the morning, I traveled almost half of Tehran to reach a meeting
I counted about 500 women on the way
From Maulavi and Ferdowsi to Inglaeb and Valiasr
The statistical community was interesting and beyond imagination.
400 women with chador or full cover *population
89 women with medium coverage
And only 11 women without hair covering

The reality of the field is different from the media!

4:32 AM · Oct 29, 2022 from Islamic Republic of Iran

502 Retweets 121 Quote Tweets 4,276 Likes



حسین عباسی‌فر 🇮🇷(داوار)🇮🇷
@h_ABBASIFAR

من با سید بودم امروز را
این آمار را صد در صد تایید میکنم
حاضرم با هر کسی که قبول ندارد بیاید و برویم مرکز شهر و بچرخیم
اصلا عددی نیستند زنهایی که حجاب را کنار گذاشته اند
اکثر قریب به اتفاق زنان ما با حیا هستند و اصل رعایت حد و حدود پوشش اعتقاد دارند
کشف حجابی ها در اقلیت محضند

Translated from Persian by Google

I was with Sid today
I confirm this statistic one hundred percent
I am ready to come with anyone who doesn't agree and go to the city center and walk around
*There are not a lot of women
There are not a number of women who have abandoned the hijab
The vast majority of our women are modest and believe in the principle of observing the limits of veiling
*Uncovering of hijab
Discovering hijabis are in the pure minority

amirhossein_hosseini.ir @hosseini_eco · Oct 29, 2022

صبح از ساعت ۷ تا ۹ تقریبا نصف تهران را برای رسیدن به جلسه ای طی کردم
حدود ۵۰۰ خانم را در مسیر شمردم
از مولوی و فردوسی تا انقلاب و ولیعصر
جامعه آماری جالب و خارج از تصور بود..
۴۰۰ خانم چادری یا پوشش کامل
۸۹ خانم با پوشش متوسط
و فقط ۱۱ خانم بدون پوشش مو

واقعیت میدان با رسانه فرق دارد!

Show this thread

3:56 PM · Oct 29, 2022

4 Retweets 3 Quote Tweets 66 Likes

6. Ethical, legal, societal implications of social media manipulation and the spread of digital propaganda during the Iran protests. (Pg. 44)

Ethical Implications	Legal Implications	Societal Implications
● Online manipulation of vulnerable populations within/outside of Iran ● Diminished user autonomy due to censorship and filtration ● Self-censorship; restricted freedom of thought and expression ● Spread of emotionally salient disinformation	● Applications of existing legal instruments for regulating neurotechnology ● Policy action for greater algorithmic sovereignty on Twitter	● Distorted public perception / limited awareness of protests ● Collective impact of manipulative technologies on decision-making ● Precedent for other totalitarian regimes and political microtargeting on social media ● International tensions; animosity towards the US and European countries

Appendix

Appendix A

Absolute and Relative Frequencies for “Enemy”/ “Devil”

Twitter Account Handle	Follower Count	Absolute Frequency of “Enemy” / “Devil” in Tweets & Replies	Average Number of Tweets / Day	Estimated Relative Frequency (%)
@Panahian_IR	255K	4	0.1	**22.73**
@khamenei_ir	961.4K	43	1.3	**18.79**
@Khamenei_fa	629.4K	20	0.7	**16.23**
@mb_ghalibaf	274.9K	8	0.3	**15.15**
@saeid_mohammad_	33.1K	2	0.1	**11.36**
@raisi_com	223.1K	2	0.1	**11.36**
@ehsan_sa	1.6K	9	0.5	**10.23**
@EnsiyehKhazali	5.5K	3	0.2	**8.52**
@Khamenei_m	8.9K	30	2.5	**6.82**
@shojaeiam	2.1K	11	1.0	**6.25**
@khameneireyhane	21.9K	2	0.2	**5.68**
@Ahmadnaderi_ir	36.1K	6	0.7	**4.87**
@Iran_GOV	26.1K	52	6.7	**4.41**
@khameneinews	8.5K	49	6.5	**4.28**
@a_dastaran	14K	51	7.1	**4.08**
@alibahaadori	89.8K	6	0.9	**3.79**
@nezammousavi	82.6K	7	1.1	**3.62**
@Tasnimnews_Fa	255.8K	236	47.0	**2.85**
@Amirabdolahian	168.9K	3	0.6	**2.84**
@TehranTimes79	37.6K	110	22.0	**2.84**
@h_ABBASIFAR	7.2K	50	10.0	**2.84**
@3eyedamir	40.4K	13	2.9	**2.55**
@mmohammadii61	96.1K	31	7.3	**2.41**
@IRNA_1313	72.2K	68	16.9	**2.29**

@jamejamCPI	60K	163	46.6	**1.99**
@MashreghNews	39K	6	1.9	**1.79**
@ab_ganji	117.1K	49	15.8	**1.76**
@IRIMFA_SPOX	7K	2	0.8	**1.42**
@mjakhavan	1.9K	8	3.4	**1.34**
@sabeti_twt	276.9K	10	4.6	**1.24**
@hosseini_eco	3.3K	24	11.2	**1.22**
@hamshahrinews	38.8K	17	8.7	**1.11**
@jamarannews	43.1K	15	8.6	**0.99**
@IranNewspaper	143.6K	61	48.3	**0.72**
@khabaronlinee	111.6K	6	5.1	**0.67**
@ilnanews	108.7K	9	8.2	**0.62**
@mehrnews_ir	32.1K	8	9.2	**0.49**
@PanahianEN	32.3K	1	1.8	**0.32**

Table A1. Twitter accounts with estimated relative frequencies > 0.0% for the keywords "Enemy" or "Devil" in English and Farsi from September 15, 2022 to March 9, 2023. Relative frequency was calculated by [(Total Keyword Count) / (Avg. Tweet Count * 176)) * 100] where Avg. Tweet Count refers to the average tweets per day for the last 1000 tweets on and before March 9, 2023 and 176 is the total number of days between the start and end date. Twitter Advanced Search was used to identify absolute frequency of keyword appearances using input "(enemy OR enemies OR devil OR devils OR دشمن OR دشمنان OR شیطان OR شیاطین) (from:[account]) until:2023-03-09 since:2022-09-15".

Twitter Account Handle	Follower Count	Name of Individual/Organization	Absolute Frequency of “Enemy” / “Devil”	Average Number of Tweets / Day	Estimated Relative Frequency (%)
@khameini_ir	961.4K	Supreme Leader of Iran	43	1.3	**18.79**
@IRIMFA_SPOX	7K	Nasser Kanaani, Spokesman of the Ministry of Foreign Affairs of Iran	2	0.8	**1.42**
@CMShehbaz	6.5M	Shehbaz Sharif, Current Prime Minister of Pakistan	2	1.9	**0.60**
@ZelenskyyUa	7.1M	Volodymyr Zelenskyy, President of Ukraine	3	3.1	**0.55**
@ImranKhanPTI	18.8M	Imran Khan, Former Prime Minister of Pakistan	2	2.5	**0.45**
@mfa_russia	579.2K	Ministry of Foreign Affairs of Russia (English)	4	6.3	**0.36**
@POTUS	29.9M	Joe Biden, President of USA	0	11.8	**0.00**
@KamalaHarris	20M	Kamala Harris, Vice President of USA	0	1.6	**0.00**

@PresidentRashad	126.9K	Rashad Muhammad al-Alimi, President of Yemen	0	0.1	**0.00**
@SahleWorkZewde	547.3K	Sahle Work Zewde, President of Ethiopia	0	0.1	**0.00**
@POEthiopia	66.9K	Office of the President, Ethiopia	0	0.8	**0.00**
@ArifAlvi	4.2M	Arif Alvi, President of Pakistan	0	0.3	**0.00**

Table A2. Absolute and relative frequencies of “enemy”/“devil” for Iranian state official accounts and state official accounts from other countries in similar states of political unrest.

Appendix B

Complete List of Twitter Accounts Analyzed

Twitter Account Handle	Name of Individual/Organization	Description of Affiliation	Type of Affiliation
@Ahmadnaderi_ir	Ahmad Naderi	Member of the Presidium of the Islamic Council	**State Official**
@alibahaadori	Ali Bahadori Jahromi	Spokesperson of the Government of the I.R.I	**State Official**
@Amirabdolahian	Hossein Amir-Abdollahian	Minister of Foreign Affairs of Iran	**State Official**
@EnsiyehKhazali	Ensiyeh Khazali	VP for Women and Family Affairs	**State Official**
@hosseini_eco	Amirhossein Hosseini	Member of Tehran Chamber of Commerce	**State Official**
@IRIMFA_SPOX	Nasser Kanaani	Spokesman of the Ministry of Foreign Affairs of Iran	**State Official**
@khabaronlinee	KhabarOnline News Agency		**State Official**
@Khamenei_fa	Supreme Leader of Iran (Farsi)		**State Official**
@khamenei_ir	Supreme Leader of Iran (English)		**State Official**
@Khamenei_m	Supreme Leader of Iran (Media Account)		**State Official**
@khameneinews	Supreme Leader of Iran (News Updates Account)		**State Official**
@khameneireyhane	Supreme Leader of Iran (Women & Family Issues)		**State Official**
@mb_ghalibaf	Mohammad Bagher Ghalibaf	Speaker of the Parliament of Iran	**State Official**
@MhmmdJamshidi	Mohammad Jamshidi	Deputy Chief of Staff for Political Affairs to I.R. Iran President	**State Official**
@mmohammadii61	Mahdi Mohammadi	Iranian National Security Analyst, Advisor to the Speaker of the Parliament in Strategic Affairs	**State Official**
@raisi_com	Ebrahim Raisi – President of Iran		**State**

			Official
@saeid_mohammad_	Saied Mohammad – Senior ranking member of Islamic Revolutionary Guard Corps (IRGC), 2021 presidential candidate		**State Official**
@syjebraily1	Seyed Yasser Jebraily	Chief, Centre for Strategic Assessment of Implementing the Macropolicies of I.R.I.	**State Official**
@yaminpour	Vahid Yaminpour	Deputy of the Ministry of Sports and Youth	**State Official**
@a_dastaran	Ahmad Dastaran	Writer, Fars News Guest Contributor	**Public Figure**
@hamedkashani_	Hamed Kashani		**Public Figure**
@Panahian_IR	Ali-Reza Panahian	Twelver Shia Scholar, Head of the Supreme Leader's Think Tank for Universities	**Public Figure**
@PanahianEN	Ali-Reza Panahian (English)		**Public Figure**
@hafezeh_tarikhi			**Pro-Govt. Content**
@3eyedamir	Seyed Amir Syah	Editor in Chief of Alef News Agency, Affiliated with the Islamic Council Research Council	**Journalist**
@ab_ganji	Abdollah Ganji	Former editor of the IRGC-linked Javan newspaper, Chief Editor of Hamshahri Newspaper	**Journalist**
@ehsan_sa	Ehsan Salehi	Media Personality, Formerly Affiliated with Raja News	**Journalist**
@h_ABBASIFAR	Hossein Abbasifar	Producer of Radio Javan, Affiliated with state-run cultural media group "Seda"	**Journalist**
@mjakhavan	Director of Javan News Agency		**Journalist**
@nezammousavi	Seyed Nizamuddin Mousavi	Former CEO of Fars News Agency, Member of Muslim Journalists Association, Appointed Head of Islamic	**Journalist**

		Propaganda Coordination Council in 2021	
@sabeti_twt	Amirhossein Sabeti - TV Host of Jahan Ara and Media Activist		**Journalist**
@shojaeiam	Mohammad Shojaeian	Managing Director of the Tehran Times and the Mehr News Agency	**Journalist**
@hamshahrinews	Hamshahri Media & News Company	Published by the Tehran City government	**Affiliated Media**
@ilnanews	Iranian Labor News Agency		**Affiliated Media**
@Iran_GOV	Government of the Islamic Republic of Iran		**Affiliated Media**
@IranNewspaper	Iran Newspaper	Daily Newspaper of the Iranian government	**Affiliated Media**
@IRNA_1313	Islamic Republic News Agency		**Affiliated Media**
@jamarannews	Jamaran News Agency		**Affiliated Media**
@jamejamCPI	Jam-e Jam Newspaper	Daily newspaper published by the ublished by Islamic Republic of Iran Broadcasting	**Affiliated Media**
@MashreghNews	Mashregh News	Affiliate of the Revolutionary Guards' Intelligence Organization (IRGC-IO)	**Affiliated Media**
@mehrnews_ir	Mehr News Agency		**Affiliated Media**
@Tasnimnews_Fa	Tasnim News Agency	Affiliated with the IRGC	**Affiliated Media**
@TehranTimes79	Tehran Times	State-Affiliated International Newspaper	**Affiliated Media**

Table B. Full appendix of the 41 Twitter accounts that were analyzed, including 18 state official accounts, 10 affiliated media accounts, 8 journalist accounts, 4 public figure accounts, and 1 pro-IR content account.