What Makes Disinformation Ads Engaging? A Case Study of Facebook Ads from the Russian Active Measures Campaign

Mirela Silva[a*], Luiz Giovanini[b*], Juliana Fernandes[c*], Daniela Oliveira[a], and Catia S. Silva[a]

[a]Herbert Wertheim College of Engineering, University of Florida, Gainesville, Florida, USA; [b]College of Education, University of Florida, Gainesville, Florida, USA; [c]College of Journalism and Communications, University of Florida, Gainesville, Florida, USA

ABSTRACT

This article examines 3,517 Facebook ads created by Russia's Internet Research Agency (IRA) between June 2015 and August 2017 in its Active Measures disinformation campaign targeting the 2016 U.S. presidential election. We aimed to unearth the relationship between ad engagement (ad clicks) and 40 features related to the ads' metadata, psychological meaning, and sentiment. The purpose of our analysis was to (1) understand the relationship between engagement and features, (2) find the most relevant feature subsets to predict engagement via feature selection, and (3) find the semantic topics that best characterize the data set via topic modeling. We found that investment features (e.g., ad spend, ad lifetime), caption length, and sentiment were the top features predicting users' engagement with the ads. In addition, positive sentiment ads were more engaging than negative ads, and psycholinguistic features (e.g., use of religion-relevant words) were identified as highly important in the makeup of an engaging disinformation ad. Linear support vector machines (SVMs) and logistic regression classifiers achieved the highest mean *F* scores (93.6%), revealing that the optimal feature subset contains 12 and six features, respectively. Finally, we corroborate the findings of previous research that the IRA specifically targeted Americans on divisive ad topics (e.g., LGBT rights) and advance a definition of disinformation advertising.

KEYWORDS
Advertisements; disinformation; Facebook; engagement; machine learning

Disinformation is any "false, inaccurate, or misleading information designed, presented, and promoted to intentionally cause public harm or for profit" (European Commission 2018, p. 3). Numerous scholars have identified disinformation's implications for society at large, such as the spread of propaganda (Tandoc, Lim, and Ling 2018), promotion of societal division (Mihailidis and Viotty 2017), decreased trust in social media and media in general (Wagner and Boczkowski 2019), and casting doubt in democratic processes and government institutions (Mcnair 2017; Morgan 2018; Recuero, Soares, and Vinhas 2020). Although previous research has not offered a formal definition of a "disinformation ad," we rely on the general definition of disinformation provided here and extend it by suggesting that it involves content that is paid for or sponsored by advertisers, organizations, or individuals. Therefore, we propose that ads or sponsored content paid for by an entity (i.e., advertiser, organization, individuals) on social media with the goal of spreading false, misleading, or inaccurate information can be considered disinformation ads.

The disinformation phenomenon is not new. The Cold War Active Measures campaigns (Sullivan 2021) bear a disturbing resemblance to what we are witnessing today (Barela and Duberry 2021). Our society has now evolved, making room for social media to become the 21st-century version of Cold War balloons

CONTACT Juliana Fernandes juliana@jou.ufl.edu Department of Advertising, College of Journalism and Communications, University of Florida, Weimer Hall 3056, Gainesville, FL 32611, USA.
Mirela Silva (PhD, University of Florida) was a PhD Candidate (at the time of this research), Department of Electrical and Computer Engineering, Herbert Wertheim College of Engineering, University of Florida.
Luiz Giovanini (PhD, Pontifical Catholic University of Paraná) is Manager, Data and Reporting, New Worlds Reading Initiative, College of Education, University of Florida.
Juliana Fernandes (PhD, University of Florida) is an assistant professor, Department of Advertising, College of Journalism and Communications, University of Florida.
Daniela Oliveira (PhD, University of California, Davis) is an associate professor, Department of Electrical and Computer Engineering, Herbert Wertheim College of Engineering, University of Florida.
Catia S. Silva (PhD, University of Florida) is an instructional assistant professor, Department of Electrical and Computer Engineering, Herbert Wertheim College of Engineering, University of Florida.
*The first three authors contributed equally.

spreading disinformation by releasing leaflets from the sky. On social media, any individual or organization can purchase ads or promote sponsored content and distribute it widely to a desired audience. While there are no physical leaflets falling from the sky, social media posts and online content can spread very quickly and reach a large audience (Vosoughi, Roy, and Aral 2018) in a matter of hours.

To understand what makes disinformation ads engaging on social media, this study analyzes a sample of more than 3,000 Facebook ads identified as disinformation by the U.S. House of Representatives Permanent Select Committee on Intelligence (HPSCI n.d.). These ads were documented in the 2019 Mueller report, which revealed that Internet Research Agency (IRA; associated with the Kremlin) employees traveled to the United States in 2014 on an intelligence-gathering mission to better understand American culture and use their findings in social media posts. Notably, during the 2016 U.S. presidential election, as many as 529 different rumors were spread on Twitter (Jin et al. 2017), and approximately 80,000 social media ads were identified by the U.S. HPSCI (n.d.) as disinformation ads released by Russian actors. The intent of these ads was to interfere with the 2016 U.S. presidential campaign and sow division in American society by exploring issues such as race (e.g., Black Lives Matter advocacy), Second Amendment rights, and immigration (Mueller 2019; DiResta et al. 2019).

Analysis of this campaign is essential for at least three reasons: First, understanding the harmful effects of such campaigns at a societal level is important to keep these detrimental campaigns from taking place in the future. For example, the U.S. Senate Select Committee on Intelligence (2019) report on Active Measures in social media highlights that IRA activity on social media did not cease but rather increased after Election Day 2016, as if the results emboldened the Russian government (Hindman and Barash 2018). Moreover, reports have shown that foreign states such as Russia, China, and Iran targeted the Donald Trump and Joe Biden 2020 election campaigns in the United States, using similar techniques as those employed by the IRA in 2016 (BBC News 2020). Second, it is important to understand how disinformation ads impact social media users and voters by analyzing what specific features of these ads garnered high engagement. Third, although much research has been conducted on deceptive content in advertising (Amazeen and Wojdynski 2019; Chaouachi and Rached 2012; Fernandes, Segev, and Leopold 2020; Gardner 1975), disinformation ads on social media have not been explored from a scholarly standpoint. Therefore, it is important to advance knowledge of what features of these ads are appealing to social media users.

The purpose of this study is to understand what features of disinformation ads individuals are more likely to engage with using the data set made available by the HPSCI (n.d.) containing 3,517 Facebook ads from June 2015 to August 2017. To investigate engagement, we rely on theoretical conceptualizations of engagement as a behavior, as previous research has established that it reflects a concrete and measurable metric of action with an ad on social media (Dolan et al. 2016; Eigenraam et al. 2018; Muntinga, Moorman, and Smit 2011; van Doorn et al. 2010). Based on this conceptualization, the number of clicks reflects an ad's pertinence and actual users' engagement. Furthermore, to predict engagement with these disinformation ads, we identified four broad categories of features that have been used in previous research (Aldous et al. 2019; Baltas 2003; Munaro et al. 2021; Rambocas and Pacheco 2018; Vosoughi et al. 2018) and have shown potential predictive value: (1) investment features, (2) the size of the ad's caption, (3) the ad caption's psycholinguistic features, and (4) the ad caption's subjectivity and sentiment.

The remainder of this article is structured as follows: First, we present the research context focusing on the findings from the Mueller report (Mueller 2019), followed by a discussion of the engagement concept in social media and advertising and the formulation of our research questions. Next, we describe our data set and analysis approach and present our results. Finally, we discuss our findings, implications for theory and research in social media and advertising, limitations of our study, and directions for future research.

Theoretical Background

Research Context: Social Media As a Tool for Russian Interference in the 2016 U.S. Presidential Election

Investigations and reports on Russian efforts to influence the 2016 U.S. presidential election emerged as early as mid-2016 via the Federal Bureau of Investigation (FBI) Crossfire Hurricane investigation and after U.S. Congress members had access to classified intelligence (Miller 2016). The intelligence community uncovered that Russian president Vladimir Putin had ordered an influence campaign using social media to hurt Hillary Clinton's election chances and undermine public faith in the U.S. democratic process (Miller and Entous 2017). Congress then sought the aid of experts and social media companies in facilitating its public hearings and investigations. In September 2017, the media started reporting (Strohm 2017) that the Mueller probe was focused on the use of social media as the main tool of the Active Measures campaign. This prompted social media companies to conduct internal audits, which led to a data set of tweets, Facebook ads and posts, and YouTube videos being released to the HPSCI.

The IRA, supported by the Kremlin, conducted a major Active Measures campaign in the years preceding the 2016 presidential election, with their social media stimuli reaching millions of Americans (Howard et al. 2018, 2019; DiResta et al. 2019). They had two main goals: (1) influence the 2016 U.S. presidential election by harming Hillary Clinton's chances of success while supporting then-candidate Donald Trump and (2) sow discord in American politics and society, especially on race issues by heavily targeting the African American population, while playing both sides of the political discourse (also corroborated by independent work from Arif et al. 2018).

As a result, the HPSCI released 3,517 Facebook ads associated with the IRA in 2018 for public access, which have been analyzed qualitatively and quantitatively in intelligence reports, by independent researchers, and by the media (Penzenstadler, Heath, and Guynn 2018). Although the ads were not the bulk of the IRA's activity on social media, the use of advertising was consistent with the IRA's modus operandi (Select Committee on Intelligence 2019): divisive subjects related to race, police brutality, Second Amendment rights, patriotism, LGBTQ + rights, and immigration (Kim 2018). In a U.S. Census–representative survey, Ribeiro et al. (2019) found that people from different socially salient groups reacted differently to the content of the IRA's Facebook ads, further positing that Facebook's ad application programming interface (API) facilitated this divisive targeting. Indeed, Facebook estimates that 11.4 million Americans saw at least one of the ads determined to have been purchased by the IRA (Select Committee on Intelligence 2019). Thus, it is important to understand what features of these ads were the most appealing to users and predicted engagement with them. The next section delves into the concept of engagement.

Predicting Engagement with Disinformation Ads

A deep understanding of the concept of engagement is imperative to effectively measure disinformation with advertising on social media. In his account of Soviet disinformation tactics, Bittman (1985) discussed the two ways by which the KGB measured the success of disinformation campaigns. One metric to measure disinformation was determining whether the message forced the target country to make any political changes that could directly or indirectly benefit the Soviet Union. The second metric was the attention that the message was drawing outside the Soviet bloc, such as the amount of public discussion generated by the message and the tone of the political discourse on the issue (Bittman 1985). In the 21st century, this metric is what online platforms call *engagement* (Meta n.d.).

The engagement concept (and its diverse definitions) has received considerable attention from scholars in a variety of fields ranging from marketing and advertising (Gavilanes, Flatten, and Brettel 2018; Greenwald and Leavitt 1984; Jiang et al. 2022) to human–robot interaction (Rich et al. 2010) to education (Reeve et al. 2004) and game-based learning (Garris, Ahlers, and Driskell 2002). In marketing and advertising, specifically, engagement was considered a research priority for 2016 to 2018 by the Marketing Science Institute (MSI 2016) due to its importance for brand communication efforts and the propagation of social media platforms.

Previous research has suggested that conceptualizing engagement is a complex endeavor (Calder, Malthouse, and Schaedel 2009; Voorveld et al. 2018), and this is evident in the different research streams that emerged since its earlier conceptualization. For instance, the pioneering work by Brodie et al. (2011) conceptualized engagement as a psychological state of mind encompassing cognitive, emotional, and behavioral characteristics. As such, the engagement concept following this research stream is seen as multidimensional (Patterson, Yu, and de Ruyter 2006; Vivek, Beatty, and Morgan 2012) and context dependent (Hollebeek 2011; Mollen and Wilson 2010) and is observed at different levels of intensity and complexity (Brodie et al. 2011).

The second line of research focused on conceptualizing engagement as an intrinsic motivation to actively engage individuals with the content on social media (Baldus, Voorhees, and Calantone 2015; Dolan et al. 2016; Muntinga, Moorman, and Smit 2011). Specifically, Baldus, Voorhees, and Calantone (2015) focused on uncovering the motivations that drive individuals to interact (and sustain interaction) with an online brand community, while Dolan et al. (2016) investigated the role of social media content in facilitating (or harming) behavioral engagement.

The third stream of research conceptualizes engagement with content on social media as a behavior (Dolan et al. 2016; Eigenraam et al. 2018; Jiang et al. 2022; Muntinga, Moorman, and Smit 2011; van Doorn et al. 2010). Specifically, van Doorn et al. (2010) suggest that "customer engagement behaviors go beyond transactions, and may be specifically defined as a customer's behavioral manifestations that have a brand or firm focus, beyond purchase, resulting from motivational drivers" (p. 254). The authors further explain that these behaviors can be positive (e.g., writing a positive comment or "reacting" to a post using a positive emoticon, such as a heart) or negative (e.g., writing a negative comment or review on the post or "reacting" to a post using a negative emoticon, such as an angry face). This conceptualization of engagement as a behavior is corroborated by research from Aldous et al.'s (2019) quantification of social media engagement, wherein engagement with social media platforms can be summarized into four levels: views, likes, comments/shares, and cross-posting to external sites. In their work, Aldous et al. (2019) curated a database of 4,000 total news articles extracted from social media and extracted language, topic, textual, and sentiment features from these articles. They found that these features were markedly useful for predicting engagement; therefore, we utilize them in our study as potential predictors of engagement with disinformation ads.

According to previous studies, these behavioral manifestations can usually reflect individuals' internal cognitive and affective evaluations of the social media content (Yang and Zhao 2021), that is, reacting to an ad using the "like" button signals a positive affective evaluation. Therefore, in this study, we adopt the definition of engagement as a behavior, as "liking" or "clicking" represent tangible outputs of performance on social media and have been found to predict engagement with social media content (Aldous et al. 2019).

However, it is important to note that previous research has proposed different levels of engagement on social media. For instance, Ji et al. (2017) propose that there are two levels of behavioral engagement: shallow and profound. According to the authors, shallow engagement (i.e., "likes" and "shares" of social media content) requires little effort because the user needs only to click on a button. Conversely, profound engagement (e.g., commenting on a social media post) requires "greater mental effort than a mere one-click action" (Yoon et al. 2018, p. 25) because the user must elaborate on what to write. Contrary to this dichotomous view of engagement level, Gavilanes, Flatten, and Brettel (2018) proposed an engagement continuum based on four levels, which is similar to Aldous et al.'s (2019) classification system. The first level is considered a weak form of engagement, where individuals click on content; however, this is an important level, as it shows perception, attention, and likely interest in the content. The second level is classified as moderate and requires a little more effort, such as clicking on the "Like" button, or any other button depicted by emoticons. The third level is classified as moderate to strong and assumes greater effort from individuals, as well as more cognitive processing; at this level, individuals write comments and elaborate on their opinions, which could be positive, neutral, or negative. The fourth and final level in the continuum is classified as strong engagement and assumes that individuals might publish content that indicates their willingness for others to see and participate.

In sum, while these different classifications of engagement level might indicate a continuum of low to high effort on the part of the individuals, they indicate an action or reaction toward the social media content. Therefore, investigating behavioral engagement (e.g., impressions, clicks, likes, shares, comments) with these ads is the most suitable conceptualization for our study. Furthermore, because the metadata made available for the data set captures only two of the aforementioned actions (ad impressions and clicks), we opted to use ad clicks as our behavioral metric for ad engagement. Ad clicks is a good measure because it indicates actual behavior beyond mere exposure (i.e., impressions) and how many users engaged with the ad—in other words, took action by clicking on the ad after exposure (Zhang and Mao 2016).

While previous works have focused on the generation and measurement of disinformation content, the goal of this research is to assess in depth the effectiveness (i.e., engagement) of such content. Thus, this article expands on these prior works by focusing on Facebook disinformation ads to find correlations between engagement (ad clicks) and 40 features investigated in previous research and found to have potential predictive power. In addition, we compared six machine learning models for feature selection to further analyze which ad features were most important for engagement. In the next section, we detail our choice of features based on previous research.

Feature Selection Rationale

Investment

These features are related to ad spend and lifetime. Baltas (2003) has shown that the more money is spent on online advertising, the higher the consumer response (i.e., clicks). Ad lifetime has been also shown to influence engagement with social media posts. Albeit De Vries, Gensler and Leeflang (2012) and Munaro et al. (2021) focused on certain times of the day (i.e., weekday versus weekends) and relationship with engagement, we focused on how long (in hours) an ad was "live" on Facebook.

Caption Length

Length refers to the verbosity of posts and has been studied in both videos and written social media posts (e.g., Facebook and Twitter; Banerjee and Chua 2019; De Vries, Gensler, and Leeflang 2012; Munaro et al. 2021; Sabate et al. 2014). Previous research in advertising suggests that message length influences engagement metrics such as click-through rate (Baltas 2003) likes, shares, and comments (Banerjee and Chua 2019; Sabate et al. 2014). However, findings from this research are mixed. For example, Baltas (2003) found that shorter messages stimulate consumer action (i.e., clicks). Corroborating this idea, Banerjee and Chua (2019) showed that lengthy posts (> 151 words) were negatively related to likes and shares, while Sabate et al. (2014)[1] found the opposite result, suggesting that longer posts can increase the number of likes.

Sentiment and Subjectivity

Sentiment analysis refers to the extraction of emotional tone (i.e., positive, negative) from texts (Liu 2012; Rambocas and Pacheco 2018), while subjectivity quantifies the amount of subjective, personal opinions and factual information. Sentiment analysis has been shown to be particularly useful in marketing and advertising as it allows for an unprecedented opportunity to collect market intelligence (i.e., understand how consumers feel about a company or brand) using raw, user-generated commentary (Erevelles, Fukawa, and Swayne 2016). While most sentiment analyses have focused on what consumers have to say about brands (Rambocas and Pacheco 2018), a few studies (Munaro et al. 2021) have focused on applying this feature to understand the sentiment of a producer's message (i.e., a brand, influencer) and their potential effects on engagement. In their study, Munaro et al. (2021) found that negative and low-arousal content was more effective in producing views, likes, and comments on YouTube. In the same vein, the analysis of subjectivity in texts has been scarce. Specifically, Munaro et al. (2021) found that videos containing influencers showing their opinions, beliefs, and feelings were more successful in generating behavioral engagement than objective, factual information.

Psycholinguistic Features

Language and word usage are powerful tools to understand thoughts, feelings, personalities, and the way individuals connect and communicate (LIWC 2022). Previous research has investigated how word usage might influence engagement with social media content. For example, Munaro et al. (2021) investigated the effects of analytical thinking, while Yoon et al. (2018) studied emotional tone on behavioral engagement.

In sum, this study employs a comprehensive set of features that have been studied separately in previous research. Given that our work seeks to unearth the ad features most predictive of engagement with disinformation ads, our article and analyses might provide in-depth insights on behavioral engagement as a key metric of disinformation impact. Therefore, considering the scarce body of research on disinformation ads and engagement, we aimed to investigate the following research questions:

RQ1: Is there a relationship between a disinformation ad's features and engagement?

RQ2: What feature set makes a disinformation ad successful?

RQ3: Given a set of the most discriminant features, how accurately can one predict engagement?

In addition to investigating what features trigger the most engagement with disinformation ads, we leveraged latent Dirichlet allocation (LDA) to detect, in an unsupervised fashion, the major topics/groups weaponized in the ads. This analysis is important because it helps reveal the types of topics/groups most used across disinformation ads as well as their scheduling timeline (i.e., times of the year that ads with a certain topic appeared). Therefore, we asked an additional question:

RQ4: Which semantic topics best characterize the Facebook IRA ad data set?

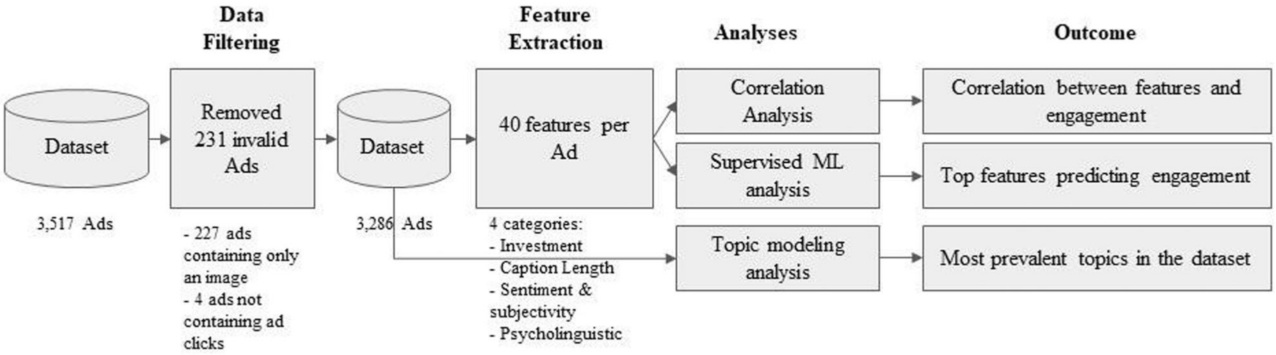
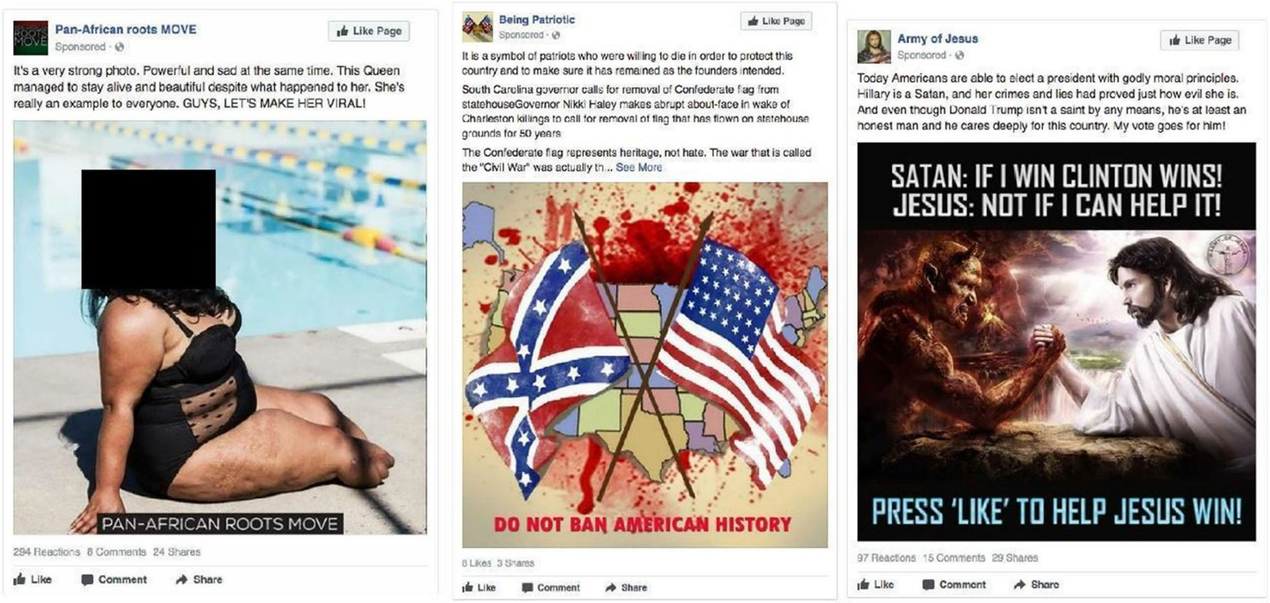


Figure 1. Overview of the data preparation, feature extraction, and analyses.



It's a very strong photo. Powerful and sad at the same time. This Queen managed to stay alive and beautiful despite what happened to her. She's really an example to everyone. GUYS, LET'S MAKE HER VIRAL!

[redacted]

It is a symbol of patriots who were willing to die in order to protect this country and to make sure it has remained as the founders intended.

South Carolina governor calls for removal of Confederate flag from statehouseGovernor Nikki Haley makes abrupt about-face in wake of Charleston killings to call for removal of flag that has flown on statehouse grounds for 50 years

The Confederate flag represents heritage, not hate. The war that is called the "Civil War" was actually th... See More

Today Americans are able to elect a president with godly moral principles. Hillary is a Satan, and her crimes and lies had proved just how evil she is. And even though Donald Trump isn't a saint by any means, he's at least an honest man and he cares deeply for this country. My vote goes for him!

Figure 2. Examples of emotional or visceral images included in the Russia's Internet Research Agency (IRA) ads.

Data Set and Feature Extraction

Data Set Description and Filtering

We leveraged a data set of 3,517 Facebook ads created by the IRA and made publicly available to the HPSCI (n.d.) by Facebook after internal audits. Estimated to have been exposed to over 126 million Americans between June 2015 and August 2017, these ads were a small representative sample of more than 80,000 pieces of content identified by the HPSCI. Of the 3,517 ads, 3,290 contained text; the remaining 227 ads (containing only an image) were purged from the data set, as we were interested in performing sentiment analysis and topic modeling based on the ads' captions. Next, we discarded four ads that did not contain a numerical value for the number of ad clicks (i.e., our metric of engagement). Therefore, our final data set contained 3,286 Facebook ads created by the IRA. Most of these ads (52.8%) were posted in 2016 (the U.S. election year), followed by 29.2% in 2017, and the remaining 18.0% in 2015. Please refer to Figure 1 for a complete overview of the data preparation, feature extraction, and analyses.

The data set consisted of one PDF file for each Facebook ad. A typical PDF datum was composed of two pages, where the first page contained ad metadata and the second page contained a screenshot of the ad as seen by Facebook users (see Figure 2 for examples). We automatically extracted engagement features from each ad: ad impressions (views)[2] and ad clicks. We opted to disregard ad impressions as a measure of engagement in our analyses because it was highly correlated with ad clicks ($r = 0.94$, $p < .01$). Thus, our main metric of engagement was ad clicks.

Table 1. Summary of all features for each group.

		Standard Engagement (Ad Clicks < 2, 188, *N* = 2,854)			High Engagement (Ad Clicks ≥ 2, 188, *N* = 432)		
Feature Category	Feature	Min	Mean	Max	Min	Mean	Max
Class label	Ad clicks	0	297.81	2,182	2,214	6,248	73,063
Investment	Ad lifetime (hours)	0	125.6	6,722.42	16.63	59	1,200.34
	Ad spend (RUB)	0	917.22	27,500	100	7311	331,675.75
Caption length	Character count	7	270.2	2,716	6	163	1,641
	Word count	0	44.72	437	1	26	274
Sentiment and subjectivity	NLTK VADER compound score	−1	0.08	1	−0.99	0.17	0.97
	NLTK negative sentiment only	−1	−0.24	0	−0.99	−0.12	0
	NLTK positive sentiment only	0	0.32	1	0	0.3	0.97
	NLTK neutral sentiment only (binary)	0	0.15	1	0	0.28	1
	TextBlob sentiment polarity	−1	0.1	1	−0.8	0.11	1
	TextBlob negative sentiment only	−1	−0.05	0	−0.8	−0.04	0
	TextBlob positive sentiment only	0	0.15	1	0	0.15	1
	TextBlob neutral sentiment only (binary)	0	0.25	1	0	0.38	1
	Flair sentiment	0	0.59	1	0	0.7	1
	Flair negative sentiment only (binary)	0	0.41	1	0	0.3	1
	Flair positive sentiment only (binary)	0	0.59	1	0	0.7	1
	TextBlob subjectivity	0	0.39	1	0	0.35	1
	TextBlob subjective scores only	0	0.23	1	0	0.23	1
	TextBlob objective Scores only	0	0.16	0.5	0	0.12	0.5
Psycholinguistic (LIWC)	Analytical thinking	0	70.27	99	1	60.1	99
	Authentic	0	27.24	99	1	23.8	99
	Clout	0	75.91	99	1	74.69	99
	Emotional tone	0	47.5	99	1	45.85	99
	Affective processes	0	7.8	100	0	7.81	100
	All punctuation	0	22.1	180	0	23.19	150
	Biological processes	0	2.06	33.33	0	1.61	25
	Cognitive processes	0	8.4	100	0	10.09	71.43
	Death	0	0.41	50	0	0.47	50
	Drives	0	14.31	100	0	13.58	100
	Future focus	0	0.67	50	0	0.82	20
	Past focus	0	1.96	50	0	2.32	50
	Present focus	0	10.59	100	0	11.18	57.14
	Home	0	0.34	28.57	0	0.22	25
	Leisure	0	1.51	33.33	0	1.21	33.33
	Money	0	1.04	20	0	0.52	20
	Perceptual processes	0	4.59	100	0	4.85	50
	Relativity	0	11.35	66.67	0	9.73	100
	Religion	0	0.98	66.67	0	0.64	33.33
	Social processes	0	13.13	80	0	14.84	66.67
	Work	0	2.72	50	0	2.47	100

Note. For a detailed list and explanation of Linguistic Inquiry and Word Count (LIWC) psycholinguistic features, see Pennebaker et al. (2015). Because ad impressions were highly correlated with ad clicks, we opted to use ad clicks as our metric of engagement and removed ad impressions from further analyses. Nonetheless, ad impressions had the following results: standard engagement: min = 0; mean = 3,715.92; max = 165,121.00; high engagement: min = 8,429.00; mean = 65,223; max = 1,334,544.00.

Feature Extraction

For each of the 3,286 ads, we extracted a total of 40 features (see Table 1) typically used in the literature to characterize social media ads and posts (e.g., Aldous et al. 2019; Munaro et al. 2021; Sabate et al. 2014; Vosoughi et al. 2018; Yoon et al. 2018). These features can be summarized into four main categories extracted from the metadata: (1) investment features, related to money spent on the ads and lifetime; (2) caption length features, related to the size of the caption; (3) sentiment and subjectivity features, describing both valence (positive versus negative) and salience (low to high arousal) of sentiment in the ad's caption; and (4) psycholinguistic features, related to emotions, mood, and cognition present in the ad's caption based on word counts (e.g., the words *crying*, *grief*, and *sad* are counted as expressing sadness). The features from categories 2, 3, and 4 are all related to the message contained in the ads.

Investment Features

We automatically extracted investment features from each ad: ad spend (money, in Russian rubles [RUB], spent on the ad), and ad lifetime (the ad's creation and end dates, in hours).

Caption Length

We summarized the length of the ad's caption using character count and word count.

Sentiment and Subjectivity Analyses

We leveraged three sentiment analysis packages: (a) VADER (Hutto and Gilbert 2014), (b) TextBlob (Loria et al. 2014), and (c) Flair (Akbik et al. 2018). These three packages yielded 14 sentiment and subjectivity features, which we opted to use in our analyses (see Table 1).

Psycholinguistic Features

To extract psycholinguistic features, we leveraged Linguistic Inquiry and Word Count (LIWC2015) (Pennebaker et al. 2015), a text analysis tool that reflects a text's emotions, thinking styles, social concerns, and grammar (e.g., parts of speech) by counting words in psychologically meaningful categories represented by dictionaries. Each dictionary contains a collection of words that defines a particular category (e.g., the category "religion" contains words such as *altar* and *church*). A total of 21 psycholinguistic features were extracted:

- Four summary variables: analytical thinking (formal, logical thinking versus informal, personal thinking), clout (expertise and confidence versus tentative or anxious), authenticity (honest, personal versus guarded, distanced), and emotional tone (positive versus anxiety, sadness, or hostility). Each of these is measured on a 100-point scale.
- Seventeen other LIWC categories, most of which are related to psychological processes. Each of these features was measured as a percentage of words (e.g., "affective process" of 10 means that 10% of all words of the ad's caption were related to emotions, such as *happy* and *cried*). See Pennebaker et al. (2015) for examples.

Topic Modeling

Topic modeling was performed on the ads' captions using latent Dirichlet allocation (LDA), an unsupervised probabilistic generative model. Simple preprocessing was done to make the text more amenable for analyses, including the removal of punctuation and stop words, and lowercasing. To transform the texts into a format that serves as input for the LDA model, we converted the texts into a simple vector representation using bag of words (BoW). Then, we converted the list of ad captions into lists of vectors, all with length equal to the vocabulary. Words were then lemmatized, keeping only nouns, adjectives, verbs, and adverbs.

The groups of people that were targeted for each advertisement were provided amongst the several metadata in our original data set. Using this information and the keywords associated with each topic, we then inspected the cleaned LDA topic results and proposed topic labels; for example, keywords such as *conservatism*, *republican*, *Fox News*, *Trump*, and *conservative* were assigned to the "conservative or Republican" category.

Evaluation Metrics

Correlation Analysis

We used the alternating conditional expectations (ACE) algorithm to find the fixed point of maximal correlation (MC) for each extracted feature. In other words, we transform the dependent and independent variables to maximize Pearson's correlation coefficient between the transformed dependent and transformed independent variables. Deebani and Kachouie (2018) tested several correlation analysis methods on several simulations of different relationship types, with and without noise, and found that MC equaled or outperformed Pearson's and Spearman's correlation. The authors thus describe MC as efficient and robust to noise and allow for nonlinear correlations to be detected. It is important to note that MC ranges from [0, 1], and does not measure the polarity of the correlation.

Predicting Engagement

To predict ad engagement, we used a classification approach, where one or more learning models or classifiers are trained to predict class labels (categorical variables—our dependent variable) represented by discrete values (Han, Pei, and Tong 2022). As such, we used six supervised classifiers with the target variable having two possible values: standard versus high

Table 2. Summary of the performance of all models used for feature selection, ranked based on mean *F* score.

		F Score (%)				
Rank	Classifier	Mean (%)	σ	Min	Max	Optimal Number of Features
1	Linear support vector machine	93.6	0.0	93.6	93.7	12
2	Logistic regression	93.6	0.1	93.4	93.8	6
3	Gradient boosting	93.4	0.1	93.1	93.8	3
4	Random forest	93.0	0.1	93.0	93.7	1
5	Adaboost	93.0	0.2	92.7	93.6	4
6	Bernoulli Naive Bayes	90.0	2.8	84.1	93.0	1

Table 3. Top five ranked features for each model tested for feature selection.

Rank	Linear SVM	Logistic Regression	Gradient Boosting	Random Forest	Adaboost	Bernoulli Naive Bayes
1	Ad spend Religion Drives Biological processes NLTK negative only Authentic Analytical thinking Emotional tone NLTK compound NLTK positive only Character count Word count	Ad spend Word count NLTK negative only Analytical thinking Character count NLTK positive only	Ad spend Character count Ad lifetime	Ad spend	Ad spend Character count NLTK compound Ad lifetime	Religion
2	Past focus	Religion	Word count	Character count	All punctuation	Ad spend
3	Affective processing	Drives	Analytical thinking	Word count	Religion	Home
4	Ad lifetime	Authentic	NLTK compound	Drives	Cognitive processes	Ad lifetime
5	TextBlob objective only	Biological processes	All punctuation	Ad lifetime	Past focus	Death

Note. NLTK = Natural Language Toolkit.

engagement ads. The models used were Adaboost, Bernoulli Naive Bayes (NB), Gradient Boosting, Linear SVM, Logistic Regression, and Random Forest (see Tables 2 and 3), all of them with default parameters. To discard irrelevant features from our set of collected features (see Table 1) and retain only those able to best predict ad engagement, we used Recursive Feature Elimination (RFE) combined with the aforementioned models. We then compared the optimal subset resulting from each model and checked for commonalities among the selected features.

Next, we standardized all features by removing the mean and scaling to unit variance. After that, to evaluate the learning models, we randomly split the data set into a training set with $\sim 2/3$ of the data and a testing set with $\sim 1/3$ of the data. The former was used for training each of the models while the latter was used for testing their effectiveness in predicting ad engagement. The models were evaluated using stratified five-fold cross-validation due to their relatively low bias and variance (Han et al. 2022). The evaluation metric used was F-score, which is well suited to handle imbalanced data sets as in our case (Han et al. 2022).

Data Analysis and Results

Correlation Analysis (RQ1: Relationship between Features and Engagement)

We opted to separate the data set into a standard group and an outlier group to better understand how low versus high engagement vary as a function of ad features. Using the 1.5 × IQR rule (i.e., values above $Q3 + 1.5 \times IQR$), we identified 432 upper outliers based on ad clicks. Therefore, ads with < 2,188 clicks ($n = 2{,}854$; 86.9%) were assigned to the Standard Engagement group (subscript *stand*) and those with ≥ 2,188 clicks ($n = 432$; 13.1%) were assigned to the High Engagement group (subscript *high*).

Before we could perform statistical analyses on the extracted features, we used Shapiro-Wilk to test for normality and found that the continuous investment features (impressions, ad spend, and lifetime) and ad clicks were not normally distributed ($p < .001$ for all variables). Prior to calculating Pearson's and Spearman's Rank Correlation Coefficients, we normalized the continuous features using the Yao-Johnson power transformation a modified Box-Cox transformation that allows values ≤ 0) because these features exhibited a heavy positive skew (Fisher-Pearson coefficient > 1).

We used Pearson's (r) and Spearman's Rank Correlation (ρ) tests to find the linear and monotonic correlations, respectively, between ad clicks and all other extracted features. We found moderate to strong positive correlations between ad clicks and investment features (Table 4). For example, for both Standard and High Engagement groups, ad clicks strongly correlated with ad spend ($\rho_{stand} = 0.79$, $r_{high} = 0.65$, $p < .001$). Strong or moderate correlations did not hold true for the remaining feature categories. For example, we found nearly no statistically significant correlations with sentiment and subjectivity features for both Standard and High Engagement groups. There were several trivial correlations between ad clicks and psycholinguistic features for the Standard Engagement group, and no statistically significant results for the majority of the LIWC features for the High Engagement group.

These overall low correlation coefficients can be explained based on the number of trivial correlations (ρ, $r < 0.1$), indicating that these variables do not exhibit a monotonic nor a linear relationship, and therefore ρ and r cannot fully describe their pairwise

Table 4. Correlation analyses for ad clicks (dependent variable) versus features.

		Standard Engagement (Ad clicks $<$ 2, 188, $N = 2{,}854$)			High Engagement (Ad clicks $\geq$ 2,188, $N = 432$)		
Category	Feature	r	ρ	MC	r	ρ	MC
Investment	Ad lifetime	−0.05***	0.21***	0.34***	−0.03***	−0.01***	0.25***
	Ad spend	0.27***	0.79***	0.70***	0.65***	0.23***	0.70***
Caption	Character count	−0.01***	0.08***	0.22***	n.s.	n.s.	0.21***
	Word count	−0.02***	0.08***	0.22***	n.s.	n.s.	0.20***
Sentiment and subjectivity	NLTK VADER compound score	n.s.	n.s.	0.15***	n.s.	n.s.	0.17***
	NLTK negative sentiment only	n.s.	n.s.	0.07***	n.s.	n.s.	0.10*
	NLTK positive sentiment only	n.s.	n.s.	0.04*	n.s.	n.s.	0.17***
	TextBlob sentiment polarity	n.s.	n.s.	0.08***	n.s.	n.s.	0.15**
	TextBlob negative sentiment only	0.01*	n.s.	0.07***	n.s.	n.s.	n.s.
	TextBlob positive sentiment only	n.s.	n.s.	0.09***	n.s.	n.s.	0.13**
	TextBlob subjectivity	0.01*	n.s.	0.10***	n.s.	n.s.	0.13**
	TextBlob subjective scores only	n.s.	n.s.	0.07***	n.s.	n.s.	0.12*
	TextBlob objective scores only	n.s.	n.s.	0.05*	n.s.	n.s.	0.11*
LIWC summary Variables	Authentic	n.s.	n.s.	0.08*** 0.09***	n.s.	n.s.	0.15**
	Analytical thinking	−0.09***	−0.08***	0.08***	n.s.	n.s.	0.12*
	Clout	−0.06***	−0.06**	0.08***	n.s.	n.s.	0.16**
	Emotional tone	n.s.	−0.04*		n.s.	n.s.	0.14**
LIWC categories	Affective processes	n.s.	n.s.	0.07***	n.s.	n.s.	0.14**
	Social processes	−0.03*	−0.04*	0.06***	n.s.	n.s.	0.16**
	Cognitive processes	0.07***	0.06**	0.11***	n.s.	n.s.	0.11*
	Perceptual processes	n.s.	n.s.	0.17***	n.s.	n.s.	n.s.
	Biological processes	n.s.	n.s.	0.06**	n.s.	n.s.	n.s.
	Drives	−0.10***	−0.13***	0.14***	0.15***	0.15**	0.21***
	Future focus	0.01***	0.08***	0.11***	n.s.	n.s.	n.s.
	Past focus	0.08***	0.13***	0.14***	n.s.	n.s.	0.14**
	Present focus	n.s.	n.s.	0.11***	n.s.	n.s.	0.15**
	Relativity	n.s.	n.s.	0.10***	n.s.	n.s.	0.12*
	Work	0.01***	0.10***	0.18***	n.s.	n.s.	0.11*
	Death	0.03***	0.07***	0.11***	−0.04*	n.s.	n.s.
	Home	n.s.	n.s.	0.06**	n.s.	n.s.	n.s.
	Leisure	−0.06*	−0.05**	0.07***	n.s.	n.s.	n.s.
	Money	−0.09***	−0.07***	0.10***	−0.08*	−0.10*	0.13**
	Religion	n.s.	n.s.	0.07***	n.s.	n.s.	0.11*
	All punctuation	n.s.	n.s.	0.10***	n.s.	n.s.	n.s.

Note. NLTK = Natural Language Toolkit; LIWC = Linguistic Inquiry and Word Count. Range for maximal correlation (MC) is [0, 1] whereas Pearson's (r) and Spearman's (ρ) coefficients range is [−1, 1]. Ad impressions results: standard engagement ($r = 0.49$***; $\rho = 0.94$***; MC = 0.93***) and high engagement ($r = 0.89$***; $\rho = 0.76$***; MC = 0.89***).
*$p < .05$; **$p < .01$; ***$p < .001$; n.s. = not significant.

correlations. This further cements our decision to rely on MC as a measure of correlation, as MC can capture both linear and nonlinear relationships, resulting in greater predictive value of the extracted features.

Investment Features

Maximal correlation greatly improved Pearson's correlation between the investment features and ad clicks. For example, ad lifetime exhibited a trivial linear correlation with clicks for both Standard and High Engagement groups ($r_{stand} = -0.05$, $r_{high} = -0.03$, $p < .001$); these values increased to a weak to moderate relationship following the ACE transformation: $MC_{stand} = 0.34$, $MC_{high} = 0.25$, $p < .001$.

Caption Length

Following the MC transformations, the character and word counts of ads increased to a small positive correlation ($MC_{stand} = 0.22$, $p < .001$; and $MC_{high} = 0.21$ and $MC_{high} = 0.20$, $p < .001$, respectively).

Sentiment and Subjectivity

As previously stated, MC does not report the direction of the relationship between the variables; however, NLTK Compound Score and NLTK Negative Sentiment Only Score exhibited the largest mean differences between the Standard and High Engagement groups: $\mu_{stand} = 0.08$ versus $\mu_{high} = 0.17$, and $\mu_{stand} = -0.24$ versus $\mu_{high} = -0.12$, respectively. Therefore, not only did High Engagement ads demonstrate higher (in terms of polarity and magnitude) MCs with sentiment and subjectivity than the Standard Engagement ads, but High Engagement ads were also more positive in sentiment, on average than Standard Engagement ads.

Sentiment and subjectivity variables were further analyzed using the Chi-Squared test (see Table 5). All sentiment features were found to be dependent on engagement as measured by the Standard and High Engagement ad click groups ($p < .001$), suggesting that sentiment features are associated with ad engagement. However, subjectivity was not statistically

Table 5. Chi-squared tests analyzing ad clicks (for standard and high engagement) versus sentiment and subjectivity features.

Feature Category	Feature	χ^2 Stat	*df*	*N*	*p* Value
Sentiment and subjectivity	NLTK VADER compound score	54.122	2	3,286	$< .001$
	TextBlob sentiment polarity	31.749	2	3,286	$< .001$
	Flair sentiment	17.965	1	3,286	$< .001$
	TextBlob subjectivity	1.574	1	3,286	0.210

Note. NLTK = Natural Language Toolkit.

significant and therefore TextBlob Subjectivity was independent of engagement (χ^2 (1, $N = 3{,}286$) $=$ 1.574, $p = .21$). Moreover, MC was statistically significant for both groups, further emphasizing the robustness of this correlation analysis method.

Psycholinguistic Features

For the four summary LIWC variables, the High Engagement group showed weak MCs ([0.12, 0.16], $p < .05$) whereas Standard Engagement showed only trivial correlations. The mean values for the summary variables were nearly unchanged across the engagement groups, with *Analytic Thinking* as the exception: the average score drops from 70.27 to 60.10 for the Standard versus High Engagement groups suggesting that High Engagement ads were more informal and personal.

The remaining LIWC categories experienced nearly no variation in mean values across the Standard and High Engagement groups. Six features were found to not be statistically significant for High Engagement, whereas all LIWC features were statistically significant ($p < .01$) for Standard Engagement. This could be due to the discrepancy between the average caption length for Standard and High Engagement: 45 words versus 26 words; it is possible that more significant results could not be found for the High Engagement group due to small sample sizes.

The feature *Drives* stands out, as it was the only LIWC feature to find nontrivial ($\geq$ 0.1) Pearson, Spearman, and MCs for both Standard and High Engagement. Both engagement groups experienced nearly the same average *Drives* value ($\mu \approx 0.14$) and this average value was the third largest mean LIWC category value for the High Engagement group. We thus can conclude that *Drives* appears to be associated with ad engagement, especially for High Engagement ads.

Feature Selection (RQ2 and RQ3: Features Predicting Ad Engagement)

Our correlation and statistical analyses used to address RQ1 relied on the individual relevance of each feature in characterizing engagement. However, individual features sometimes fail in predicting the target variable accurately. Machine learning models can combine multiple features to predict the target, sometimes revealing promising features that do not have relevant pairwise correlation results. Our results are summarized in Table 2.

Investment Features

The investment features mirrored the results from the MC analysis (RQ1)—that is, both ad spend and lifetime were selected as important features for predicting engagement, particularly in distinguishing between Standard and High Engagement. Ad spend was ranked the most important in five models, and ad lifetime was ranked top five by five models.

Caption Length

The size of the ad's caption appeared in the top two most important features for all models tested, corroborating our MC results. Based on the average values for caption length, we can infer that shorter ad texts were more engaging (e.g., character count: $\mu_{stand} =$ 270.20, $\mu_{high} = 163$).

Sentiment and Subjectivity

At least one sentiment or subjectivity feature was ranked top five by five models (see Table 3). Notably, NLTK's VADER Compound Score was ranked first by three models: Linear SVM, Adaboost, and Linear SVM, whereas NLTK Positive Only scores and NLTK Negative Only scores were selected by both Linear SVM and logistic regression. Based on Table 1, we see that NLTK Negative Only scores were more negative for Standard Engagement than for High Engagement ($\mu_{stand} = -0.24$, $\mu_{high} = -0.12$), and NLTK Compound Scores were more positive for High Engagement ads ($\mu_{stand} = 0.08$, $\mu_{high} = 0.17$), which is in accordance with our earlier observations that positive sentiment disinformation ads in this data set were more engaging.

Psycholinguistic Features

Two summary variables appeared in three out of the six models tested for feature selection: Analytic Thinking (Linear SVM, Logistic Regression, and

Gradient Boosting) and Authentic (Linear SVM and Logistic Regression). High Engagement ads were, therefore, more informal and personal than Standard Engagement ads.

Two other LIWC features were selected among the top five by the feature selection models: Drives and Religion, chosen by three and four of the six models, respectively. Whereas *Drives* experienced nearly the same min, mean, and max values for both High and Standard Engagement, Religion differed in max values for Standard ($max = 0.66$) and High ($max = 0.33$); therefore, Religion showed greater variety and range of values for the Standard group, yet all *Religion* scores for both Standard and High Engagement group were relatively low ($\mu_{stand} = 0.98$, $\mu_{high} = 0.64$), indicating low use of religious language across all ads. Conversely, both *Drives* ($MC_{stand} = 0.14$, $MC_{high} = 0.21$, $p < .001$) and *Religion* ($MC_{stand} = 0.07$, $p < .001$; $MC_{high} = 0.11$, $p < .05$) showed higher MC values for the High Engagement group as compared to Standard Engagement. LIWC features dominated four of the six feature selection models: 8/16 for Linear SVM, 5/10 for Logistic Regression, 4/8 for Adaboost, and 3/5 for Bernoulli NB.

Topic Modeling

We validated LDA's topic modeling performance using topic coherence, as described in Röder, Both, and Hinneburg (2015). Using the C_v coherence, i.e., the coherence computed as the average similarity between the top word context vectors and their centroid, we find the set of parameters with maximum coherence value of 0.58 for the entire data set: $\beta = 0.01$ and $\alpha = 0.91$, yielding a total of eight topics. The coherence measure C_v is a continuous value in the range [0,1], where 1 indicates the highest degree of semantic similarity between high scoring words in the topic. In the LDA implementation, we have experimented with different values for α and β. In Figure 3, we present the largest topic coherence C_v score as a function of the number of topics. The selected number of topics (8) returned a coherence score of $C_v \approx 0.58$. We then reduced the number of repeated keywords across different topics and proposed the following eight overarching topic categories (see Table 6): (1) American patriotism, (2) justice/African-American, (3) perseverance/liberal/democrat, (4) female rights/education, (5) peace/guns, (6) police/military, (7) community integration/LGBT, and (8) capitalism/conservative/republican.

Our results, as analyzed and validated using machine learning algorithms, are in agreement with the qualitative analyses presented in prior works (Howard et al. 2018; DiResta et al. 2019). Figure 4 shows the occurrence of each summary topic derived by the LDA topic modeling from June 2015 to August 2017. We see that the majority of Topic 7 (community integration/LGBT) has the largest ad count preceding the election (May 2016). Interestingly, Topic 3 (perseverance/liberal/democrat) closely mirrors Topic 8 (capitalism/conservative/republican). We also observe several interesting occurrences when considering the median number of ad clicks for each summary topic during this same period. Topic 3 (perseverance/liberal/democrat) stands out in engagement before the election (February–July, 2016) as well as the significant impact during office takeover. Topic 5 (peace/guns) has some significant engagement in the months preceding the election and some impact during office takeover. Topic 1 (American patriotism) and Topic 7 (community integration/LGBT) follow each other throughout this timeline. Topic 2 (justice/African American) experiences relatively low

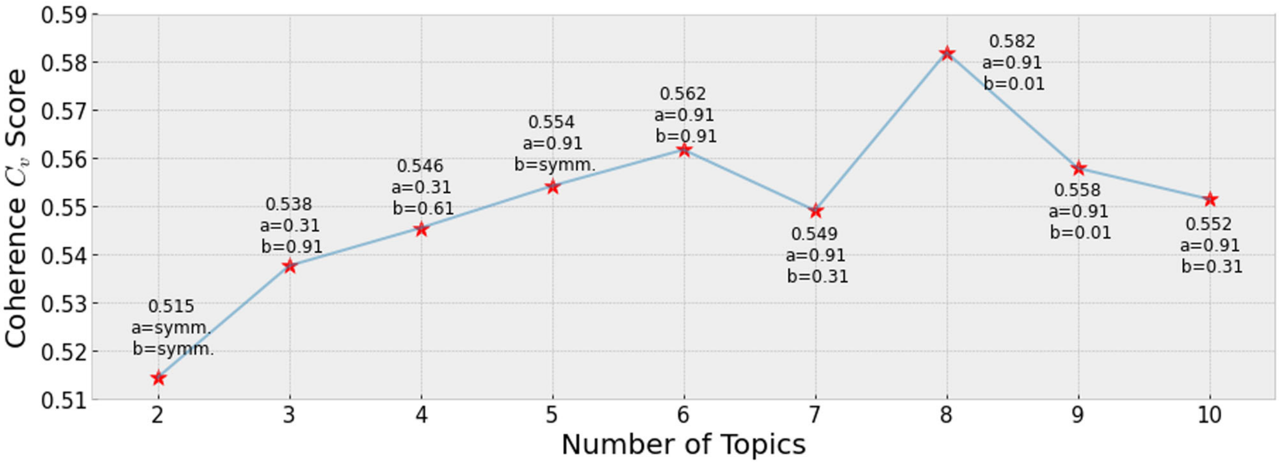


Figure 3. This figure illustrates the value of the C_v coherence as the number of topics increases; the graph displays only the best set of hyperparameters for each number of topics.

Table 6. Example keywords with the largest weight contribution to each topic.

	N		
Proposed Summary Topic	Standard Engagement	High Engagement	Example Keywords
(1) American patriotism	411	70	support, follow, vote, go, veteran, always, give
(2) Justice and African American rights	333	53	justice, year, group, racism
(3) Perseverance, liberal political movement, Democratic Party	376	51	people, right, stop, join, take, good, think, war
(4) Female rights and education	261	34	girl, woman, matter, student, young, people
(5) Peace and guns	308	54	let, think, war, need, right
(6) Police/military	360	81	cop, life, brutality, shoot, video
(7) Community integration, LGBT rights	286	63	stand, stay, nation, proud
(8) Capitalism, conservative political movement, Republican Party	490	55	free, self-defense, class, safe, world, white
Total	2,854	432	

Note. Topics were based on latent Dirichlet allocation (LDA) topic modeling.

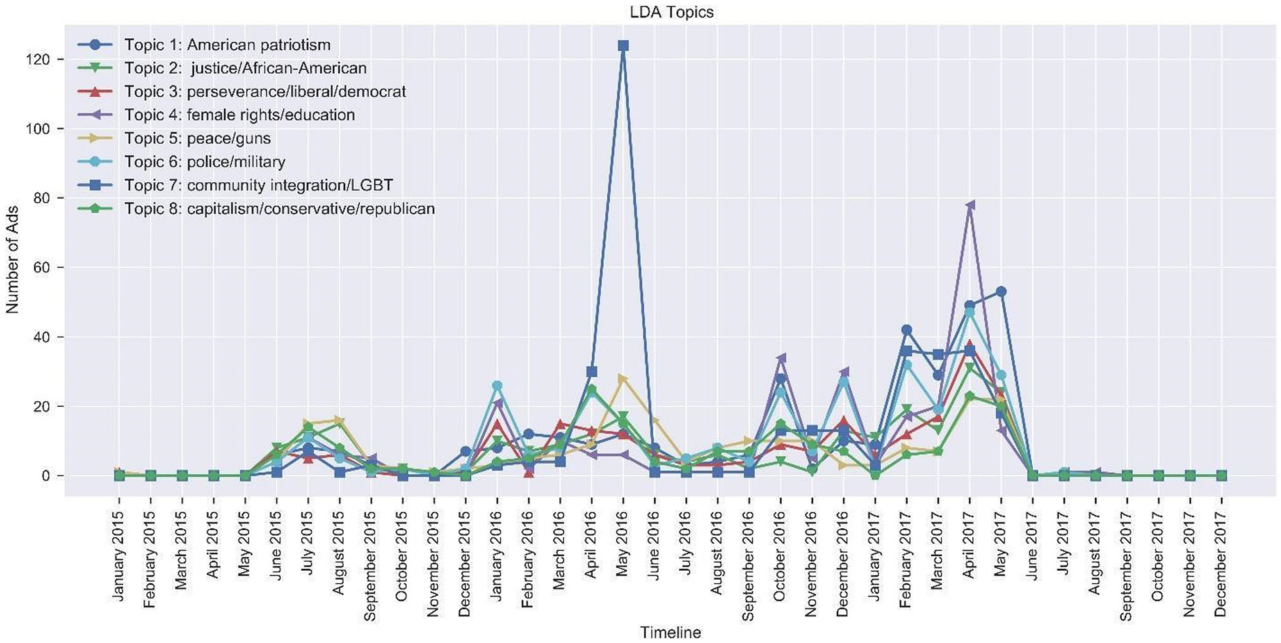


Figure 4. Total number of ads for each summary topic predicted by latent Dirichlet allocation (LDA).

median engagement numbers except for a spike during the office takeover period. In January 2017, there was a surprisingly big significant spike in engagement in both figures.

General Discussion

Summary of Findings

Using a unique data set released by the U.S. House of Representatives Select Committee on Intelligence (HPSCI n.d.), this study investigated the ad features more likely to lead to engagement with disinformation ads. Due to the large number of features examined in the study, we present a summary of findings based on four categories: investment features, caption length, sentiment and subjectivity, and psycholinguistic features.

Investment Features: Ad Spend and Lifetime

Our MC results (Table 4) show that ad spend and lifetime had a moderate to strong relationship with both Standard and High Engagement. This was supported during our feature selection analysis. The average ad spend for the High Engagement group was notably higher than that of the Standard Engagement group (7,311 versus 917 RUB). It is possible that paying more for an ad might be associated with a better targeting service from social media platforms, potentially causing the ad to reach more people who will be more interested in the ad (Baltas 2003).

Caption Length

Another important feature was the length of the ad's caption. Facebook truncates posts greater than 477 characters (Gessler 2016). High Engagement ads had, on average, nearly 110 fewer characters (and nearly 20 fewer words) than the Standard Engagement group; therefore, our results revealed that shorter ads are more engaging. Research on deception detection shows that deceivers embed influence cues in their content to blur people's decision-making (Kahneman 2011). In fact, accounts from Cold War disinformation point to the use of pictures, short texts, sexual appeal, sensationalism, and high-arousal emotion in disinformation stimuli (Rid 2020). However, we only considered the textual content of each ad and disregarded the presence of images. It is possible that ads with shorter texts used emotionally visceral images (see Figure 2) to communicate a message, likely increasing users' engagement with the advertisement.

Sentiment and Subjectivity

We found that sentiment features were highly important for predicting engagement, with High Engagement ads being more positive in sentiment than Standard Engagement ads. Corroborating this finding, there is indeed a wealth of cognitive and behavioral sciences research that points to the impact of affective states in decision-making (Forgas and George 2001; Isen and Baron 1991), where positive emotions have been shown to be more detrimental to rational decision-making than negative emotions. Positive affect states may cause an increase in trust and a decrease in social vigilance (Kahneman 2011; Kircanski et al. 2018); therefore, a user's good mood indicates a safe environment (Kahneman 2011), and can thus increase one's susceptibility to deception. Several works have also shown that con artists leverage high emotional arousal to persuade victims to comply with their requests (Kircanski et al. 2018; Loewenstein 1996) by focusing the victims' attention on reward cues (Langenderfer and Shimp 2001).

Psycholinguistic Features

We found that High Engagement ads were more personal and informal than Standard Engagement ads. Evans and Krueger (2009), and Cialdini's principles of persuasion (Cialdini 2006) offer plausible explanations for this: people who are perceived as familiar or similar are more likely to be trusted by others (a phenomenon termed the *in-group trust disposition*) and are more likely to have their requests obeyed. Therefore, ads whose authors masqueraded themselves as part of the targeted community may have achieved higher engagement levels.

In total, LIWC features dominated four of the six feature selection models. From this, we see that the content of the advertisement itself, along with the use of (or lack thereof) certain topics (e.g., religion) influences engagement. In this paper, we found that LIWC features such as *Authentic* (which measures how authentic a writer appears to be) suggest that the authors of the Facebook ads may have impacted users' engagement (e.g., an African American user posting about #BlackLivesMatter). The IRA has been shown to groom real users (Schifrin 2020) into writing their disinformation articles; as such, future works should analyze the accounts of users responsible for spreading disinformation in our data set.

Theoretical and Methodological Contributions to Advertising Research

Our findings contribute to the literature in at least three ways. First, to the best of our knowledge, this study is the first to formally conceptualize the notion of *disinformation ads* on social media. While disinformation campaigns have been around for many years (Bittman 1990; Martin 1982; Romerstein 2001), the term "disinformation" has been primarily used in the context of news (i.e., fake news) and propaganda (Steinfeld 2022; Tandoc, Lim, and Ling 2018). This differentiation is important because social media disinformation ads involve ads that are paid for or sponsored by an entity (i.e., any individual or organization can buy an ad or sponsored content on social media). Second, this study contributes to the advertising engagement literature by comprehensively investigating the relationship between 40 features and behavioral engagement (ad clicks). While some of these features have been studied in previous research, they have been studied separately (Alvarez, Choi, and Strover 2020; Munaro et al. 2021; Rambocas and Pacheco 2018) or using different stimuli (Aldous, An, and Jansen 2019). Therefore, our study design, techniques, and results can offer a baseline of how to investigate behavioral engagement in a more comprehensive manner. Furthermore, our classification of Standard and High Engagement ads provides nuanced insights into how these ads appeal to social media users and has the potential to guide future categorizations. Third, this study responds to the call for more research using AI tools to examine the relationship between textual content and individuals' engagement with social media ads (Li 2019).

The methodological contributions of this study are multifaceted. First, we combined a wide variety of features from the literature to understand engagement with disinformation ads. Besides features that are directly obtained from Ad metadata (e.g., spend and lifetime), we also extracted features with the ability to describe more nuanced characteristics of ads such as the ad's caption, sentiment, and psycholinguistic properties. This allowed us to investigate what makes for engagement with disinformation ads from different perspectives. Second, we designed a more efficient methodology for our correlation analysis by employing the Maximal Correlation (MC), which is more robust to noise and allows for detecting nonlinear correlations compared to other metrics more commonly used in the literature, such as Pearson and Spearman.

Third, we designed a robust machine learning methodology by considering multiple classifiers with different levels of comprehensiveness and complexity, combined with a feature selection stage for optimal effectiveness in predicting ad engagement. To illustrate, as exhibited in Table 3, the Linear SVM model reached the best average $F1$-score among all evaluated classifiers, but it took 12 features into account, which implies a more complex prediction model compared to the Logistic Regression classifier, which achieved nearly the same average F-score using half of the number of features. Random Forest, which has better interpretability than the other assessed models, achieved slightly inferior results using a single feature. This not only helps to foster more research using AI tools to analyze individuals' engagement with social media ads but also provides at least initial directions to future work on what models can be more promising in terms of accuracy and interpretability. Fourth, we conducted a feature importance analysis. In other words, we went beyond analyzing the effectiveness of machine learning models in predicting engagement based on a set of handcrafted features and investigated the contribution of each of those features to the prediction process. This allowed us to understand the power of each analyzed feature for predicting ad engagement, which oftentimes is not possible through the models themselves. For example, by glancing at Table 3, one can see that the feature Ad Spend was top ranked by five of the six models considered, meaning that this is a very relevant feature for predicting disinformation ad engagement with AI tools.

Finally, we leveraged LDA for topic modeling using unigrams, bigrams, and trigrams BoW. We utilized the topic coherence performance measure to select the number of topics to maximize this score. Final topics were then qualitatively summarized based on the list of words assigned by the model. This work helps reveal the types of topics/groups most used across disinformation ads. Because documents are also labeled in time, we can observe which topics are targeted in different timelines, providing a preliminary insight into disinformation schedules, and targeting strategies.

Limitations and Future Research

As with any research endeavor, this study also has limitations. First, though a valuable and unique data set, the HPSCI does not detail how this representative sample of 3,500 ads was selected and redacted. DiResta et al. (2019), who were given access to 61,500 Facebook ads, allude to the bias inherent in the data set: the social media platforms did not report a methodology, did not include anonymized user comments, and gave minimal metadata. This data set nonetheless has its merits; given Russia's long Active Measures campaign, we conjecture that Russia's highly skilled intelligence community is adept at composing galvanizing messages and targeting the audience most susceptible to engaging with their ads. This data set is therefore our peephole into the IRA's modus operandi. This also emphasizes the desperate need for rich and diverse disinformation data sets for future research.

Second, although we extracted a comprehensive number of features from the data set and utilized advanced analytical methods to understand what makes a disinformation ad engaging, our study is limited to the textual content of these ads and discarded any images associated with the ad, overlooking the presence of emotionally charged visual stimuli used in combination or as its own malicious ad product. Although analyzing images associated with the ads would have benefited this work, it is nonetheless important for the research community to be able to discern which textual and metadata features can be leveraged to automatically predict engagement. However, future research can leverage deep-learning architectures such as neural networks for image captioning to characterize the content of an image (Hossain et al., 2019) and pair them with the ad caption and data set's features. Similarly, if an ad contains only a video, future works can make use of video summarization and image captioning with attention-based mechanisms (Fajtl et al. 2019) to leverage all available information. In fact, this treatment of media files has its standalone merit and is well suited to be integrated within social media platforms. In addition, future research could combine experimental research that investigates consumer responses to such ads and their effects on outcomes such as trust and emotions.

Third, we leveraged LDA, a powerful tool for topic modeling, though it suffers from major drawbacks similar to many unsupervised models (e.g., data set size). We nonetheless corroborated prior works by Howard et al. (2018) and DiResta et al. (2019) showing that the IRA purposefully targeted communities to polarize political discourse in the United States. Based on this, the unsupervised LDA performed surprisingly well and may serve as additional features in future work. Moreover, we used BoW to convert the ad text into a vector representation to serve as an input for the LDA model. Given that BoW disregards certain properties of the text, such as grammar, semantic meaning, and word ordering, the use of other word vectorization techniques able to capture semantic meaning and other relevant properties, such as Word2Vec and GloVe, is therefore another research direction.

Finally, transforming ad engagement from a continuous variable into categorical variables led to an imbalanced data set, which may have caused our models to better predict standard engagement ads. Future work can address this imbalance by curating a larger data set by either (a) reproducing our coding methodology or (b) generating an artificial but balanced data set (e.g., by applying generative adversarial network [GAN]; Shamsolmoali et al. 2021). A downsampling strategy may also be adopted to reduce data imbalance, though a larger data set would be highly beneficial when applying machine learning.

Understanding what makes for high engagement in disinformation ads paves the way for countermeasures in several respects. Future research can evolve social media labels and potentially expose deceptive cues in posts from suspicious or biased accounts to better inform users. Future works using survey and experimental methods could test different types of social media labels (i.e., “This post is rated false” versus “Partly false information. Check other sources”) to assess their reception and effectiveness in labeling disinformation. This is particularly important when we consider that Russian Active Measures did not stop after 2016 and in fact intensified after the election (Select Committee on Intelligence 2019). For example, in 2018, the *Washington Post* reported that Russian trolls inflamed the U.S. debate over climate change (Timberg and Romm 2018). In June 2020, the Associated Press reported that U.S. officials confirmed that Russia was behind the spread of disinformation about the coronavirus pandemic (Tucker 2020). Disinformation campaigns have also been generated from their own nation-state figures (Guynn 2020), as we have witnessed in the aftermath of the U.S. 2020 presidential election. The success of such campaigns has even prompted the business of disinformation as a service, which key stakeholders, including disinformation researchers, should pay a closer look (Grossman and Ramali 2020).

Notes

1. Please note that the authors did not disclose the number of words or characters in their article.
2. Ad impressions can be considered a measure of “private level of engagement” (Aldous et al. 2019). While it is a signal of interest from the part of social media users, it does not indicate a stronger physical action, such as clicking or sharing. Because this feature was highly correlated with ad clicks, we opted to remove it from further analyses.

ORCID

Mirela Silva http://orcid.org/0000-0001-5021-0311
Juliana Fernandes http://orcid.org/0000-0002-8391-8460

References

Akbik, A., D. Blythe, and R. Vollgraf. 2018. “Contextual String Embeddings for Sequence Labeling.” In *Proceedings of 27th International Conference on Computational Linguistics, August,* 1638–1649. Santa Fe, NM: Association for Computational Linguistics.

Aldous, K., J. An, and B. Jansen. 2019. “View, Like, Comment, Post: Analyzing User Engagement by Topic at 4 Levels Across 5 Social Media Platforms for 53 News Organizations.” *Proceedings of the International AAAI Conference on Web and Social Media*, *13*(1), 47–57.

Alvarez, G., J. Choi, and S. Strover. 2020. “Good News, Bad News: A Sentiment Analysis of the 2016 Election Russian Facebook Ads.” *International Journal of Communication* 14: 3027–3053.

Amazeen, M. A., and B. W. Wojdynski. 2019. “Reducing Native Advertising Deception: Revisiting the Antecedents and Consequences of Persuasion Knowledge in Digital News Contexts.” *Mass Communication and Society* 22 (2): 222–47. doi:10.1080/15205436.2018.1530792.

Arif, A., L. Stewart, and K. Starbird. 2018. “Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse.” *Proceedings of the ACM on Human Computer Interaction* 2 (CSCW): 1–27.

Baldus, B. J., C. Voorhees, and R. Calantone. 2015. “Online Brand Community Engagement: Scale Development and Validation.” *Journal of Business Research* 68 (5): 978–85. doi:10.1016/j.jbusres.2014.09.035.

Baltas, G. 2003. “Determinants of Internet Advertising Effectiveness: An Empirical Study.” *International Journal of Market Research* 45 (4): 1–9. doi:10.1177/147078530304500403.

Barela, S. J., and J. Duberry. 2021. “Understanding Disinformation Operations in the 21st Century.” In *Defending Democracies: Combating Foreign Election*

Interference in a Digital Age, edited by edited by D. B. Hollis and J. D. Ohlin, 1–40. New York, NY: Oxford University Press. doi:10.2139/ssrn.3757022.

Banerjee, S., and A. Y. Chua. 2019. "Identifying the Antecedents of Posts' Popularity on Facebook Fan Pages." *Journal of Brand Management* 26(6): 621–633. doi:10.1057/s41262-019-00157-7.

BBC News. 2020. "Russia, China and Iran Hackers Target Trump and Biden, Microsoft Says." *BBC*, September 11. https://www.bbc.com/news/world-us-canada-54110457

Bittman, L. 1985. *The KGB and Soviet Disinformation: An Insider's View*. London: Pergamon-Brassey's.

Bittman, L. 1990. "The Use of Disinformation by Democracies." *International Journal of Intelligence and Counterintelligence* 4(2): 243–261. doi:10.1080/08850609008435142.

Brodie, R. J., L. D. Hollebeek, B. Jurić, and A. Ilić. 2011. "Customer Engagement: Conceptual Domain, Fundamental Propositions, and Implications for Research." *Journal of Service Research* 14(3): 252–271. doi:10.1177/1094670511411703.

Calder, B. J., E. C. Malthouse, and U. Schaedel. 2009. "An Experimental Study of the Relationship Between Online Engagement and Advertising Effectiveness." *Journal of Interactive Marketing* 23(4): 321–331. doi:10.1016/j.intmar.2009.07.002.

Chaouachi, S. G., and K. S. B. Rached. 2012. "Perceived Deception in Advertising: Proposition of a Measurement Scale." *Journal of Marketing Research and Case Studies* : 1–15. doi:10.5171/2012.712622.

Cialdini, R. 2006. *Influence: The Psychology of Persuasion*. Revised ed. New York: Collins.

Deebani, W., and N. Kachouie. 2018. "Ensemble Correlation Coefficient." In *International Symposium on Artificial Intelligence and Mathematics.*

De Vries, L., S. Gensler, and P. S. Leeflang. 2012. "Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing." *Journal of Interactive Marketing* 26 (2): 83–91. doi:10.1016/j.intmar.2012.01.003.

DiResta, R., K. Shaffer, B. Ruppel, D. Sullivan, R. Matney, R. Fox, J. Albright, and B. Johnson. 2019. *The Tactics & Tropes of the Internet Research Agency*. Austin, TX: New Knowledge.

Dolan, R., J. Conduit, J. Fahy, and S. Goodman. 2016. "Social Media Engagement Behaviour: A Uses and Gratifications Perspective." *Journal of Strategic Marketing* 24 (3–4): 261–277. doi:10.1080/0965254X.2015.1095222.

Eigenraam, A. W., J. Eelen, A. van Lin, and P. W. J. Verlegh. 2018. "A Consumer-Based Taxonomy of Digital Customer Engagement Practices." *Journal of Interactive Marketing* 44 (1): 102–121. doi:10.1016/j.intmar.2018.07.002.

European Commission. 2018. "A Multi-Dimensional Approach to Disinformation." Accessed 28 April 2022. https://coinform.eu/wp-content/uploads/2019/02/EU-High-Level-Group-on-Disinformation-A-multi-dimensionalapproachtodisinformation.pdf

Erevelles, S., N. Fukawa, and L. Swayne. 2016. "Big Data Consumer Analytics and the Transformation of Marketing." *Journal of Business Research* 69 (2): 897–904. doi:10.1016/j.jbusres.2015.07.001.

Evans, A., and J. Krueger. 2009. "The Psychology (and Economics) of Trust." *Social and Personality Psychology Compass* 3 (6): 1003–1017. doi:10.1111/j.1751-9004.2009.00232.x.

Fajtl, J., H. S. Sokeh, V. Argyriou, D. Monekosso, and P. Remagnino. 2019. "Summarizing Videos With Attention." In *Computer Vision – ACCV 2018 Workshops*, edited by Gustavo Carneiro, Shaodi You, 39–54. ACCV 2018. Lecture Notes in Computer Science, vol. 11367. Cham: Springer. doi:10.1007/978-3-030-21074-8_4

Fernandes, J., S. Segev, and J. K. Leopold. 2020. "When Consumers Learn to Spot Deception in Advertising: Testing a Literacy Intervention to Combat Greenwashing." *International Journal of Advertising* 39 (7): 1115–1149. doi:10.1080/02650487.2020.1765656.

Forgas, J., and J. George. 2001. "Affective Influences on Judgments and Behavior in Organizations: An Information Processing Perspective." *Organizational Behavior and Human Decision Processes* 86 (1): 3–34. doi:10.1006/obhd.2001.2971.

Gardner, D. M. 1975. "Deception in Advertising: A Conceptual Approach." *Journal of Marketing* 39 (1): 40–46. doi:10.1177/002224297503900107.

Garris, R., R. Ahlers, and J. E. Driskell. 2002. "Games, Motivation, and Learning: A Research and Practice Model." *Simulation & Gaming* 33 (4): 441–467. doi:10.1177/1046878102238607.

Gavilanes, J. M., T. C. Flatten, and M. Brettel. 2018. "Content Strategies for Digital Consumer Engagement in Social Networks: Why Advertising is an Antecedent of Engagement." *Journal of Advertising* 47 (1): 4–23. doi:10.1080/00913367.2017.1405751.

Gessler, K. 2016. "Stop Mindlessly Following Character Count Recommendations on Facebook Posts." Medium, October 13. https://kurtgessler.medium.com/stop-mindlessly-following-character-count-recommendations-on-facebook-posts-e01103b4d349

Greenwald, A. G., and C. Leavitt. 1984. "Audience Involvement in Advertising: Four Levels." *Journal of Consumer Research* 11 (1): 581–592. doi:10.1086/208994.

Grossman, S., and K. Ramali. 2020. "Outsourcing Disinformation." *Lawfare Blog*, December 13. https://www.lawfareblog.com/outsourcing-disinformation.

Guynn, J. 2020. "From COVID-19 to Voting: Trump is Nation's Single Largest Spreader of Disinformation, Studies Say." *USA Today*, October 5. https://www.usatoday.com/story/tech/2020/10/05/trump-covid-19-coronavirus-disinformation-facebook-twitter-election/3632194001/.

Han, J., J. Pei, and H. Tong. 2022. *Data Mining: Concepts and Techniques*. Cambridge, MA: Morgan Kaufmann.

Hindman, M., and V. Barash. 2018. *Disinformation, "Fake News" and Influence Campaigns on Twitter*. Miami: Knight Foundation.

Hollebeek, L. D. 2011. "Demystifying Customer Engagement: Exploring the Loyalty Nexus." *Journal of Marketing Management* 27 (7–8): 785–807. doi:10.1080/0267257X.2010.500132.

Hossain, M. Z., F. Sohel, M. F. Shiratuddin, and H. Laga. 2019. "A Comprehensive Survey of Deep Learning for Image Captioning." *ACM Computing Surveys* 51 (6): 1–36. doi:10.1145/3295748.

Howard, P., B. Ganesh, D. Liotsiou, J. Kelly, and C. François. 2019. *The IRA, Social Media and Political Polarization in the United States, 2012–2018*. Oxford, UK: University of Oxford.

Howard, P., B. Kollanyi, S. Bradshaw, and L. M. Neudert. 2018. "Social Media, News and Political Information During the US Election: Was Polarizing Content Concentrated in Swing States?" arXiv:1802.03573 [cs.SI].

Hutto E., and C. Gilbert. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the International AAAI Conference on Web and Social Media* 8 (1): 216–225. doi:10.1609/icwsm.v8i1.14550.

Isen, A., and R. Baron. 1991. "Positive Affect as a Factor in Organizational Behavior." *Research in Organizational Behavior* 13: 1–53.

Ji, Y. G., C. Li, M. North, and J. Liu. 2017. "Staking Reputation on Stakeholders: How Does Stakeholders' Facebook Engagement Help or Ruin a Company's Reputation?" *Public Relations Review* 43 (1): 201–210. doi:10.1016/j.pubrev.2016.12.004.

Jiang, M., J. Yang, E. Joo, and T. Kim. 2022. "The Effect of Ad Authenticity on Advertising Value and Consumer Engagement: A Case Study of COVID-19 Video Ads." *Journal of Interactive Advertising* 22: 1–9. doi:10.1080/15252019.2022.2035282.

Jin, Z., J. Cao, H. Guo, Y. Zhang, Y. Wang, and J. Luo. 2017. "Detection and Analysis of 2016 US Presidential Election Related Rumors on Twitter." In *Social, Cultural, and Behavioral Modeling*, edited by D. Lee, Y. R. Lin, N. Osgood, and R. Thomson. Lecture Notes in Computer Science, vol 10354. Cham: Springer. doi:10.1007/978-3-319-60240-0_2.

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kim, Y. 2018. "Uncover: Strategies and Tactics of Russian Interference in US Elections." *Young* 9 (04).

Kircanski, K., N. Notthoff, M. DeLiema, G. Samanez-Larkin, D. Shadel, G. Mottola, L. Carstensen, and I. Gotlib. 2018. "Emotional Arousal May Increase Susceptibility to Fraud in Older and Younger Adults." *Psychology and Aging* 33 (2): 325–337. doi:10.1037/pag0000228.

Langenderfer, J., and T. Shimp. 2001. "Consumer Vulnerability to Scams, Swindles, and Fraud: A New Theory of Visceral Influences on Persuasion." *Psychology and Marketing* 18 (7): 763–783. doi:10.1002/mar.1029.

Li, H. 2019. "Special Section Introduction: Artificial Intelligence and Advertising." *Journal of Advertising* 48 (4): 333–337. doi:10.1080/00913367.2019.1654947.

Liu, B. 2012. "Sentiment Analysis and Opinion Mining." *Synthesis Lectures on Human Language Technologies* 5 (1): 1–167. doi:10.2200/S00416ED1V01Y201204HLT016.

LIWC 2022. "Introducing LIWC-22". https://www.liwc.app/

Loewenstein, G. 1996. "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes* 65 (3): 272–292. doi:10.1006/obhd.1996.0028.

Loria, S., P. Keen, M. Honnibal, R. Yankovsky, D. Karesh, and E. Dempsey. 2014. "TextBlob: simplified Text Processing." *Secondary TextBlob: Simplified Text Processing* 3. Accessed 1 June 2021. https://buildmedia.readthedocs.org/media/pdf/textblob/latest/textblob.pdf

MSI (Marketing Science Institute) 2016. "Research Priorities." http://www.msi.org/research/2016-2018-research-priorities/

Martin, L. J. 1982. "Disinformation: An Instrumentality in the Propaganda Arsenal." *Political Communication* 2 (1): 47–64. doi:10.1080/10584609.1982.9962747.

McNair, B. 2017. *Fake News: Falsehood, Fabrication and Fantasy in Journalism*. London: Routledge.

Post engagement in Facebook ads. n.d. "Meta." Accessed 2 May 2022. https://www.facebook.com/business/help/735720159834389?helpref=uf_permalink

Mihailidis, P., and S. Viotty. 2017. "Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in 'Post-Fact' Society." *American Behavioral Scientist* 61 (4): 441–454. doi:10.1177/0002764217701217.

Miller, G., and A. Entous. 2017. "Declassified Report Says Putin 'Ordered' Effort to Undermine Faith in U.S. election and Help Trump." The Washington Post, January 6. https://www.washingtonpost.com/world/national-security/intelligence-chiefs-expected-in-new-york-to-brief-trump-on-russian-hacking/2017/01/06/5f591416-d41a-11e6-9cb0-54ab630851e8_story.html?utm_term=.4ee1ebad851d.

Miller, G. 2016. "Key Lawmakers Accuse Russia of Campaign to Disrupt U.S. Election." *The Washington Post,* September 16. https://www.washingtonpost.com/world/national-security/key-lawmakers-accuse-russia-of-campaign-to-disrupt-us-election/2016/09/22/afc9fc80-810e-11e6-b002-307601806392_story.html

Mollen, A., and H. Wilson. 2010. "Engagement, Telepresence, and Interactivity in Online Consumer Experience: Reconciling Scholastic and Managerial Perspectives." *Journal of Business Research* 63 (9–10): 919–925. doi:10.1016/j.jbusres.2009.05.014.

Morgan, S. 2018. "Fake News, Disinformation, Manipulation and Online Tactics to Undermine Democracy." *Journal of Cyber Policy* 3 (1): 39–43. doi:10.1080/23738871.2018.1462395.

Mueller, R. 2019. *Report on the Investigation into Russian Interference in the 2016 Presidential Election*. Washington, DC: US Department of Justice.

Munaro, A. C., R. H. Barcelos, E. C. F. Maffezzolli, J. P. S. Rodrigues, and E. C. Paraiso. 2021. "To Engage or Not Engage? The Features of Video Content on YouTube Affecting Digital Consumer Engagement." *Journal of Consumer Behaviour* 20 (5): 1336–1352. doi:10.1002/cb.1939.

Muntinga, D. G., M. Moorman, and E. G. Smit. 2011. "Introducing COBRAs." *International Journal of Advertising* 30 (1): 13–46. doi:10.2501/IJA-30-1-013-046.

Patterson, P., T. Yu, and K. de Ruyter. 2006. "Understanding Customer Engagement in Services." In Advancing Theory, Maintaining Relevance, *Proceedings of ANZMAC 2006 Conference*, Brisbane, 4–6 December.

Pennebaker, J. W., R. L. Boyd, K. Jordan, and K. Blackburn. 2015. *The Development and Psychometric*

Properties of LIWC2015. Austin, TX: University of Texas at Austin.

Penzenstadler, N., B. Heath, and J. Guynn. 2018. "We Read Every One of the 3,517 Facebook Ads Bought by Russians. Here's What we Found." *USA Today,* May 11. https://www.usatoday.com/story/news/2018/05/11/what-we-found-facebook-ads-russians-accused-election-meddling/602319002/.

Perkins, A. 2018. *Soviet Active Measures Reborn for the 21st Century: What is to Be Done*. Monterey, CA: Naval Postgraduate School.

Rambocas, M., and B. G. Pacheco. 2018. "Online Sentiment Analysis in Marketing Research: A Review." *Journal of Research in Interactive Marketing* 12 (2): 146–163. doi:10.1108/JRIM-05-2017-0030.

Recuero, R., F. Soares, and O. Vinhas. 2020. "Discursive Strategies for Disinformation on WhatsApp and Twitter During the 2018 Brazilian Presidential Election." *First Monday* 26 (1). doi:10.5210/fm.v26i1.10551.

Reeve, J., H. Jang, D. Carrell, S. Jeon, and J. Barch. 2004. "Enhancing Students' Engagement by Increasing Teachers' Autonomy Support." *Motivation and Emotion* 28 (2): 147–169. doi:10.1023/B:MOEM.0000032312.95499.6f.

Ribeiro, F., K. Saha, L. Babaei, J. Messias, F. Benevenuto, O. Goga, K. Gummadi, and E. Redmiles. 2019. "On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook." In *Proceedings of the Conference on Fairness, Accountability, and Transparency,* 140–149. Atlanta, GA: ACM. doi:10.1145/3287560.3287580.

Rich, C., B. Ponsleur, A. Holroyd, and C. L. Sidner. 2010. "Recognizing Engagement in Human-Robot Interaction." In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, 375–82. Piscataway, NJ: IEEE Press. doi:10.1109/HRI.2010.5453163.

Rid, T. 2020. *Active Measures: The Secret History of Disinformation and Political Warfare*. New York, NY: Profile Books Limited.

Röder, M., A. Both, and A. Hinneburg. 2015. "Exploring the Space of Topic Coherence Measures." In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399–408. New York, NY: Association for Computing Machinery. doi:10.1145/2684822.2685324.

Romerstein, H. 2001. "Disinformation as a KGB Weapon in the Cold War." *Journal of Intelligence History* 1 (1): 54–67. doi:10.1080/16161262.2001.10555046.

Sabate, F., J. Berbegal-Mirabent, A. Cañabate, and P. R. Lebherz. 2014. "Factors Influencing Popularity of Branded Content in Facebook Fan Pages." *European Management Journal* 32 (6): 1001–1011. doi:10.1016/j.emj.2014.05.001.

Shamsolmoali, P., M. Zareapoor, L. Shen, A. H. Sadka, and J. Yang. 2021. "Imbalanced Data Learning by Minority Class Augmentation Using Capsule Adversarial Networks." *Neurocomputing* 459:481–493. doi:10.1016/j.neucom.2020.01.119.

Schifrin, N. 2020. "Russia 'Launders' Disinformation By Using Fake Personas, U.S. Writers". *PBS NewsHour*.

Select Committee on Intelligence. 2019. *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election. U.S. Senate.*

Steinfeld, N. 2022. "The Disinformation Warfare: How Users Use Every Means Possible in the Political Battlefield on Social Media." *Online Information Review* 46 (7): 1313–34. doi:10.1108/OIR-05-2020-0197.

Strohm, C. 2017. "Mueller Probe Has 'Red-Hot' Focus on Social Media, Officials Say." *Bloomberg News,* September 13. https://www.bloomberg.com/news/articles/2017-09-13/mueller-probe-is-said-to-have-red-hot-focus-on-social-media#xj4y7vzkg

Sullivan, J. 2021. "Active Measures: The Secret History of Disinformation and Political Warfare." *International Affairs* 97 (1): 244–245. doi:10.1093/ia/iiaa211.

Tandoc, E. C., Jr, Z. W. Lim, and R. Ling. 2018. "Defining "Fake News" a Typology of Scholarly Definitions." *Digital Journalism* 6 (2): 137–153. doi:10.1080/21670811.2017.1360143.

Theodoridis, S., and K. Koutroumbas. 2008. "Pattern Recognition." *IEEE Transactions on Neural Networks and Learning Systems* 19 (2): 376.

Timberg, C., and T. Romm. 2018. "These Provocative Images Show Russian Trolls Sought to Inflame Debate Over Climate Change, Fracking and Dakota Pipeline." *The Washington Post,* March 1. https://www.washingtonpost.com/news/the-switch/wp/2018/03/01/congress-russians-trolls-sought-to-inflame-u-s-debate-on-climate-change-fracking-and-dakota-pipeline/

Tucker, E. 2020. "US Officials: Russia Behind Spread of Virus Disinformation." *AP News,* July 28. https://apnews.com/article/virus-outbreak-ap-top-news-health-moscow-ap-fact-check-3acb089e6a333e051dbc4a465cb68ee1

U.S. House of Representatives Permanent Select Committee on Intelligence (HPSCI). n.d. *Social Media Advertisements*. Accessed 12 April 2021. https://intelligence.house.gov/social-media-content/social-media-advertisements.htm

van Doorn, J., K. N. Lemon, V. Mittal, S. Nass, D. Pick, P. Pirner, and P. C. Verhoef. 2010. "Customer Engagement Behavior: Theoretical Foundations and Research Directions." *Journal of Service Research* 13 (3): 253–266. doi:10.1177/1094670510375599.

Vivek, S. D., S. E. Beatty, and R. M. Morgan. 2012. "Customer Engagement: Exploring Customer Relationships beyond Purchase." *Journal of Marketing Theory and Practice* 20 (2): 122–146. doi:10.2753/MTP1069-6679200201.

Voorveld, H. A. M., G. van Noort, D. G. Muntinga, and F. Bronner. 2018. "Engagement with Social Media and Social Media Advertising: The Differentiating Role of Platform Type." *Journal of Advertising* 47 (1): 38–54. doi:10.1080/00913367.2017.1405754.

Vosoughi, S., D. Roy, and S. Aral. 2018. "The Spread of True and False News Online." *Science (New York, N.Y.)* 359 (6380): 1146–1151. doi:10.1126/science.aap9559.

Wagner, M. C., and P. J. Boczkowski. 2019. "The Reception of Fake News: The Interpretations and Practices That Shape the Consumption of Perceived Misinformation." *Digital Journalism* 7 (7): 870–885. doi:10.1080/21670811.2019.1653208.

Yang, J., and X. Zhao. 2021. “The Impact of Information Processing Styles and Persuasive Appeals on Consumers’ Engagement Intention Toward Social Media Video Ads.” *Journal of Promotion Management* 27 (4): 524–546. doi: 10.1080/10496491.2020.1851846.

Yoon, G., C. Li, Y. G. Ji, M. North, C. Hong, and J. Liu. 2018. “Attracting Comments: Digital Engagement Metrics on Facebook and Financial Performance.” *Journal of Advertising* 47 (1): 24–37. doi:10.1080/00913367.2017.1405753.

Zhang, J., and E. Mao. 2016. “From Online Motivations to Ad Clicks and to Behavioral Intentions: An Empirical Study of Consumer Response to Social Media Advertising.” *Psychology & Marketing* 33 (3): 155–164. doi:10.1002/mar.20862.