Attachment 2

Normative and Digital Solutions to Counter Threats during National Election Campaigns (RightNets)

MILESTONE 2 - Open source report on existing technical solutions and technologies and on the pilot case analysis

Project Title	Normative and Digital Solutions to Counter Threats during National Election Campaigns
Project Acronym	RightNets
Call Identifier	PRIN 2022 PNRR - Decreto Direttoriale n. 1409 del 14-9-2022
Grant. no	P2022MCYCK
CUP	D53D23022340001
Start of Project	30 November 2023
Duration	24 months
Work Package	2
Abstract	In this report for the RightNets project, we delve into the dynamic interface between digital technology and democratic elections, particularly focusing on technological aspects of digital campaigns. We address two main research questions that explore both the technical and data-driven landscapes of online political campaigns, by means of a targeted literature review. Firstly, we examine various algorithms, methods, and technologies currently utilized to study and mitigate the influence of digital campaigning on democratic processes. The collected papers reveals a mix of traditional statistical methods and advanced Artificial Intelligence (AI) techniques employed to parse vast data volumes, enhancing our understanding of digital campaign dynamics and their impact on voter behavior. Secondly, we investigate the availability of data connected to digital campaigns about the 2022 Italian political elections. Given the goals of the RightNets' project, our targeted review aims to provide a state-of-the-art analysis that should serve as a foundation for developing tools and strategies that enhance transparency and accountability in digital campaigning.
Authors	Paolo Sernani, Emanuele Frontoni
PI	Giovanni Di Cosimo

Il Progetto *Normative and Digital Solutions to Counter Threats during National Election Campaigns – RightNets* è finanziato dall'Unione europea - Next Generation EU, nell'ambito del bando PRIN 2022 PNRR Missione 4, Componente 2, Investimento 1.1 (P2022MCYCK; CUP D53D23022340001).
I punti di vista e le opinioni espresse sono solo quelli degli Autori e non riflettono necessariamente quelli dell'Unione europea o della Commissione europea. Né l'Unione europea né la Commissione europea possono essere ritenute responsabili per essi.

A survey about innovations and algorithms to monitor the impact of social media campaigning on democracy

Paolo Sernani[1], Emanuele Frontoni[2]

[1]Department of Law, University of Macerata, Italy
paolo.sernani@unimc.it

[2]Department of Political Sciences, Communication and International Relations, University of Macerata, Italy
emanuele.frontoni@unimc.it

Abstract

In this report for the RightNets project, we delve into the dynamic interface between digital technology and democratic elections, particularly focusing on technological aspects of digital campaigns. We address two main research questions that explore both the technical and data-driven landscapes of online political campaigns, by means of a targeted literature review. Firstly, we examine various algorithms, methods, and technologies currently utilized to study and mitigate the influence of digital campaigning on democratic processes. The collected papers reveals a mix of traditional statistical methods and advanced Artificial Intelligence (AI) techniques employed to parse vast data volumes, enhancing our understanding of digital campaign dynamics and their impact on voter behavior. Secondly, we investigate the availability of data connected to digital campaigns about the 2022 Italian political elections. Given the goals of the RightNets project, our targeted review aims to provide a state-of-the-art analysis that should serve as a foundation for developing tools and strategies that enhance transparency and accountability in digital campaigning.

1 Introduction

In the evolving landscape of digital campaigning, the absence of explicit regulations presents a significant risk to the integrity of democratic systems [27]. This risk is exacerbated by the potential for foreign entities to influence election outcomes through online political advertising, alongside the challenge of covert or illicit political financing [22]. Social networks, serving

as primary platforms for digital campaigning, not only facilitate targeted propaganda, often referred to as micro-targeting, but also play a crucial role in shaping the dynamics of voter-politician relationships [25].

The RightNets ("Normative and Digital Solutions to Counter Threats during National Election Campaigns") project[1] is conceived to address these challenges by leveraging the collaborative efforts of data scientists and legal scholars. The objective is to create robust tools that can effectively monitor the origins of political advertisements on social networks and assess the extent of foreign interference in elections, with a particular focus on the Italian political landscape. Furthermore, the project will delve into regulatory measures designed to mitigate these threats and promote a fair and transparent online political competition. In addition, the project will explore the contemporary concept of political representation, reflecting on how it is influenced by digital media.

In pursuit of these goals, this report primarily serves as a literature analysis covering two significant aspects crucial to our understanding and development of digital solutions for democracy.

- First, we analyze existing technical solutions and innovations that are pertinent to social media and its overarching impact on democratic processes. This analysis not only evaluates current methodologies but also explores innovative approaches that could potentially safeguard the integrity of electoral systems.

- Second, we identify and discuss various algorithms documented in the literature that are relevant for the automatic collection of data from social media platforms. These algorithms are essential for efficiently processing and analyzing vast amounts of data, thereby enabling us to monitor and evaluate the dynamics of digital campaigning and political engagement more effectively.

Through this analysis, we aim to provide a foundation for developing tools and strategies that enhance transparency and accountability in digital campaigning.

The rest of this report is organized as follows. Section 2 describes the methodology that we used in our analysis.

2 Research questions and methodology

In this report, we embark on an exploration aimed at addressing two primary research questions that probe into the intersection of digital technologies and democratic processes:

[1]The RightNets project's website: https://rightnets.unimc.it/

RQ1. What algorithms, techniques, and methods have been applied to study the impact of digital online campaigns on elections and democracy?

RQ2. What data about the 2022 Italian political elections are available (and accessible)?

Our approach is founded on a targeted literature review rather than a systematic literature review. Unlike systematic reviews, which involve a structured collection of primary studies to systematically map a research area [14], our targeted review focuses specifically on selected facets of the broader topic. This approach is particularly suited to our objectives, as it allows us to delve deeply into specialized areas of interest without the exhaustive breadth required by systematic reviews [4]. As such, we have adopted a narrative style that is more interpretative and less structured than systematic reviews. We have curated our sources selectively, choosing studies and reports that specifically discuss the use of Artificial Intelligence (AI) and other analytical techniques in monitoring and evaluating digital election campaigns. This selection is based on a criteria that prioritizes relevance to our research questions and the potential to contribute significant insights into the use of technology in influencing democratic outcomes.

To answer the two research question through this targeted literature review, we focus on identifying and analyzing the most pertinent algorithms and methods that have been employed to understand the dynamics of online campaigns and their implications on voter behavior and election results. This involves a detailed examination of various AI-based tools and techniques, such as sentiment analysis, content filtering, and data aggregation technologies, which might have been utilized to parse and interpret large volumes of data generated during election campaigns. Moreover, we pay special attention to the availability and types of data pertaining to the 2022 Italian political elections. This includes publicly accessible datasets, proprietary data collections, and the methodologies employed by researchers to gather, clean, and analyze this data. Our analysis aims to outline the scope and limitations of the existing data, providing a foundation for understanding the digital landscape of recent elections.

In this regard, we applied the following methodology:

1. **Search on scholar indexing services**. The resource for gathering papers to analyze in our study was Google Scholar, focusing on peer-reviewed journal articles published in English. We did not restrict our search to specific journals, aiming to capture a wide range of potentially relevant studies. We restricted the search to primary studies, excluding meta-analysis. We differentiated the search terms based on the the research questions. Specifically:

(a) For RQ1, we did to searches. The first incorporated the following terms: "digital campaign" AND "elections" AND ("evaluation" OR "impact" OR "analysis" OR "empirical") targeted within the content of the articles (titles, abstracts, keywords, and text). The second added, in AND conjunction with the previous ("Artificial Intelligence" OR "Machine Learning)." We restricted the search to articles published after 2018.

(b) For RQ2, the terms "2022" AND "Italian" AND "political" AND "elections" AND ("data" OR "dataset" OR "database" OR "repository") were used.

2. **Application of exclusion criteria**. We downloaded all identified papers from the searches. Duplicate entries were removed to streamline the selection process. Each article was then evaluated by reviewing its title, abstract, and keywords, which helped in discerning their relevance to our review's thematic scope. Then, based on our research questions:

   (a) For RQ1, articles lacking empirical study techniques with objective evaluation metrics were excluded from further analysis.

   (b) For RQ2, any article that did not describe a data collection process related to the 2022 Italian political elections was omitted from our selection.

3 The study of the impact of digital campaign on election and democracy

Using the methodology described in Section 2, we collected 16 papers from January 2018 to may 2024 that described algorithms, techniques, and methods that have been applied to study the impact of digital online campaigns on elections and democracy, in order to shed a light onto RQ1. Among the included studies with identified some works which applied qualitative or statistical analysis (Described in Subsection 3.1, and other studies that exploited Artificial Intelligence (AI) and Machine Learning (ML) techniques (Subsection 3.2.

3.1 A description of the studies employing qualitative analysis and/or statistics

For example, Keller and Klinger [13] investigated the prevalence and activities of social bots among Twitter followers of seven German political parties before and during the 2017 national election campaigns. The researchers used BirdSong Analytics[2] to download Twitter account data of followers, including

[2]At the time of writing (May 2024) BirdSong Analytics is not active anymore.

metadata such as Twitter ID, screen name, number of followers, following, and tweets. The Botometer X tool[3], developed by computer scientists at the University of Indiana [26], was employed to identify bots. Its probability indicated the likelihood of a Twitter account being a bot (a threshold score of 0.76 was set to classify accounts as bots). The analysis also involved comparing the distribution of Botometer scores across the seven political parties' followers using probability density functions. Additionally, the study examined the content of tweets from the most active and most popular bots, focusing on the frequency and nature of political hashtags used during both periods. In this way, the study revealed that the share of social bots among Twitter followers of German political parties increased from 7.1% before the 2017 election campaigns to 9.9% during the campaigns. However, the percentage of active social bots remained low, with a slight increase from 2% to 1.4%. Such findings suggested that while social bots quantitatively distorted the appearance of popularity, their impact on spreading political messages during the election campaign was minimal. In addition, the authors highlighted an important research gaps: the need for standardized bot detection methods in political communication research to ensure the validity of social media data analysis.

Dommett and Temple [10] focused on digital campaigning in the 2017 UK general election, particularly examining Facebook advertising and satellite campaigns. They highlighted the importance of these online campaigns in volunteer mobilization, voter identification, and message targeting, including, in their study, data for the analysis of party activities on platforms like Twitter, Facebook, YouTube, Snapchat, and Instagram, alongside the examination of intermediary organizations such as Campaign Together and CrowdPac. However, the study used a qualitative approach, drawing on interviews, media reports, and existing literature to explore the implications of these digital practices on party structures, campaign strategies, and public expectations, with no o explicit quantitative metrics are provided in the paper to measure the impact of digital campaigning. The authors concluded that 2017 election demonstrated the evolving role of digital media in political strategies, raising questions about data usage, campaign control, and resource disparities. Nevertheless, the authors claim that digital tools extend traditional campaign activities but do not fundamentally transform them, highlighting, as a research gap the need for further exploration of the long-term implications on democratic processes.

Bright et al. [5] leveraged both cross-sectional and panel data from the 2015 and 2017 UK national elections. The colletcted dataset included almost 6,000 political candidates from 631 constituencies in England, Scotland, and Wales. To assess the impact of Twitter use on vote outcomes, the study tracks candidate Twitter activity during the campaign period, including the

[3]The project is archived since May 31, 2023.

number of tweets, replies, and retweets. The primary dependent variable is the percentage of overall votes gained by each candidate. The study uses Ordinary Least Squares (OLS) regression models to analyze the relationship between Twitter activity and vote share, incorporating control variables such as campaign spending, incumbency, and party affiliation. Additionally, first difference models are employed for candidates who competed in both elections, allowing the analysis to control for time-invariant unobserved variables. The topic modeling of tweets using Non-negative Matrix Factorization (NMF) helps classify the tweets into broadcasting or interacting communication styles. The authors concluded that Twitter use has a modest but statistically significant impact on vote share outcomes in UK national elections. Candidates with active Twitter accounts typically gained between 1 and 2 percentage points more in vote share compared to those without. The analysis also shows that increased Overall, this research provides evidence that Twitter can be an effective tool for political campaigning, though its impact is mediated by factors such as party size and the nature of the tweets.

Ohme [16] used a pre- and post-election online panel survey combined with an eleven-wave smartphone diary study during the Danish national election campaign in 2015. This mixed-method approach aimed to capture the influence of exposure to political information on campaign participation. Respondents were invited to complete a short survey on their smartphones every second day, with questions focusing on their media exposure for that day. This high-frequency measurement allowed for a detailed estimation of political media use and a granular assessment of social media content exposure. The authors used "Exposure to Political Information" (EPI) as the primary metrics, computing the frequency of exposure, calculated relative to the number of days each respondent participated in the survey, producing a metric ranging from 0 to 1. An index of campaign participation was created with values ranging from 0 to 12, and a measure of intended campaign participation was obtained before the election on an 11-point scale. Control variables included age, gender, education, income, political interest, and whether the respondent was a first-time voter. These metrics were used in OLS regression models to analyze the impact of media exposure on campaign participation, including interaction effects to differentiate between first-time and experienced voters. With these metrics, the study concluded that social media platforms play a crucial role in the media diet of first-time voters during election campaigns.

Ceccobelli [8] employed a comparative analysis to study the Facebook pages of main political leaders from advanced industrial democracies (from North America, South America, Australia, New Zealand, and the European Union). The analysis focuses on the 60 days leading up to Election Day, chosen to ensure a homogeneous comparison across 18 different election campaigns. This period was compared with a non-campaign period for each leader (of parties with more than 15% of votes) to observe differences in

communication strategies. The data collection resulted in 76,510 posts, of which 15,864 were manually coded. The posts were analyzed for personalization, policy issues, negative rhetorical strategies, and privatization of political communication. The evaluation metrics included the number of posts published per day, the degree of personalization (explicit references to the leader's persona), the percentage of posts about policy issues, and the frequency of posts with a negative rhetorical strategy (for example, personalization increased from 69.7% to 81.8% during election campaigns, and the number of posts per day rose significantly, with leaders like Capriles publishing up to 17.8 posts per day during campaigns compared to 4.0 posts per day outside campaigns). The author concluded that election campaigns significantly impact how political leaders use Facebook.

Sandoval-Almazan and Valle-Cruz [20] studied the relationship between Facebook emoticons and electoral behavior in the local government elections of the State of Mexico in 2017. The researchers collected 4128 Facebook posts from four major political parties. The methodology involved classifying Facebook reactions into positive, neutral, and negative sentiments. The metrics used in the study include the number of Facebook reactions, comments, and shares for each post. The total number of reactions for each category was calculated for each political party and the sentiment impact percentage was determined by calculating the share and comment impact as a percentage of the total for each party. This study highlighted the complex relationship between social media sentiment and electoral success, suggesting that while Facebook reactions provide insights into voter perceptions, they do not necessarily predict election results.

Coppock et al. [9] a randomized field experiment to evaluate the effects of digital advertising on vote choice during the 2018 U.S. midterm elections in Florida. The experiment was conducted over Facebook and Instagram, targeting ZIP codes in four congressional districts. Ads created by a Democratic-leaning political action committee were designed to promote the Democratic Party's values and criticize the Republican Party's stance on gun control, without mentioning specific candidates. The ads accumulated over 1.1 million impressions and over 100,000 full views. The experimental units were 210 ZIP codes, which were randomly assigned to one of three conditions: no ads (control), pro-Democratic ads emphasizing gun control (treatment video 1), and another version of pro-Democratic ads (treatment video 2). The impact of these ads was measured based on the two-party vote share for Democrats at the precinct level, using historical election data from 2012, 2014, and 2016 as covariates to enhance precision. The experiment adhered to a pre-registered analysis plan, with data clustered at the ZIP code level and standard errors adjusted accordingly. By analysing these data mthe authors concluded that digital advertising via Facebook and Instagram had negligible effects on vote choice in the context of the 2018 U.S. midterm elections. The ads tested did not significantly increase Democratic vote share.

These results emphasize the difficulty of influencing voter behavior through online ads, highlighting the limited persuasive power of such interventions in general elections. Despite the high volume of ad impressions and full views, the substantive effect on vote choice was close to zero.

Bossetta [3] presented a comparative framework to analyze political campaigning on four social media platforms: Facebook, Twitter, Instagram, and Snapchat, during the 2016 U.S. election. The methodology combines qualitative data from interviews with three Republican digital strategists and quantitative social media data from Facebook, Instagram, and Snapchat. The study incorporates interviews from the Social Media and Politics Podcast and uses as metrics the reach and engagement rates of social media posts across the analyzed platforms. Bossetta [3] also examined the overlap of content across platforms, comparing direct and edited overlaps between Instagram and Facebook posts. Additionally, qualitative metrics are derived from interviews, providing insights into the effectiveness of social media strategies and platform-specific nuances in political communication. The study concludes that digital architectures significantly influence political communication strategies on social media platforms.

Concerning the effects of social media feed algorithms, Guess et al. [12] focused on the 2020 US election. Researchers conducted randomized controlled experiments, assigning a sample of consenting users to reverse-chronologically-ordered feeds instead of the default algorithmic feeds. Participants were surveyed five times and their on-platform activities and off-platform internet usage were passively tracked. The treatment involved changing the feed ranking to chronological order, which affected the visibility of content from friends, Pages, and Groups, but not advertisements. Approximately 80% of the material seen by participants was manipulated as part of the experiment. The study aimed to understand how the chronological feed affected users' exposure to content, engagement, political attitudes, and behaviors. Time spent on the platforms, engagement (likes and comments), exposure to various types of content, and changes in political attitudes and behaviors were analyzed through unweighted sample statistics and weighted population average treatment effects (PATE). The authors found that moving users to a chronological feed substantially decreased their time on Facebook and Instagram and reduced their engagement with content. The chronological feed increased exposure to political and untrustworthy content and decreased exposure to uncivil content and content with slur words on Facebook. These results suggested that while algorithmic feed ranking strongly influences on-platform user experiences, it does not directly cause changes in political attitudes or offline behaviors within the study period.

3.2 A description of the studies employing Artificial Intelligence

Valle-Cruz et al. [24] described a hybrid approach combining affective computing and classic statistical analysis to analyze the emotional charge of Twitter posts related to the 2020 U.S. presidential elections. The study focuses on tweets from presidential candidates Donald Trump and Joe Biden, as well as trending topics that emerged during the election period. On a total of 166,637 tweets, the emotional content was extracted using the SenticNet lexicon [2] based on the Hourglass of Emotions model. This model categorizes emotions into eight primary groups: trust, anger, disgust, fear, anticipation, joy, sadness, and surprise. Morevoer, for each tweet, the emotional charge was calculated by summing the multimodal values of the words in the text. This resulted in a multi-emotional profile for each tweet, which was then averaged to determine the general emotional charge for the candidates and voters. In addition, the t-Student test and Welch's t-test, were used to compare the emotional charges between the candidates and between the candidates and the voters. This helped determine the level of emotional agreement or divergence between these groups. With this data the authors concludes that voters tended to mimic the emotional tones of the candidates they supported. This study highlights the potential of affective computing in understanding political campaign dynamics and voter behavior on social media.

Concerning the influence of algorithmic decision-making (ADM) systems on citizens' perceptions of input, throughput, and output legitimacy in EU political decision-making, Starke and Lünich [21] proposed a preregistered online experiment. A total of 572 respondents were recruited: the experiment utilized a between-subjects design with one factor and three levels: (a) independent human decision-making by EU politicians, (b) independent algorithmic decision-making by AI-based systems, and (c) hybrid decision-making by EU politicians and AI-based systems together. The primary tool for analysis was structural means modeling, which accounted for measurement errors and provided a comprehensive comparison of legitimacy perceptions across the three decision-making arrangements. The authors evaluated four key metrics: input legitimacy, throughput legitimacy, goal attainment, and decision favorability. The papers' findings suggest that citizens perceive the existing EU decision-making arrangements, involving only human politicians, as the most legitimate in terms of input legitimacy, while automatic decision-making systems alone were viewed as the least legitimate across all three legitimacy dimensions.

In addition to evalutating the impact of the use of social media on the election process, a bunch of work used AI, especially Machine Learning (ML) to forecast election results. For instance, Mameli et al. [15] presented a novel social media analytics system, called SocMINT, designed to monitor

and analyze social media activities. TData were collected from Facebook, Instagram, and Twitter using public/authenticated APIs, static data files, or web crawlers. These data were processed to extract raw post information, such as text, publication date, author, hashtags, mentions, likes, reposts, and geolocation. The authors used various KPIs, incuding Trend KPIs, like the number of published posts, engagement, a network analysis KPI, closeness centrality, that measured the average distance between nodes in the social network. In addition, entropy, borrowed from information theory, was used to gauge the uncertainty and information richness of events and interactions. Moreover, the authors used the sentiment analysis, assessing both visual and textual content. Sentiment classification used a three-value scale (0 for neutral, 1 for negative, and 2 for positive) and applied deep learning models like VGG16 for images and BERT for text. The overall sentiment of a post was a weighted average of its visual and textual sentiments, with user-defined weights. The framework was demonstrated in the case study of the Marche regional elections, that highlighted the system's ability to predict election outcomes and reflect public opinion trends.

Similarly, Brito and Adeodato [6, 7] proposed the SoMEN framework, designed to nowcast election results from social media performance metrics and ML approaches. The experiments were conducted on the presidential elections of Colombia, Mexico, Brazil (2018), and Argentina (2019), covering the first round only. Data collection commenced 300 days before the election, aggregating over 65,000 posts from the candidates' Facebook, Twitter, and Instagram profiles, alongside 195 presidential polls. The authors derived the primary metrics used in this study from social media performance, specifically user interactions on candidate posts, including likes, comments, and shares across Facebook, Twitter, and Instagram. The prediction model's performance was evaluated by comparing prediction errors with the Mean Absolute Error (MAE) threshold of 2.7 percentage points. The study used statistical tests to compare the predicted errors against poll errors. Then, the authors compared two neural networks namely a model based on General Regression Neural Networks (GRNN) and a Multi.-Layer Perceptron (MLP). The results showed that the proposed outperformed traditional polls.

Ali et al. [1] also presented a framework for predicting election results using Twitter data, specifically focusing on Pakistan's general election in 2018. The authors collected a dataset that includes 2,100 tweets used for training and 900 for testing. A deep learning model, implemented in RapidMiner, is trained on the labeled dataset. This model's architecture includes multiple layers, including convolutional, max-pooling, and fully connected layers, with dropout values set to 0.5 to promote generalization. The model is validated against traditional ML techniques, specifically Naive Bayes and Support Vector Machines (SVM), demonstrating the proposed method's comparative efficacy. The proposed model achieved a precision of 88.24%, a recall of 97.14%, and an F1-score of 92.47%. The accuracy of the model was compared

against Naive Bayes and SVM, with the deep learning approach showing an average accuracy of 71%, while Naive Bayes and SVM achieved 65% and 67%, respectively. Beyond the presented results, the authors highlighted the significant role of social media, particularly Twitter, in reflecting public opinion and predicting electoral outcomes.

Finally, Safiullah and Parveen [19] evaluated the impact of AI and ML on election campaigns, particularly focusing on their application in understanding voter behavior and influencing voter decisions. Their work evaluated internet penetration and social media's role in the 2019 Indian General Election, the use of deepfakes and automated social media bots, and, the future of AI and ML in electoral processes. However, no specific metrics are provided in the paper regarding the quantitative assessment of AI and ML tools' effectiveness in election campaigns. The study primarily discusses qualitative outcomes, such as the influence of deepfake videos and social media bots on voter perceptions and behaviors. It mentions the reach of these technologies, such as the 15 million voters contacted via WhatsApp groups with deepfake videos in the Delhi elections and the extensive data collection capabilities of applications like the Trump 2020 app, but does not provide detailed statistical data or metrics to measure the success or impact of these methods quantitatively. The authors concluded that AI and ML have significantly transformed election campaigns by enabling highly targeted and data-driven strategies. As a research gaps, the authors highlighted that the ethical implications and potential for misuse of these technologies, however, remain a significant concern for the integrity of democratic processes.

4 The case of the 2022 Italian political elections

The query on Google Scholar to find papers to answer RQ2, returned 3 manuscripts that matches the inclusion criteria. Specifically, Pierri et al. [17] collected and described ITA-ELECTION-2022, a comprehensive multi-platform dataset of social media conversations around the 2022 Italian political election. The dataset encompasses millions of posts from Facebook, Instagram, and Twitter, including metadata for TikTok and YouTube videos shared on these platforms. Data were collected over four months (July to October 2022) using public APIs and keyword-based searches. The collection includes Facebook and Instagram posts gathered through CrowdTangle[4], a public tool by Meta. Additionally, the dataset contains social media handles of Italian political representatives. Data collection involved a snowball sampling procedure, starting with seed terms like "elezioni2022," (which can be translated as *elections2022*) and "elezioni," (*elections*) and eventually expanded to include 62 keywords relevant to the election. The Twitter

[4]CrowdTanble, which can be found here https://www.crowdtangle.com/ (accessed on May 29th, 2024) will not be available starting from August 2024.

dataset alone comprises 19,087,594 tweets from 618,089 unique users. The Facebook dataset includes 1,142,812 posts from 445,461 accounts, while the Instagram dataset contains 68,078 posts from 5,274 accounts. Metadata for 22,754 unique YouTube videos and 1,903 unique TikTok videos were also collected. The dataset is publicly accessible via GitHub[5]. The authors carried out no experiments on this manuscript, even if they aim for it to enhance the understanding of social media's role in democratic processes.

Trastulli and Mastroianni [23] collected a different dataset abount 2022 Italian Political election Specifically, they collected 14 party manifestos, representing 97.7 percent of the list votes in the election. The parties included are "Alleanza Verdi e Sinistra," "Alternativa per l'Italia," "Forza Italia," "Fratelli d'Italia," "Impegno Civico," "Italexit," "Italia Sovrana e Popolare," "Lega," "Movimento 5 Stelle," "Noi Moderati," "Partito Democratico," "+Europa," "Terzo Polo (Azione – Italia Viva)," and "Unione Popolare." Optical Character Recognition (OCR) was used to convert those manifestos who were not searchable pdfs into machine-readable text. The corpus was then tokenized, cleaned of non-informative entities, and transformed into a document-feature matrix (DFM) for corpus linguistic analysis. The authors also performed an experiment corpus linguistic analysis performed on the 14 party manifestos from the 2022 Italian election. The analysis was executed using the R programming language with the quanteda family of packages for quantitative textual analysis. The corpus underwent tokenization and cleaning to remove non-informative entities and stop words. Lemmatization and part-of-speech tagging were applied to simplify the variation in words. A document-feature matrix (DFM) was created, and tools such as relative frequency analysis, collocation, and Key Word In Context (KWIC) analysis were employed. The metrics used in the experiment include the frequency of tokens, relative emphasis of keywords, and substantively meaningful collocations. The results highlight socio-economic issues, administrative reforms, and thematic keywords related to current events like the environment and energy crises. The authods highlighted thatt the 2022 Italian party manifestos present a balanced emphasis on various themes, including socio-economic issues, administrative reforms, and socio-cultural topics. It also finds that while environmental and energy issues are prominently featured, the Russian–Ukrainian war is not emphasized due to its controversial nature. The analysis presented in the manuscript confirmed that parties focus on themes historically and ideologically aligned with their core values. Unfortunately, a link to the collected dataset is not provided.

Giglietto et al. [11] introduced a workflow devised to detect, monitor, and update lists of coordinated social media accounts during peak activity periods and beyond. This workflow was applied to the 2022 Italian political

[5]Link to the dataset https://github.com/frapierri/ita-election-2022 (Accessed on May 29th, 2024)

elections, leveraging previous research on coordinated inauthentic behavior during the 2018 and 2019 Italian elections. The initial dataset included a list of 435 coordinated accounts comprising 238 Facebook Pages, 196 Facebook public groups, and 1 Instagram account. The monitoring process was conducted every 6 hours from July 28 to September 25, 2022. This yielded 1,022 highly shared or commented political posts and 272 coordinated links. Additionally, 66 new coordinated political accounts and 554 generic coordinated accounts were detected. The data collection involved posts, images, and links from these social media accounts to identify new actors and update the monitored pool. In addition, the authors described an experiment to show the monitoring and the detection of coordinated social media actors during the election campaign. Metrics used in the experiment included the number of detected coordinated accounts, the number of overly shared or commented political posts, and the number of coordinated links. According to the authors, this approach allows for near-real-time monitoring and updating, highlighting the dynamic nature of coordinated inauthentic behavior on social media. The study emphasized the importance of continuous monitoring as well, in order to capture the evolving strategies of information operations.

5 Conclusions

In conclusion, our analysis addressed two research questions through a targeted literature review. Concerning RQ1, we discovered that the majority of the collected manuscripts rely on traditional statistical methods. These methodologies serve crucial functions such as identifying bots on social media platforms [13], estimating the vote percentages that candidates receive [5, 9], and evaluating the impact of digital campaigns by administering surveys to voters [16]. Furthermore, some studies (see, for instance, [3]) enrich these surveys with quantitative data to provide a more nuanced understanding of digital campaign influences.

Additionally, our findings include a significant representation of AI and ML based techniques within the reviewed literature. While a limited number of studies [19, 21] explored the roles of automated tools in elections, a broader set of works focused on employing AI techniques to predict election results. For examples, studies such as [1, 6, 15] demonstrated the potential of AI in forecasting electoral outcomes.

As we continue to explore these complex interactions between technology and electoral processes, a deeper understanding of the tools and methods that are shaping democratic practices in the digital age is needed. The research did for this targeted literature review not only highlights the current state of the field but also underscores the diverse applications of statistical and AI methodologies in understanding and potentially enhancing the democratic process, making a first step towards future inquiries into the nexus of social

media, technology, and democracy. Future steps will include an analysis of the use of AI tools to create misinformation and fake news during election campaigns, as "the impact of disinformation has been particularly evident in recent years with reference to the influence that such informational disorder has exerted in the context of some important democratic processes at the European and global levels" [18].

Bibliography

[1] H. Ali, H. Farman, H. Yar, Z. Khan, S. Habib, and A. Ammar. Deep learning-based election results prediction using twitter activity. *Soft Computing*, 26:7535–7543, 2022. doi: 10.1007/s00500-021-06569-5.

[2] R. Biagioni. *The SenticNet Sentiment Lexicon: Exploring Semantic Richness in Multi-Word Concepts*. Springer, 2016. doi: 10.1007/978-3-319-38971-4.

[3] M. Bossetta. The digital architectures of social media: Comparing political campaigning on facebook, twitter, instagram, and snapchat in the 2016 u.s. election. *Journalism & Mass Communication Quarterly*, 95(2):471–496, 2018. doi: 10.1177/1077699018763307.

[4] J. F. P. Bridges, Z. Berger, M. Austin, N. Nassery, R. Sharma, Y. Chelladurai, T. D. Karmarkar, and J. B. Segal. Public reporting of cost measures in health: An environmental scan of current practices and assessment of consumer centeredness. Technical Briefs 19, Agency for Healthcare Research and Quality (US), Rockville (MD), 2015. URL https://www.ncbi.nlm.nih.gov/books/NBK280191/. Appendix A, Targeted Literature Review.

[5] J. Bright, S. Hale, B. Ganesh, A. Bulovsky, H. Margetts, and P. Howard. Does campaigning on social media make a difference? evidence from candidate use of twitter during the 2015 and 2017 u.k. elections. *Communication Research*, 47(7):988–1009, 2020. doi: 10.1177/0093650219872394.

[6] K. Brito and P. J. L. Adeodato. Machine learning for predicting elections in latin america based on social media engagement and polls. *Government Information Quarterly*, 40(1):101782, 2023. doi: 10.1016/j.giq.2022.101782.

[7] K. d. S. Brito and P. J. L. Adeodato. Predicting brazilian and u.s. elections with machine learning and social media data. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2020. doi: 10.1109/IJCNN48605.2020.9207147.

[8] D. Ceccobelli. Not every day is election day: a comparative analysis of eighteen election campaigns on facebook. *Journal of Information Technology & Politics*, 15(2):122–141, 2018. doi: 10.1080/19331681.2018.

1449701.

[9] A. Coppock, D. P. Green, and E. Porter. Does digital advertising affect vote choice? evidence from a randomized field experiment. *Research & Politics*, 9(1):20531680221076901, 2022. doi: 10.1177/20531680221076901.

[10] K. Dommett and L. Temple. Digital Campaigning: The Rise of Facebook and Satellite Campaigns. *Parliamentary Affairs*, 71:189–202, 2018. doi: 10.1093/pa/gsx056.

[11] F. Giglietto, G. Marino, R. Mincigrucci, and A. Stanziano. A workflow to detect, monitor, and update lists of coordinated social media accounts across time: The case of the 2022 italian election. *Social Media + Society*, 9(3):20563051231196866, 2023. doi: 10.1177/20563051231196866.

[12] A. M. Guess, N. Malhotra, J. Pan, P. Barberá, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Dimmery, D. Freelon, M. Gentzkow, S. González-Bailón, E. Kennedy, Y. M. Kim, D. Lazer, D. Moehler, B. Nyhan, C. V. Rivera, J. Settle, D. R. Thomas, E. Thorson, R. Tromble, A. Wilkins, M. Wojcieszak, B. Xiong, C. K. de Jonge, A. Franco, W. Mason, N. J. Stroud, and J. A. Tucker. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656): 398–404, 2023. doi: 10.1126/science.abp9364.

[13] T. R. Keller and U. Klinger. Social bots in election campaigns: Theoretical, empirical, and methodological implications. *Political Communication*, 36(1):171–189, 2019. doi: 10.1080/10584609.2018.1526238.

[14] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman. Systematic literature reviews in software engineering – a systematic literature review. *Information and Software Technology*, 51(1):7–15, 2009. doi: 10.1016/j.infsof.2008.09.009.

[15] M. Mameli, M. Paolanti, C. Morbidoni, E. Frontoni, and A. Teti. Social media analytics system for action inspection on social networks. *Social Network Analysis and Mining*, 12(1):33, 2022. doi: 10.1007/s13278-021-00853-w.

[16] J. Ohme. When digital natives enter the electorate: Political social media use among first-time voters and its effects on campaign participation. *Journal of Information Technology & Politics*, 16(2):119–136, 2019. doi: 10.1080/19331681.2019.1613279.

[17] F. Pierri, G. Liu, and S. Ceri. Ita-election-2022: A multi-platform dataset of social media conversations around the 2022 italian general election. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, CIKM '23, pages 5386—-5390, New York, NY, USA, 2023. Association for Computing Machinery. doi: 10.1145/3583780.3615121.

[18] O. Pollicino and P. Dunn. Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità). *federalismi.it*, 12:4–22, 2024. ISSN 1826-3534.

[19] M. Safiullah and N. Parveen. *Big Data, Artificial Intelligence and Machine Learning: A Paradigm Shift in Election Campaigns*, chapter 11, pages 247–261. John Wiley & Sons, Ltd, 2022. doi: 10.1002/9781119884392.ch11.

[20] R. Sandoval-Almazan and D. Valle-Cruz. Facebook impact and sentiment analysis on political campaigns. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, dg.o '18, New York, NY, USA, 2018. Association for Computing Machinery. doi: 10.1145/3209281.3209328.

[21] C. Starke and M. Lünich. Artificial intelligence for political decision-making in the european union: Effects on citizens' perceptions of input, throughput, and output legitimacy. *Data & Policy*, 2:e16, 2020. doi: 10.1017/dap.2020.19.

[22] C. Tenove, J. Buffie, S. McKay, and D. Moscrop. Digital threats to democratic elections: How foreign actors use digital techniques to undermine democracy. Research report, Centre for the Study of Democratic Institutions, University of British Columbia, 2018.

[23] F. Trastulli and L. Mastroianni. What's new under the sun? a corpus linguistic analysis of the 2022 italian election campaign themes in party manifestos. *Modern Italy*, 29(1):51—72, 2024. doi: 10.1017/mit.2023.45.

[24] D. Valle-Cruz, A. Lopez-Chau, and R. Sandoval-Almazan. How much do twitter posts affect voters? analysis of the multi-emotional charge with affective computing in political campaigns. In *DG.O2021: The 22nd Annual International Conference on Digital Government Research*, DG.O'21, page 1–14, New York, NY, USA, 2021. Association for Computing Machinery. doi: 10.1145/3463677.3463698.

[25] M. Vergeer, L. Hermans, and S. Sams. Online social networks and microblogging in political campaigning: The exploration of a new campaign tool and a new campaign style. *Party Politics*, 19(3):477–501, 2013. doi: 10.1177/1354068811407580.

[26] K.-C. Yang, E. Ferrara, and F. Menczer. Botometer 101: social bot practicum for computational social scientists. *Journal of Computational Social Science*, 5:1511–1528, 2022. doi: 10.1007/s42001-022-00177-5.

[27] F. J. Zuiderveen Borgesius, J. Möller, S. Kruikemeier, R. Ó Fathaigh, K. Irion, T. Dobber, B. Bodo, and C. de Vreese. Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, 14(1):82–96, 2018. doi: 10.18352/ulr.420.