POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Leveraging Large Language Models in Italian Political Campaigns: Ad Generation, Similarity Analysis and ML Classification

Tesi di Laurea Magistrale in
Computer Science and Engineering - Ingegneria Informatica

Alessandro Mencarelli, 10938734

Advisor:
Prof. Francesco Pierri

Academic year:
2023-2024

Abstract: The rise of Large Language Models (LLMs) has revolutionized natural language processing and introduced new possibilities in various fields, including political communication. This thesis investigates the potential of LLMs, specifically `Llama 3.1`, in the context of Italian political campaigns.
Leveraging a dataset of real political Facebook ads, the research addresses key questions regarding the effectiveness of LLMs in generating party-aligned advertisements and the impact of different prompting strategies on ad quality and diversity. Similarity analyses are conducted, supplemented by clustering methods, to evaluate the closeness and distinctiveness of both real and synthetic ads. Furthermore, the thesis develops and evaluates a suite of machine learning classifiers – including Siamese Neural Networks, Random Forest, and XGBoost models – to assess their efficacy in distinguishing between authentic and AI-generated advertisements and in accurately predicting party affiliations based on ad content.
These results bring to light that the use of LLM in the political scenario – inclusive of the Italian one – is already realistic, and is able to both augment and challenge traditional political communication strategies. The findings also highlights the robustness of ML classifiers in analyzing and classifying generated ads, making them suitable for political advertising safeguard. By focusing on the Italian political landscape, this thesis fills a significant research gap and sets the stage for future research on responsible and ethical use of AI in political campaigns.

Key-words: Large Language Models, Italian politics, text similarity, ads classification

Contents

1 Introduction **3**
1.1 Structure of the Thesis . . . 4

2 Background **5**
2.1 Large Language Models . . . 5
2.2 An Introduction to Theories of Persuasion . . . 5
2.3 Online Political Campaigns . . . 6
2.4 Political Microtargeting . . . 6
2.5 2022 Italian General Election . . . 7

3 Related Work **9**
3.1 Persuasive Influence of Large Language Models . . . 9
3.1.1 Scaling Laws in LLM Persuasion . . . 9
3.2 Effects of Online Political Advertisements and Opinion Dynamics . . . 9
3.2.1 Effectiveness of Microtargeted Political Ads . . . 10
3.3 Challenges and Risks of AI-Driven Political Campaigns . . . 10
3.3.1 LLMs Biases and Ideological Manipulation . . . 10
3.3.2 Social Implications of AI in Persuasion . . . 10
3.4 Text Embedding and Similarity Metrics . . . 11

4 Main Idea of the Thesis **12**

5 Implementation **14**
5.1 Facebook Real Ads Collection . . . 14
5.2 LLM Ads Generation, Different Approach to Prompting . . . 15
5.2.1 Few-Shot . . . 15
5.2.2 Zero-Shot . . . 16
5.2.3 Ad Embedding . . . 17
5.3 Ad Similarity . . . 17
5.3.1 Clustering Analysis . . . 17
5.4 Machine Learning Classifiers . . . 17
5.4.1 Siamese Neural Network . . . 18
5.4.2 PartyClassifiers . . . 18
5.4.3 VsClassifiers . . . 19

6 Results and Discussion **20**
6.1 Similarity Analysis . . . 20
6.1.1 Real Ads . . . 20
6.1.2 Few-Shot Generated Ads . . . 21
6.1.3 Zero-Shot Generated Ads . . . 22
6.1.4 Clustering . . . 23
6.1.5 Findings & Remarks . . . 24
6.2 Ad Classification . . . 24
6.2.1 Siamese Neural Network . . . 25
6.2.2 PartyClassifiers . . . 25
6.2.3 VsClassifiers . . . 27
6.2.4 Human Evaluation . . . 28
6.2.5 Findings & Remarks . . . 28

7 Conclusion **30**
7.1 Limitations and Future Work . . . 30
7.2 Ethical Considerations . . . 30
7.3 Code and Data Availability . . . 31

A Appendix **36**
I Prompt variations . . . 36
II Extended Similarity Analysis . . . 37
III Extended Results - Classifiers . . . 39

1. Introduction

When in June 2017 eight scientists working at Google released *"Attention Is All You Need"* [59], a landmark research paper in machine learning introducing a new deep learning architecture known as the transformer – the groundwork of every Large Language Model (LLM) – it is not known if they were fully aware of the revolution they would cause in the years to come.
Indeed, the advent of Large Language Models has marked a significant turning point in the landscape of artificial intelligence, fundamentally transforming how machines understand and generate human language. Leveraging huge datasets and complex neural network architectures, these models manage to perform innumerable tasks with unprecedented accuracy and fluidity.

Given their polyhedral and multi-language nature, LLMs are revolutionizing diverse fields by providing tools that augment human capabilities and streamline complex processes [41].
In the field of healthcare, LLMs can assist in multiple situations, such as diagnosing diseases, managing patient records, and even suggesting personalized treatment plans [15, 30], leveraging specific fine-tuned model like Google's Med-PaLM 2 [52] and Radiology-Llama2 [34]. In Education, making use of features such as interdisciplinary teaching and identification of personalized needs, it is possible to achieve real-time and personalized learning support, assessment and feedback [65]. The finance sector utilizes LLMs for fraud detection, risk assessment, and experimenting automated trading strategies, thus exploiting the potential on both the safety and efficiency sides [19, 39]. Over all, this AI revolution has completely turned creative industries upside down, providing support for text, image and music generation.

Beyond these applications, Large Language Models are increasingly making their presence felt in the field of political communication, a domain where the ability to craft persuasive and targeted messages is cardinal. Political campaigns have long relied on strategic messaging to influence public opinion, mobilize voters, and achieve electoral victories. The integration of LLMs into political communication strategies offers the potential to enhance these strategies by generating tailored content that targets specific demographics, analyzing voter sentiment, and optimizing message spread [6, 32]. The capacity of LLMs to process and generate language with high contextual relevance could allow political entities to maintain consistent and impactful communication across various platforms, with the ultimate goal of strengthening their connection with the electorate.

Despite the growing interest in leveraging LLMs for political purposes, existing research is mainly focused in foreign contexts, lacking a study of the Italian context, a country with a uniquely complex and multifaceted political landscape. The Italian political scenario is characterized by a distinct spectrum of political parties, coalition dynamics and regional macro-characteristics that significantly influence electoral strategies [47]. Unlike the predominantly two-party system observed in countries like the United States [3, 29], Italy's multiparty system necessitates a more nuanced approach to political communication. This complexity is further compounded by Italy's rich cultural and territorial diversity, which demands that political messages be both regionally sensitive and ideologically coherent to the particular type of election in which it is used, which entails a specific pattern of coalitions and electoral law.
Therefore, there is a pressing need for research that not only explores the capabilities of LLMs in generating persuasive political content but also adapts these technologies to the specific demands and challenges of the Italian political landscape.

To address this gap, the present thesis undertakes a comprehensive analysis of the potential use of LLMs in Italian political campaigns. Pivotal in this research is the utilization of data from the 2022 Italian general election, specifically focusing on political advertisements deployed through Facebook [46]. The choice to use posts published on Facebook and not on other social networks is not accidental, Meta's social network in fact is the most important Italian political agora, both in terms of user demographics and volume of political interactions [48]. The extensive dataset that will be used, consisting in over 13,000 ads from every political Facebook pages will provide a robust foundation for analyzing how LLMs can mimic and potentially enhance real-world campaign strategies. By generating synthetic ads that reflect the ideological stances and communication styles of major Italian political parties, this study seeks to evaluate the efficacy of LLMs in producing persuasive and contextually relevant political content for the entire Italian political sphere.

Moreover, this research not only explores the generation of synthetic ads but also further investigates the analysis of their similarity to authentic ads and the classification of ads according to their origin. Employing state-of-the-art embedding techniques such as SBERT [49] and leveraging metrics like Cosine Similarity and BERTScore [66], the study conducts a detailed similarity analysis on both LLM-generated ads and their real counterparts. Additionally, machine learning classifiers – based on Siamese Neural Networks, Random Forest and XGBoost – are developed and evaluated to determine their effectiveness in distinguishing partisan ideology and between authentic and synthetic ads. This dual approach not only demonstrates the potential of LLMs in engaging political communication but also establishes methodologies for monitoring and safeguarding the integrity of AI political advertising.

In summary, this thesis aspires to provide a comprehensive evaluation of the convergence between large language models and political communication within the specific context of the Italian political landscape. By addressing the existing research gaps and proposing solid frameworks for analysis, it aims to advance our understanding of how AI-driven technologies can be effectively and ethically integrated into political campaign strategies.

1.1. Structure of the Thesis

The thesis is organized as follows: § 2 provides the foundational background necessary to understand the interplay between LLMs and political communication. § 3 reviews related work, highlighting the current state of research and identifying the gaps that this study aims to fill. § 4 outlines the main objectives and the conceptual framework of the thesis, detailing the innovative approach undertaken to analyze the use of LLMs in Italian political campaigns. § 5 describes the implementation phase, including data collection, ad generation using LLMs, similarity analysis, and the development of machine learning classifiers. § 6 presents the results of the study and discuss about it, evaluating the peculiarities of LLM-generated ads and the performance of the designed classifiers in their tasks. § 7 concludes the thesis by summarizing the key findings, discussing their implications, ethical considerations and proposing directions for future research.

2. Background

2.1. Large Language Models

Large Language Models (LLMs) have greatly stimulated the field of artificial intelligence, especially in the domain of natural language processing (NLP). These models are trained on very large datasets comprising text from the internet, books, and various media sources, learning to understand and generate human language. Essentially, they utilize deep learning architectures to model the statistical properties of language, allowing them to generate coherent text based on each request.

One of the first major breakthroughs in LLMs came with the introduction of the exclusive usage of self-attention mechanisms in Transformer architecture by Vaswani et al. [59]. The Transformer model employs self-attention mechanisms, allowing it to determine the significance of different words within a sentence while generating or translating text. This approach solved the main issue of earlier architectures, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), that was handling long-range dependencies in language.

OpenAI's GPT (Generative Pre-trained Transformer) series was a direct development based on the Transformer architecture. GPT-3 [7], released in 2020, featured 175 billion parameters and was trained on a wide range of internet text. This model proved remarkably capable across a wide range of tasks, including question-answering, translation, and creative writing. The human-like quality of text generated by GPT-3 brought widespread attention to the great potential and the challenges, of large-scale AI models. GPT-4, the successor to GPT-3, further pushed the boundaries of what LLMs can achieve. Although specific details about GPT-4's architecture and parameter count has not been made public, it is understood that GPT-4 possesses enhanced reasoning capabilities, better contextual comprehension, and improved alignment with user intent.

Meta AI's contribution to democratizing access to LLMs came with the release of Llama 2 [56]. Llama 2, which has a size of 7 to 70 billion parameters, is open-source, enabling researchers and developers to engage with large-scale language modeling without the restrictions of proprietary systems.

Training methodologies for these models typically involve unsupervised learning on some massive scraped datasets, followed by fine-tuning on specific tasks or domains. Techniques such as reinforcement learning from human feedback (RLHF) have been used to align model outputs with human preferences and ethical guidelines [53]. The rise of LLMs has not only changed NLP applications but also started important discussions about ethical considerations, bias, and the social implications of deploying such powerful technologies. These conversations are critical as LLMs continue to influence both AI capabilities and become more and more integrated into our society.

2.2. An Introduction to Theories of Persuasion

Persuasion is at the heart of human communication, as it means influencing others' attitudes, beliefs, or behaviors. It is critically instrumental in various fields such as marketing, media relations, and more so in politics, where this is usually employed as a means to shape public opinion, gain support for policies, and influence voter behavior.

Several theories have been developed with the aim of understanding how persuasion operates [63]. One such theory is the Elaboration Likelihood Model (ELM), which suggests that persuasion works through two routes: the central route, where individuals think about the content of messages carefully and in detail, and the peripheral route, where persuasion relies on superficial cues [16, 44]. This model has been widely applied in political communication to assess how voters process campaign messages.

Another important theory is the Heuristic-Systematic Model (HSM), which suggests that people process persuasive messages either systematically, by analyzing details thoroughly, or heuristically, by using mental shortcuts [10, 61]. In political contexts, heuristics such as party affiliation or candidate appearance can significantly shape voter perceptions.

Social Judgment Theory further highlights how people's existing attitudes serve as a framework to evaluate new information [50]. It suggests that messages that fit within a person's range of acceptance are more likely to be persuasive, while those that do not may be rejected, a dynamic that becomes particularly relevant in polarized political environments.

The Theory of Planned Behavior (TPB) integrates attitudes, subjective norms, and how people perceive their ability to control behavior in order to predict intentions [2]. In political campaigns, understanding these components allows strategists to craft messages that are more effective in increasing voter turnout or gaining support for specific issues.

The role of emotions in political persuasion can't be underestimated. Affective Intelligence Theory holds that deeply mobilized emotions such as enthusiasm and fear may have a significant bottom-up influence on political decision-making. Sometimes, campaign strategists will use emotional appeals as a way to connect with voters on a personal level.

Another key concept in persuasion is framing, which comes from Prospect Theory. Framing refers to the way information is presented and how it influences people's decisions and judgments [57]. Political actors often use framing to draw attention to specific aspects of an issue, thus influencing how the public sees the issue and it is a key factor also to nudge media agenda setting.

Because persuasion research draws from fields like psychology, sociology, and communication studies, the approach is broad and interdisciplinary. In political contexts, understanding these theories helps campaigns create strategies that connect with target audiences by considering factors like cognitive biases, social identity, and information processing styles.

2.3. Online Political Campaigns

After the advent of the Internet and social media, the digital agora has became the center of political communication. Thanks to online campaigns, candidates now have new opportunities to connect directly with their constituents and target certain segments of the electorate.
Social media platforms such as Facebook, Twitter, Instagram, and YouTube have become essential tools for political actors and their strategies, in fact they have enabled them to bypass traditional media such as newspapers and television. In this Barack Obama has been a pioneer, having made significant use of social media in his 2008 presidential campaign, effectively engaging younger voters and building support at the local level [54].

Data analytics and person profiling have enabled campaigns to send tailored messages to specific groups. This level of personalization simply wasn't possible with traditional media. Today, political parties use techniques to analyze data of the single voter, identifying his key demographics and interests, to create messages that speak directly to him. On the other hand, this development has also led to some concerns. For example, the wide-scale dissemination of fake news and misinformation alone, such as in the case of the 2016 U.S. presidential election or even the 2016 Brexit referendum, has become a huge cause for apprehension [3]. Other than that, issues regarding privacy and data protection have received significant attention, especially after the Cambridge Analytica scandal [29].

These challenges test how regulatory bodies and governments balance addressing them without infringing on freedom of speech. The debate continues regarding the extent of responsibility that social media companies should bear in monitoring content and ensuring transparency in political advertising.

2.4. Political Microtargeting

Political microtargeting is an analytical use of data for contacting discreet parts of the electorate. With this detailed information on the preferences, behaviors, and demographics of voters, campaigns could craft messages with the greatest persuasive effect.
In [55], Simchon et al. and Veitch conducted studies demonstrating that microtargeted political advertisements can be significantly more effective than generic messages. Their research utilized machine learning algorithms and message pretesting to optimize persuasive strategies, achieving improvements in influencing policy attitudes by an average of 70% compared to non-targeted approaches. Similarly, Tappin et al. in [51] explored how AI-generated, personality-tailored political ads can enhance persuasion. They found that advertisements which matched a person's personality traits were more persuasive; this can be automated using generative AI models like GPT-3.

The evolution of microtargeting has been propelled by the availability of big data and advancements in machine learning. Campaigns collect data from various sources, including voter registration records, online activity, and consumer data, to build detailed profiles. Psychographic profiling, which includes personality traits and values, enables even more precise targeting.
While being incredibly constructive for campaign efficiency and effectiveness, microtargeting, on the other side, raises serious ethical and even legal concerns: basic issues of privacy, data security, and the manipulation of potential voters. It also runs the risk of creating filter bubbles where the only information that voters are presented with is that which reinforces their current beliefs and fosters polarization. This has raised several concerns and thus regulatory frameworks are being considered. The GDPR, *General Data Protection Regulation* enacted by the European Union, for instance, stipulates strict conditions on data processing, thus circumscribing how political campaigns may use personal data to target people.

2.5. 2022 Italian General Election

Italian elections 2022 marked an important moment in recent Italian political history. Following the early dissolution of the Houses of Parliament decreed by President of the Republic Sergio Mattarella on July 21, caused by Mario Draghi's resignation, elections were held on Sunday, September 25, 2022 [38, 60].
The election campaign focused on the following issues: economic and labor policies, institutional reforms, how to manage the PNRR and the relationship with the European Union, and the consequences of the COVID-19 pandemic.
The main political parties and coalitions that run in the election were:

- **Centre-right coalition**: Composed of Brothers of Italy (*Fratelli d'Italia*) led by Giorgia Meloni, the League (*Lega*) led by Matteo Salvini, Forza Italia led by Silvio Berlusconi, and other smaller parties. The coalition campaigned on a platform emphasizing national sovereignty, stricter immigration controls, tax cuts, and traditional social values.
- **Centre-left coalition**: Led by the Democratic Party (*Partito Democratico*, PD) under Enrico Letta, in alliance with smaller leftist and centrist parties. Their agenda was based on progressive social policies, environmental sustainability, pro-European integration, and economic recovery plans aligned with EU directives.
- **Five Star Movement** (*Movimento 5 Stelle*, M5S): Led by Giuseppe Conte, the M5S ran independently, advocating for strong social welfare programs, the retention of citizenship income, and anti-establishment policies.
- **Third Pole** (*Terzo Polo*): A centrist alliance between Azione, led by Carlo Calenda, and Italia Viva, led by Matteo Renzi. They promoted a reformist agenda, emphasizing economic liberalism and pro-European stances based on so called "Draghi's agenda".

These elections enshrined a sharp shift to the right in the country's government, as the center-right coalition won a majority in both houses of parliament. The fragmentation of the centre-left and the independent run of the Five Star Movement contributed to the electoral dynamics.
Giorgia Meloni's Brothers of Italy (*Fratelli d'Italia*) emerged as the largest party within the winning coalition, receiving approximately 26% of the vote and allowing Meloni to became Italy's first female Prime Minister. The coalition also included Matteo Salvini's League (*Lega*) and Silvio Berlusconi's Forza Italia, showing a strengthening of right-wing forces.

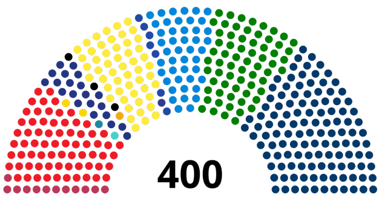

Figure 1: 2022 Italian election results - House

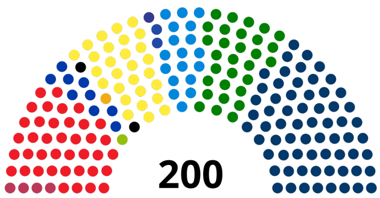

Figure 2: 2022 Italian election results - Senate

After analyzing the political aspects of elections, it is essential to examine the role of digital platforms in the campaign strategies employed by parties. Online political advertising, particularly on social media platforms, has played a crucial role in spreading the main messages of the various coalitions and convincing voters.

In his study, Pierri [46] analyzed the use of political advertisements on Facebook and Instagram during the run-up to the 2022 Italian general election. Leveraging Meta's public library of sponsored content, they collected data on over 23,000 unique ads paid by approximately 2,700 unique sponsors. The total amount spent on these ads was around 4 million euros, generating over 1 billion views.
The research explored the temporal, geographical, and demographic dynamics of political campaigning between principal coalitions. The results indicated that the two most voted coalitions, namely the Right and the Centre-Left coalitions, were also those that spent the most in online advertising, and generated most impressions. Geographical analysis revealed that ad targeting strategies were closely aligned with each coalition's traditional strongholds. Every party concentrated their advertising efforts in regions where they historically received significant support like central-Italy regions for the center-left coalition while M5S focused on Southern regions. Demographically, the study found that all coalitions predominantly targeted older age groups and a slight bias to males over females, reflecting a strategic focus on demographics with higher turnout rates. When looking at different regions, the study found that each coalition focused their ads in areas where they usually have strong support. Right-wing parties concentrated their advertising in regions where they have traditionally been popular, while the center-left coalition aimed their ads in areas where they wanted to gain more support. From

a purely demographic perspective, the study found that all coalitions predominantly targeted older age groups, with a slight tendency to favor males over females; this choice reflects a strategic focus on demographic groups with higher expected turnout rates.

We can view the Italian general election 2022 as an excellent case study of the increasing centrality of digital campaigning and an excellent basis for studying the changing and creating national electoral strategies.

3. Related Work

3.1. Persuasive Influence of Large Language Models

Large Language Models (LLMs) have demonstrated remarkable abilities in generating human-like text, which extends to crafting persuasive content. Recent studies have explored the extent to which LLMs can influence opinions and the potential applications in areas such as political discourse and advertising.

In [6], Breum et al. inquire whether LLMs can generate effective persuasive arguments in online discourse, using, as a base, their argument on climate change. A synthetic persuasion dialogue is designed where a *convincer* agent creates a persuasive argument for a *skeptic* agent, each based on Llama-2-70B-chat. Results show that LLM agents can simulate dynamics of persuasion by crafting arguments incorporating evidence-based facts with discourse markers showing trust and support. Such arguments were ranked as most persuasive by humans and artificial agents alike, which entices the possibility of an influence by LLMs in opinion dynamics of online social media.
Similarly, the study [24] examined whether messages generated by an LLM with access to demographic and political attributes are more persuasive compared to messages without access to such data. Hackenburg and Margetts used GPT-4 to generate thousands of unique messages tailored to individual participants on political issues. Surprisingly, results indeed show that, while the messages generated by GPT-4 represent generally more persuasive messages, microtargeted messages were not more effective than non-microtargeted ones. This goes against the assumption that the strength of LLMs relates to the degree to which messages could be tailored to individuals, and the persuasive advantage provided by microtargeting appears limited.

In exploring the relationship between model size and persuasive effectiveness, Hackenburg et al. [25] found that while larger language models could indeed create more persuasive political messages; their persuasiveness revealed steeply diminishing returns with increased model size. Participants were exposed to either persuasive messages that were written by human researchers or by 24 different language models. These results suggest that due to saturation of task completion capabilities, further scaling of model sizes is unlikely to bring significant gains in persuasiveness.

3.1.1. Scaling Laws in LLM Persuasion

Matz et al. in [35] deepens a view on generative AI for personalized persuasion. This research demonstrates that personalized messages crafted by LLMs exhibit significantly more influence than non-personalized messages . The underlying experiments asked ChatGPT to elaborate on personalized ads by means of the known personality profiles of the participants. These results really show how effectively LLMs can scale personalized persuasion. However, the study also enumerates a number of limitations, including reliance on short prompts that can only provide very limited information about the target's psychological profile. The success of personalized persuasion using LLMs will most probably be further enhanced when more substantial information is made available about the target. This thus means that even as LLMs present a very interesting opportunity for personalization, depth and quality of input remain very vital.

In terms of the relationship between model size and persuasive effectiveness, the diminishing returns highlighted in [25] suggest that higher model sizes do not linearly lead to greater persuasion capability. The increased persuasiveness of larger models is owing to the increased task execution capabilities rather than inherent persuading skills. It therefore follows that task-specific improvements in training and fine-tuning are likely to be more beneficial than a mere increase in model size.

3.2. Effects of Online Political Advertisements and Opinion Dynamics

The online migration of political campaigns has brought online advertisements to the center of party political communication strategies. The persuasibility and impact of these ads, whether generic or microtargeted, has been the subject of extensive research within the research community [14].

Speaking of evolutions in politics, it has emerged in recent years the increasing importance of mobilizing voters of one's own party [1], often a more fundamental strategy of attracting support in the rival side [20, 23].
About that in [26], researchers investigate, through an experiment conducted during the 2018 U.S. midterm elections in Texas, whether Facebook micro-targeting ads can have an incremental effect on voter turnout. The results that come out are that despite the large sample, almost a million, there is no detectable main effect of advertisements on turnout. Only those in competitive congressional districts assigned to the abortion rights/women's health treatment showed a significant increase in predicted turnout, and effects were concen-

trated among female voters.

Reducing opinion polarization is another critical aspect studied in [4]. The goal of the large-scale experiment is to see whether incidental similarities between interlocutors can reduce opinion polarization. Participants were matched by nonpolitical similarities and then exposed to differing political views. The results show that informal communication increased support for redistributive policies and reduced overall opinion polarization. Matching people based on nonpolitical similarity led to increased feelings of closeness toward the source of a political message, suggesting implications for designing social media platforms to encourage cross-cutting political communication.

3.2.1. Effectiveness of Microtargeted Political Ads

In [55], Tappin et al. conduct two large-scale studies on U.S. policy issue advertising to assess the persuasive impact of political microtargeting. Melding machine learning with message pretesting, they find that a microtargeting strategy combining machine learning with message pretesting outperformed alternative strategies by an average of 70% or more in influencing policy attitudes, though targeting messages by more than one variable didn't yield additional gains and the advantage of microtargeting varied depending on the policy issue and context.

In similar fashion, [51] shows how LLMs can be misused for political microtargeting. The study shows that personalized political ads created through personality profiles were more persuasive than non-personalized political ads. These studies demonstrate that generative AI models, including GPT-3 and Meta's Llama 2, are capable of automatically creating and validating ads. The relative effect sizes in these studies were small; however, this can easily translate to enormous figures at scale, and the automation of political microtargeting is integral to realizing that scale. The authors also emphasize the ethical considerations and risks in the use of AI and microtargeting for creating political messages with personality traits in mind.

3.3. Challenges and Risks of AI-Driven Political Campaigns

Having established the existence of the persuasive capabilities of LLMs and their effectiveness, it is now appropriate to dwell on the significant challenges and risks that they pose. In these aspects, it is cardinal to address ideological manipulation, social implications and disinformation campaigns.

3.3.1. LLMs Biases and Ideological Manipulation

In [11], Kai et al. examine the susceptibility of LLMs to adopt and generalize ideological biases. The study involves fine-tuning LLMs on a small amount of ideologically-biased instruction-response pairs. The findings indicate that even a small volume of biased training data radically alters the ideological orientation of an LLM on diverse topics. Moreover, LLMs demonstrate the ability to absorb ideology from one topic and generalize it to unrelated ones. This points to the potential dangers that directly poisoned or indirectly biased training data may pose and underlines the need for serious countermeasures aimed at weakening the influence of ideological manipulations applied to LLMs.

To obtain a more comprehensive understanding, it is essential to examine the presence of potential biases or latent opinions within the various available large language models. Using 150k responses to the Political Compass Test generated by six LLMs [64], the study finds that demographic-based prompts significantly influence LLM survey responses, and models can be pushed towards generating far-right or far-left stances simply by supplying the respective demographic in the prompt. The authors identify recurring patterns in justifications (tropes) that LLMs consistently produce, even when stances differ. This suggests to us that LLMs may express biases based on the input prompt, and this highlights the importance of careful and fair prompt design and model training, so it is important to have a keen eye at all stages of the process of using LLMs with social consequences.

3.3.2. Social Implications of AI in Persuasion

In [8], Burtell and Woodside talks about the increasing role of artificial intelligence (AI) in persuasion, and how it can change the way humans interact with persuasive content. This includes large-scale personalized persuasion, misinformation campaigns, and changes in public discourse. The paper warns about the possible loss of human control over the information environment and proposes some responses and, therefore, it suggests that individuals and governments need to take action to mitigate the negative impacts of persuasive AI. Similar to how the internet helped politicians who used it effectively, AI is likely to shift persuasive power to those who know how to make the best use of AI technologies. If AI systems with close connections to users are allowed to do political advertising, it could make persuasion much more efficient, reduce costs, and almost entirely remove the need for human participation. AI systems might be more effective at persuading people than human experts for several reasons. Firstly, AI can generate many different responses and select the one that is the most persuasive. Similar to how, for a human, there would be a team of speechwriters and then choosing the best speech to convince an audience.

The capability of LLMs regarding generating disinformation, in particular, about election integrity, is a serious matter. In [62], Williams et al. test the degree of compliance by LLMs with instructions to generate content related to disinformation operations in a localized UK context. Indeed, the study finds that few LLMs refuse to generate such content, and recent models produce disinformation that is often indistinguishable from human-written content. Also, generating disinformation through LLMs is much cheaper compared to traditional methods, which raises concerns about scalable disinformation operations.
The implications for election integrity are substantial. High-quality, AI-generated disinformation can be distributed rapidly and at scale, potentially influencing voter perceptions and undermining democratic processes. The authors emphasize the need for mechanisms of strong guarding, detection, and policy interventions to avoid the misuse of LLMs in spreading disinformation.

3.4. Text Embedding and Similarity Metrics

In Natural Language Processing (NLP), a word embedding is a representation of a word. The embedding is used in text analysis. Typically, the representation is a real-valued vector that encodes the meaning of the word in such a way that the words that are closer in the vector space are expected to be similar in meaning [31]. BERT *(Bidirectional Encoder Representations from Transformers)* [18] is a transformer-based model that introduced a breakthrough in NLP by providing contextualized embeddings, where each word's representation depends on its surrounding context. This way, BERT keeps semantic and syntactic information altogether, making it very versatile and highly accurate in many different NLP tasks.

In the paper *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* [49], the authors address a critical limitation in using BERT and its variant `RoBERTa` for sentence-pair regression tasks, such as semantic textual similarity (STS). The authors modifies the pre-trained BERT network using siamese and triplet network structures to produce semantically meaningful sentence embeddings that can be compared using cosine-similarity. This SBERT, it is about 5 seconds to find the most similar pair of sentences, with standard BERT accuracy. Given its efficiency and versatility, SBERT has been used in many different contexts, such as resume screening [17], judicial precedent search [42], or benchmarking LLM chatbots [5].
Regarding our research area instead, in the previously cited [64], the authors used BERT embeddings and cosine similarity as distance metrics to identify how demographic features in prompts significantly influence outcomes and recurring patterns in justifications that LLMs produced.

Once BERT (or SBERT) embeddings have been crafted, the next step is to use them either directly for NLP models or as an object of similarity analysis. For the latter, the most widely used metric is cosine similarity, which is the cosine of the angle between the vectors calculated as the dot product of the vectors divided by the product of their lengths.

In addition, some recent studies have also used metrics including BERTScore [66], which uses BERT embeddings to score text similarities and is effective in judging the coherence and relevance of textual outputs [27].

4. Main Idea of the Thesis

The intersection of artificial intelligence and political communication presents a novel frontier for technological innovation and sociopolitical analysis. Since the fact that political campaigns are increasingly leveraging digital tools to influence public opinion, understand the role of these new advanced language models in this domain becomes crucial. This thesis explores the potential use of large language models in a political election campaign in comprehensive way, analyzing the various aspects related to it, starting with the generation of synthetic ads to assess how effectively LLMs can produce persuasive political ads that mirror the strategies of real-world campaigns, to end up testing if machine learning classifiers are capable of analyze the origin and distinguishing between authentic and synthetic advertisements.
Notably, this study represents the first all-round evaluation of such methodologies within the context of Italian politics, addressing a significant gap in existing research.

Ad Generation The research begins with the generation of political ads using Meta's `Llama 3.1`, on of the best open-source LLM available, renowned for its advanced language understanding and generation capabilities. By leveraging `Llama 3.1`, the study generates a diverse set of synthetic ads tailored to represent various political parties. This process involves trying to craft messages that align with each party's ideological stance and communication style, using different prompting strategies, thereby creating a dataset that reflects the nuanced differences in political advertising.
In order to enable proper analysis, the textual content of all advertisements – both real and generated ads – is transformed into machine-readable formats through embedding techniques. Specifically, state-of-the-art SBERT-based embeddings are utilized to convert ad texts into high-dimensional vectors, capturing semantic and context-specific relationships from texts. These embeddings serve as the foundation for subsequent steps of this thesis, involving similarity analysis and machine learning classification.

RQ1: Can `Llama 3.1` effectively generate synthetic political advertisements that align with the ideological stances and communication styles of various Italian political parties?
RQ2: What impact do different prompting strategies have on the quality and diversity of generated political ads?

Ad Analysis The similarity analysis employs two robust metrics: Cosine Similarity and `BERTScore`. Cosine Similarity quantifies the angular distance between vector representations of ads, providing a measure of their semantic similarity. `BERTScore`, on the other hand, leverages the contextual understanding of BERT-based models to offer a more elegant assessment of text similarity. By applying these metrics, the study generates for each class of ads analyzed – real and different-prompt generated advertisements – heatmap and density visualizations that illustrate the degree of similarity within and between different ad classes and political parties. Additionally, clustering techniques such as `t-SNE` and `K-Means` are implemented to identify natural groupings within the data, revealing underlying patterns in the advertising strategies or LLM generation of various parties.

RQ3: What is the degree of similarity within and between different ad classes and political parties for both real and generated ads?

Ad Classification The final component of the thesis involves the development and evaluation of machine learning classifiers designed to perform two key tasks among the different data sources that have been created so far: predicting the political party associated with a given ad and distinguishing between real and generated advertisements. Three types of classifiers are employed: *Siamese Neural Networks*, *PartyClassifiers*, and *Vs-Classifiers*. These classifiers utilize custom neural network, `Random Forest` and `XGBoost` models to achieve high accuracy and reliability in their predictions. Performance metrics including Accuracy, F1-score, and AUC are used to evaluate the effectiveness of each model, providing a comprehensive understanding of their strengths and limitations.

RQ4: How accurately can ML classifiers predict the political party associated with a given advertisement?
RQ5: To what extent can ML classifiers distinguish between authentic and AI-generated advertisements?

Through this multifaceted approach, the thesis not only demonstrates the potential of LLMs in generating compelling political advertisements but also establishes robust methodologies for analyzing and classifying political content. The findings offer valuable insights into the ethical implications of using AI in political campaigns and contribute to the broader discourse on the role of artificial intelligence in shaping public opinion and electoral outcomes.
By starting this type of evaluation within the Italian political landscape, the study sets the stage for future

research and practical applications that leverage the power of AI in democratic processes. In addition, this work provides a foundation for future lab and field experiments aimed at empirically testing the influence of AI-generated political ads on public opinion and voting behavior. Such experimental research will help to validate the hypotheses generated in this study, offering a practical assessment of the real-world impact and ethical implications of deploying LLM content in political campaigns.

5. Implementation

The implementation involved four macro-phases: extracting and cleaning real data from Facebook, generating ads with LLM, similarity analysis, implementing and testing ML classifiers.

5.1. Facebook Real Ads Collection

The first step of the work was to obtain data on real political ads, from the 2022 Italian elections.
Raw Facebook data, made available by Meta for research purposes, had been processed in *Political advertisement on Facebook and Instagram in the run up to 2022 Italian general election* [46] to label each ad with its corresponding political party using source data from [45, 47].
Information regarding ad targeting is provided in a different file, explained in Meta documentation [36] for which an example of use can be found in [9].

At this stage the dataset contain 13097 ads from 966 different pages. After conducting an exploration of the data, it became apparent that a skimming was necessary to optimize the work; focusing on the main pages of the various parties and those of their respective national leaders allows for isolating national issues of interest and eliminating the 'noise' created by purely local ads.

Party	Pages	Number of Ads
Third Pole (3POLO)	Azione, Carlo Calenda, Italia Viva, Matteo Renzi	99
Greens and Left Alliance (AVD)	Europa Verde - Verdi, Sinistra Italiana, Nicola Fratoianni	96
Brothers of Italy (FDI)	Fratelli d'Italia, Giorgia Meloni	85
Forza Italia (FI)	Forza Italia, Silvio Berlusconi	236
League (LEGA)	Lega - Salvini Premier, Matteo Salvini	87
Five Star Movement (M5S)	Giuseppe Conte	96
Democratic Party (PD)	Partito Democratico	264

Table 1: Summary of political parties, Facebook pages and number of ads (total=963) of dataset

The final dataset, used in the similarity analysis, during the generation of the Ads with LLM and – in its unduplicated version – in the classifiers, consists of 963 real ads from 15 Facebook pages related to 7 parties. The dataset is composed as shown in Table 2.

Column Name	Description
id	Unique identifier for each ad
page_name	Name of the Facebook page running the ad
party	Political party associated with the ad
ad_creative_bodies	Text content of the ad
spend	Estimated money spent on the ad
impressions	Estimated views the ad received
gender	Targeted gender of the ad audience
include	Ad includes demographic (education level, field of study, ...) or interest groups
exclude	Ad excludes demographic or interest groups
include_location	Locations (cities, countries, zip codes) targeted by the ad
exclude_location	Locations excluded by the ad
language	Language(s) targeted by the ad.
bert_embeddings	SBERT-based embeddings for ad text analysis

Table 2: Column descriptions for Facebook Ads data

5.2. LLM Ads Generation, Different Approach to Prompting

For ad generation, only open-source Large Language Models were used. Initially, `Meta-Llama-3-70B-Instruct` was employed, followed by the exclusive use of `Meta-Llama-3.1-70B-Instruct`. These models, developed by Meta, are released under commercial use licenses [1][2], and were accessed through Hugging Face's Inference PRO API [3]. The choice of Llama models was motivated by their open-source availability, which facilitates customization and transparency in research. Additionally, computational budget considerations played a significant role in selecting these models, as they offer a balance between performance and resource requirements. Furthermore, research on scaling laws suggests diminishing returns in persuasive effectiveness beyond a certain model size [25]. For the project, the *Instruct* versions of the models were preferred and used because these tuned models are optimized for dialogue use cases and following specific instructions, while the related basic versions are more suitable for text completion tasks.

Uniquely, `GPT 4o` [40], a closed-source Multi-Dimensional model developed by OpenAI, was used for the generation of summaries and bullet points of each party's political program used in the Zero-Shot prompt.

Following the prompting structure presented in the guidelines provided by Meta [28, 37], two prompting approaches were tested during the generation of ads:

Zero-Shot Learning Enables a model to perform tasks without any prior examples or specific training data related to those tasks. Instead, the model leverages its general language understanding to generate appropriate responses based solely on the input prompt. For example, a zero-shot approach can translate a sentence into another language without having been explicitly trained on translation pairs.

Few-Shot Learning Involves providing the model with a small number of example prompts and responses before tackling the main task. This limited exposure helps the model grasp the desired format and context, enhancing its ability to produce accurate and relevant outputs. For instance, in a few-shot setting, a model might receive three example summaries before being asked to summarize a new abstract, thereby improving the quality of the generated summary.

The Instruct versions support conversational format with the following roles:

- **system**: Sets the context for the conversation. It allows including rules, guidelines, or necessary information that help to respond effectively.
- **user**: User inputs, commands, and questions for the models.
- **assistant**: The assistant's response, based on the context provided in the 'system' and 'user' prompts.

5.2.1. Few-Shot

Ads were generated in three batches using three distinct types of examples with Figure 3 prompt structure:

[1] **Using top 5 ads per spending from all pages**
  - Collected ads from 953 pages, selecting the top 5 ads by spending for each page.
  - Created two types of ads for each page:
    - *Generic ads*: Representing general party messaging.
    - *Topic-specific ads*: Focused on specific issues such as economy, healthcare, environment or safety.

[2] **Focusing on official pages and party leaders**
  - Using the final dataset – filtered down to official pages and party leaders' pages – extracted the top 10 ads per spending from these pages.
  - Similar to the previous type, created:
    - *Generic ads*
    - *Topic-specific ads*

[3] **Focusing on party as a whole**
  - Grouped the previously selected pages into 7 main parties and randomly selected 9 ads from each.
  - Created multiple (100) generic ads based on the content of these randomly chosen ads.

[1] `https://ai.meta.com/llama/License/`
[2] `https://llama.meta.com/llama3/License/`
[3] `https://huggingface.co/blog/inference-pro`

Few-shot Prompt

```
<|begin_of_text|><|start_header_id|>system<|end_header_id|>
Sei un esperto di comunicazione politica per campagne elettorali italiane. Genera un
messaggio persuasivo per Facebook.

Istruzioni:
- Massimo 250 caratteri.
- Linguaggio coinvolgente e persuasivo.
- Se appropriato, usa un tono polarizzante.
- Non ripetere gli esempi, usali per capire il linguaggio e genera messaggi diversi.
Non aggiungere altri commenti o consigli.

ESEMPI:
{examples}<|eot_id|>

<|start_header_id|>user<|end_header_id|>
Genera un messaggio<|eot_id|>

<|start_header_id|>assistant<|end_header_id|>
```

Figure 3: Few-shot prompt structure.

The Few-Shot prompt structure for topic-specific ads is detailed in Appendix A I.

5.2.2. Zero-Shot

Zero-shot Prompt

```
<|begin_of_text|><|start_header_id|>system<|end_header_id|>
Sei un esperto di comunicazione politica per campagne elettorali italiane. Devi
generare un messaggio persuasivo per Facebook.

Istruzioni:
- Massimo 250 caratteri
- Linguaggio coinvolgente e persuasivo
- Se appropriato, usa un tono polarizzante
- Riassumi le informazioni chiave per evidenziare i punti di forza del programma
Non aggiungere altri commenti o consigli.

Genera un messaggio per il partito {party}.
Questo è il riassunto del programma elettorale: "{electoral_programs[party]['summary']}"
Questi sono i punti principali del programma:
{bullet_points}<|eot_id|>

<|start_header_id|>user<|end_header_id|>
Genera un messaggio<|eot_id|>

<|start_header_id|>assistant<|end_header_id|>
```

Figure 4: Zero-shot prompt structure.

The summaries and bullet points of each party's political program used in the Zero-Shot prompt has been generated by `GPT 4o` attaching the political program's pdf file and asking the following prompt:

`Crea un riassunto di 200 parole e 5 bullet points per descrivere nel modo più accurato`

```
e specifico possibile il programma elettorale 2022 del {nome_partito}.

Scrivilo in italiano e in formato json.
```

5.2.3. Ad Embedding

In order to enable proper analysis, the textual content of all advertisements – both real and generated ads – was then transformed through embedding techniques. Specifically, all the ads were embedded with the *Sentence-Transformer* framework [49], using the specific model `nickprock/sentence-bert-base-italian-xxl-uncased`, tuned on Italian sentences. The model maps sentences and paragraphs to a 768 dimensional dense vector space, which can then be used for similarity, clustering and classification tasks.

Within the dataset, the embedding vector for each ad is contained in the `bert_embeddings` column.

5.3. Ad Similarity

To understand the relationships and patterns among the generated and real political ads, a similarity analysis was conducted using the corresponding embeddings. The study of those similarities is necessary firstly, in order to know just how similar advertisements are between each other, both in the real and generated case, and secondly, to find underlying clusters within each party's advertising strategy.

The two metrics in measuring similarities used in the work are Cosine Similarity and BERTScore.
Cosine Similarity measures the cosine of the angle between the vectors calculated as the dot product of the vectors divided by the product of their lengths, quantifying the similarity between the embeddings of different ads. A higher cosine similarity score indicates greater similarity between two ads. The scores were computed using the `cosine_similarity` function from the `sklearn` module.

BERTScore makes use of the deep contextual understanding from BERT-based to evaluate the similarity between texts. It computes a partial similarity score for each token in the candidate sentence with each token in the reference sentence, providing a measure of similarity that accounts for the context and meaning of the words in the sentences. The `evaluate-metric/bertscore` library, as referenced in [66], was utilized for this analysis.

For each ad class for each similarity metric an Heatmap had been generated to visualize and compare the similarity scores. These will intuitively lead to a view of how the ads within the same classes compare against one another, highlighting large clusters or outliers.
Additionally, both intra-party and inter-party similarities were analyzed to understand how ads within the same party compared to those from different parties. Intra-party similarity focused on the consistency of messaging within a single party, while inter-party similarity refers to the distance of one party compared to all the others. Combined, the two give insight into exactly how distinctive or common each party's advertising approach is and in the case of generated ads, whether the LLM is able to particularize the messages of the various parties or tends towards excessive uniformity.

5.3.1. Clustering Analysis

To further explore the structure of the datasets, for each of the aforementioned ad classes, it were tested two clustering techniques: t-SNE and K-Means.

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a dimensionality reduction technique that allows for the visualization of high-dimensional data within a two-dimensional space. The `TSNE` class from the `sklearn` module was used to project the ad embeddings and, thus, detect clusters that represent similar advertising messages.
K-Means clustering was employed to try to partition the ads into distinct clusters based on their embeddings. By specifying the number of clusters, the ads were categorized into groups that shared common characteristics. For the implementation, also in this case the `sklearn` module was used, through its `KMeans` class.

These clustering analyses, combined with the results of the similarity analysis, were aimed at assessing possible positive use by ML classifiers, and an estimate on the ad classes and tasks in which they could perform best.

5.4. Machine Learning Classifiers

As the last step of the research work, three different families of ML models were developed: Siamese Neural Networks, PartyClassifiers, and VsClassifiers. Each model was designed to address specific classification tasks using different subsets of the dataset.

5.4.1. Siamese Neural Network

The first model designed was a Siamese neural network, since it is an approach that has proven to be very suitable for tasks involving similarity and relationship prediction high-dimensional data representations [13]. In this case the goal is given a pair of ads, determine if they originate from the same party. The dataset used were: `real-ads`, `real-ads-cap`, `few-shots`, and `zero-shot`; a different SNN was trained for each dataset.

The architecture consisted of a custom three-layer Siamese network trained using contrastive loss, in which each ad in the pair – via its embedding representation – were passed through identical subnetworks that shared weights. The contrastive loss function aims at minimizing the difference between embeddings from ads of the same party by maximizing the embedding distance of ads from different parties. This training strategy aims to capture the intrinsic and underlying similarities and differences between ads based on their party affiliations. The network has a total of 238,016 parameters, and its structure is illustrated in Figure 5.

After training the models with cross-validation – implemented stratifying the data, to ensure that the proportion of each party label is maintained in every fold and training/test subsets – the performance of the models was then evaluated using metrics such as *F1-Score* and *Accuracy*, and it was also analyzed how the performance varied considering only the ads related to one party at a time.

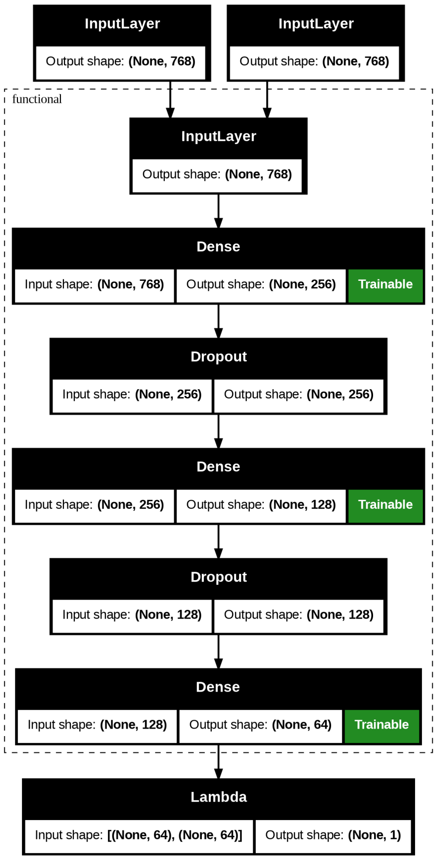

Figure 5: Siamese Neural Network architecture

5.4.2. PartyClassifiers

PartyClassifiers were designed to predict the political party associated with a given advertisement testing two type of classifier: `RandomForestClassifier` from `sklearn` [43] library and `XGBClassifier` from `xgboost` [12]. The dataset used were the same as for the SNN: `real-ads`, `real-ads-cap`, `few-shots`, and `zero-shot`; a different classifier was trained and hypertuned for each dataset.

The two classifiers, *RFClassifier* and *XGBClassifier* were hypertuned using GridSearchCV for `[n_estimators, max_depth]` and `[n_estimators, learning_rate]` respectively. Subsequently, they were trained using stratified cross-validation, to ensure performance optimization but always keeping in mind to avoid biased predictions or overfitting.

The model outputs its prediction, i.e. the name of the party, and, if desired, its percentage of sureness.
The performance of the models was then evaluated using metrics such as *F1-Score* and *Accuracy*, and it was also analyzed how the performance varied considering only the ads related to one party at a time.

5.4.3. VsClassifiers

VsClassifiers were designed for a distinct task, specifically to distinguish between real and generated advertisements. Similar to the PartyClassifiers, the two base models that has been employed are Random Forest and XGBoost Classifier.
In this scenario, the datasets used to train and fine-tune the classifiers differ, with each having two distinct classes:

- `real-vs-fs`: In order to maintain a perfect balanced dataset 1:1 Real/Gen and within each party subclass, for every party has been selected `min(num_real_ads, num_generated_ads)` real and few-shot generated ads, ultimately resulting in 940 total ads, 470 for each class.
- `real-vs-zs`: Same as before, this time due to sightly larger `num_generated_ads` we end up with 984 ads, 492 for each class.
- `real-vs-gen`: For this dataset, a different choice has been made, it has been selected all unique real ads available (635) and it has randomly selected 635 generated ads from few-shot and zero-shot combined. By doing so, a larger dataset of 1270 total ad was obtained.

This group of models underwent the same training and hyper-tuning process as described in the previous section.
The model outputs its prediction – this time binary: *Real/Generated* – and, if desired, its percentage of sureness.
The performances are evaluated using the same methods and metrics as the PartyClassifiers, as outlined in section 5.4.2.

6. Results and Discussion

This study aimed to evaluate the capability of large language models in generating persuasive political advertisements and to assess the effectiveness of if machine learning classifiers are capable in analyzing partisan origin and distinguishing between authentic and synthetic ads. The following section present the findings related to these objectives.
The results of this study are presented into two primary sections: similarity analysis and ad classification.

In the similarity analysis section, the relationships and patterns among both real and generated political ads are examined using Cosine Similarity and BERTScore metrics. Heatmap visualizations and clustering techniques, such as t-SNE and K-Means, are utilized to illustrate the extent of similarity within and between different ad classes and political parties. This analysis makes it possible to highlight similarity processes among ads both within the same party and among others and how these are reflected in the counterpart of generated ads.

The ad classification section evaluates the effectiveness of the developed machine learning models, namely Siamese Neural Networks, PartyClassifiers, and VsClassifiers. In this regard, each model's ability to accurately classify party affiliations and distinguish between real and generated ads has been evaluated with metrics such as Accuracy, F1-score, and AUC. The discussion highlights the strengths and limitations of each classifier in relation to practical applicability and further improvement.

6.1. Similarity Analysis

6.1.1. Real Ads

The first ad class examined was the Real-Ads category.
After calculating with cosine similarity and BERTScore the similarities between the ads, they were plotted in two heatmaps to visualize the results and give us an opportunity for analysis them, visible in Figure 6 and 7.

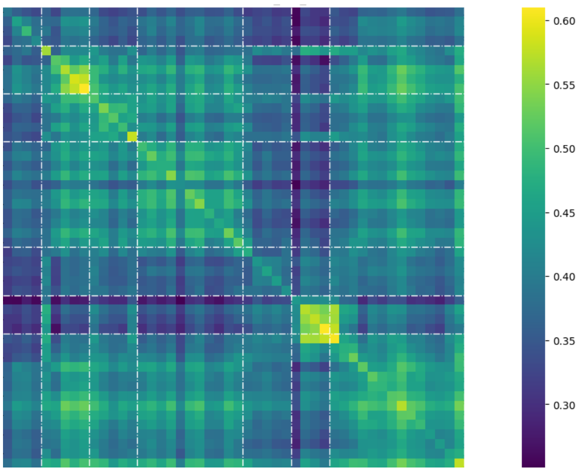

Figure 6: Real-Ads Similarity Heatmap (Cosine Similarity)

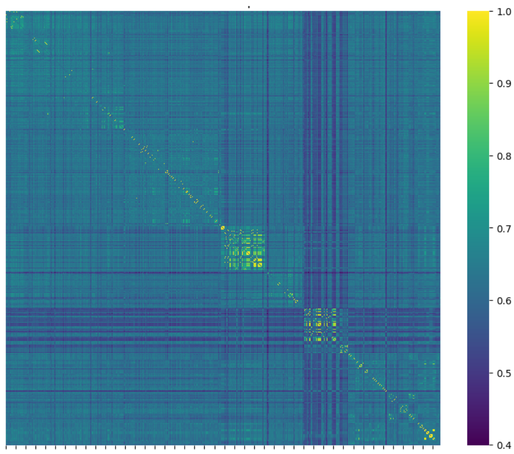

Figure 7: Real-Ads Similarity Heatmap (BERTScore F1)

Looking at the heatmaps, it is possible to discern areas of higher similarity along the diagonal, which indicate, predictably and hopefully, a slight increase in similarity between ads of the same party.

At this point the next step was to identify a way to analyze this behavior in more detail. To do this, the intra- and inter-similarity distributions for each party were derived. Shown in Figure 8, they confirm a slight tilt where the intra similarities – i.e. the similarities calculated between ads of the same party – are shifted to the right compared to the inter similarities – similarities calculated between ads of the party in relation to ads of other parties -. This behavior is more pronounced for some parties, such as the PD and M5S (above left and below left of the figure) and less so in others, such as the 3rd Pole (top center), signaling greater differentiation of their political ads.

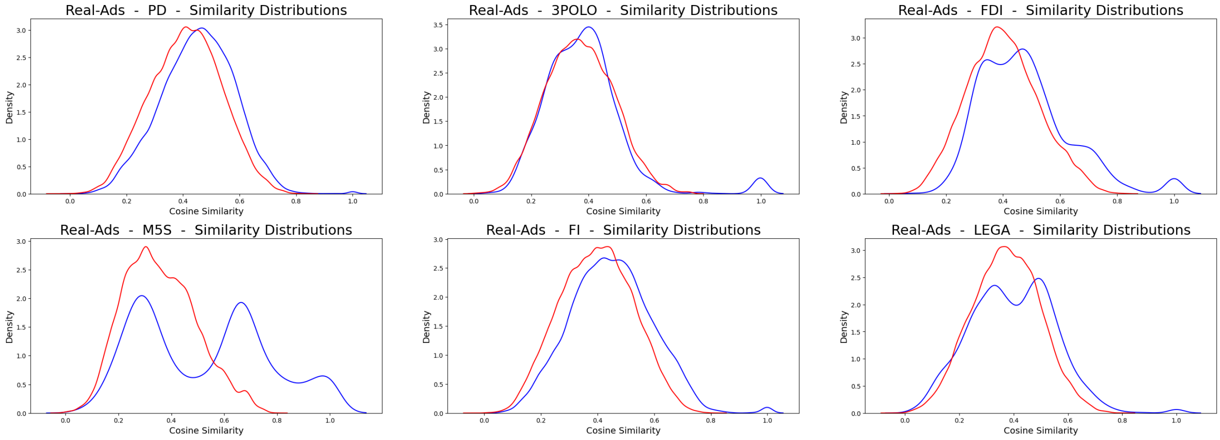

Figure 8: Real-Ads per-party intra-similarity(in blue) and inter-similarity(in red)

6.1.2. Few-Shot Generated Ads

Following the analysis of real advertisements, the next focus was on the generated ads produced by `Llama 3.1`. The similarity metrics, Cosine Similarity and BERTScore, were applied to assess how the LLM behaves, whether it follows the same dynamics as real ads, amplifying or attenuating them.

The similarity analysis was conducted across all three types of Few-Shot generated ads as described in Section 5.2.1. These types include:

[1] One generic and four topic-specified ads for each page (953)
[2] One generic and four topic-specified ads for each official pages and party leaders' pages (15)
[3] Multiple (100) generic ads for each party (7)

However, for the remainder of the research, only case [3] (Multiple generic ads for each party) was explored and utilized in the classification tasks. The results and heatmaps for cases [1] and [2] are available in Appendix II.

As before, the results were visualized using heatmaps, as shown in Figure 9 and Figure 10.

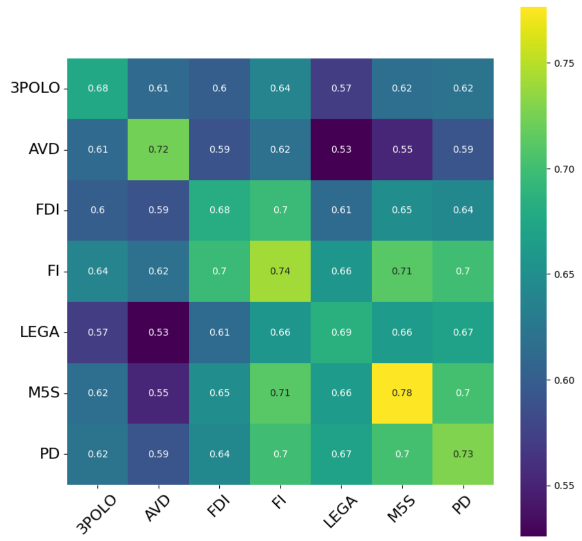

Figure 9: Few-Shot[3] Similarity Heatmap (Cosine Similarity)

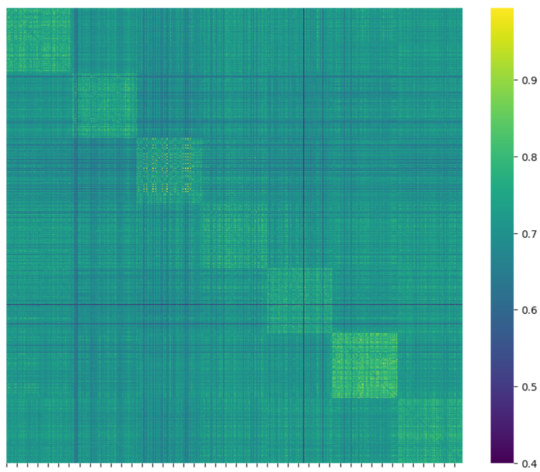

Figure 10: Few-Shot[3] Similarity Heatmap (BERTScore F1)

The heatmaps reveal distinct patterns in the similarity scores of generated ads. Similar to the real ads, higher similarity scores are observed along the diagonal, suggesting that generated ads within the same class exhibit greater likeness to each other. However, the overall similarity scores of the generated ads are generally higher than those of the real ones, indicating that although LLMs are able to produce consistent and contextually relevant messages, the way they work means that they are able to generalize less and tend to focus more on certain themes in the way these ads mimic the real ones.

To elaborate on these observations, intra-party and inter-party similarity distributions were calculated as for the precursor class, but this time two additional noteworthy data were added to the plots: the intra-similarity of the actual party ads and the similarity between the ads generated and the examples used in the prompt – calculated as the average of the similarities between the generated ad and each of "its" examples -, shown in green and yellow in Figure 11, respectively.

The results obtained confirm what the heatmap anticipated, following the same trend, and indeed amplifying it, as the predecessor case with Real-Ads. In this case, all parties have a rightward shift by taking intra-similarity into account.
Interestingly, taking a party, the ads tend to be more similar to each other than to the ads used as few-shot examples to generate them. This points to an ability of LLMs to abstract from the mere examples provided and create a message based on the predicted ideology of the party in question.

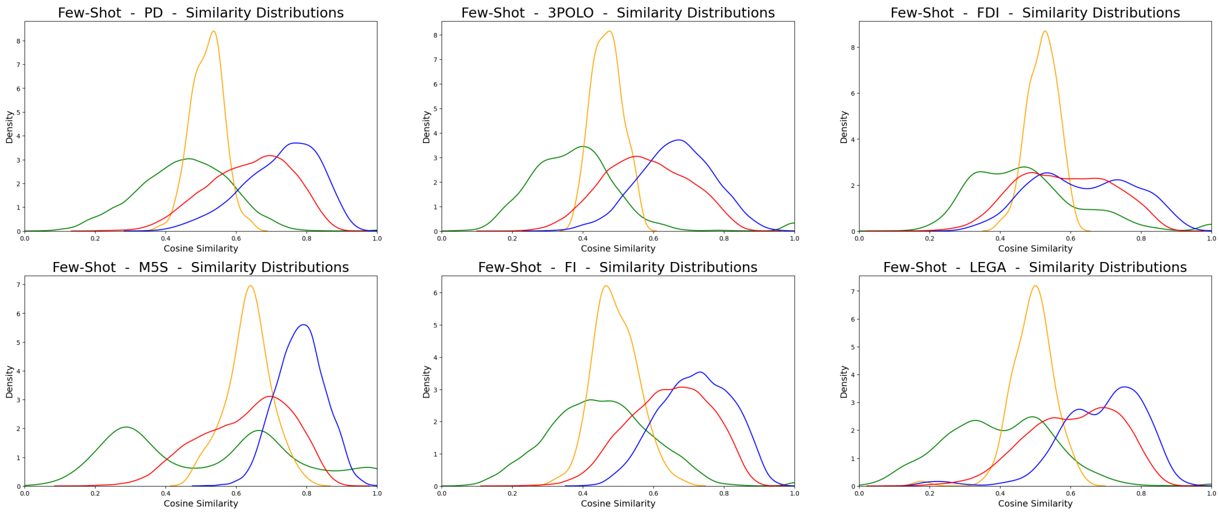

Figure 11: Few-Shot[3] per-party intra-similarity(in blue), inter-similarity(in red), intra-similarity of the real party ads(in green), similarity between the ads generated and the examples used in the prompt(in yellow)

6.1.3. Zero-Shot Generated Ads

To finish the similarity analysis, we now turn the attention to the last class of Ads, the ads generated via Zero-Shot prompt. Once more, the first step was to calculate the similarity of the ads using the same metrics: Cosine Similarity and BERTScore. The results were visualized using heatmaps, illustrated in Figures 12 and 13.

Figure 12: Zero-Shot Similarity Heatmap (Cosine Similarity)

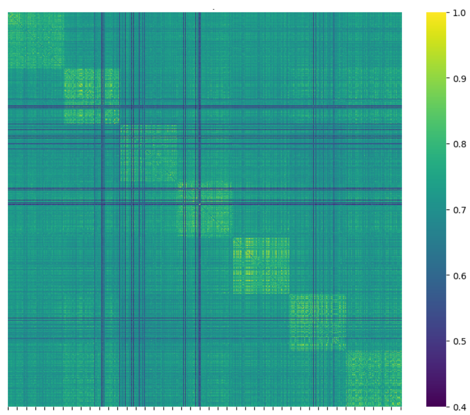

Figure 13: Zero-Shot Similarity Heatmap (BERTScore F1)

The heatmaps for Zero-Shot generated ads show uniformly higher similarity scores across all classes. This indicates that the LLM produces ads that are more alike to one another, regardless of the political party. However, the distinction between intra-party and inter-party similarities remains clear and is even more pronounced than in the Few-Shot scenario.

Further examination through intra-party and inter-party similarity distributions, as shown in Figure 14, confirms that both types of similarities – for all parties – are elevated in the Zero-Shot case. The higher intra-similarity scores indicate that ads within the same party are more consistent, while the increased inter-similarity scores reveal greater overlap in themes and language across different parties. This behavior suggests that the LLM is less effective in differentiating ads regardless of their party affiliation, generating more generic and similar ads. Despite this, the clear gap between intra-party and inter-party similarities still allows, perhaps even more easily than previous cases, ads to be effectively grouped by party, despite an overall shift toward less diversity in messaging.

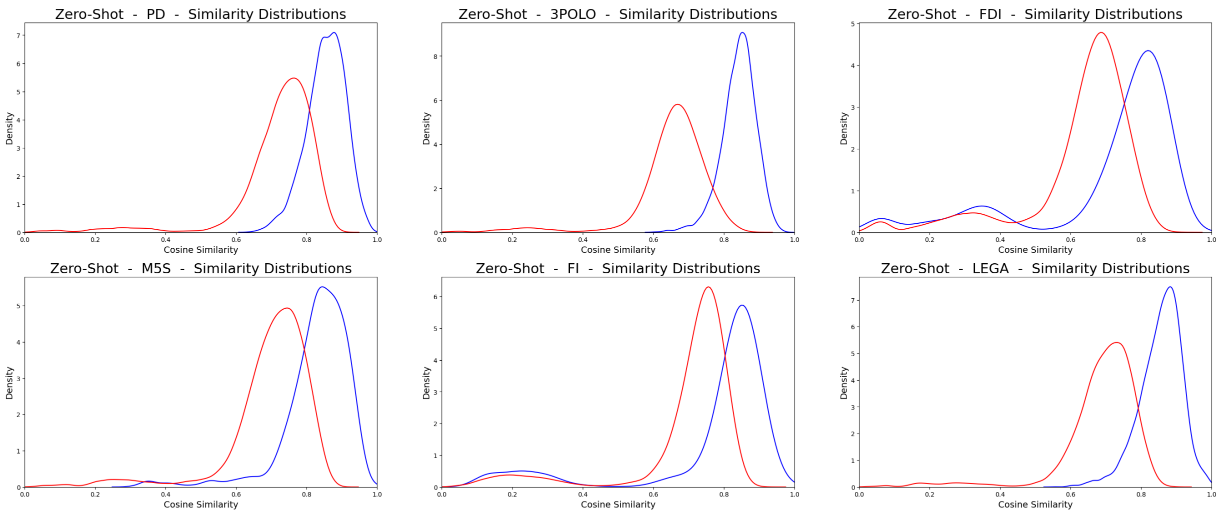

Figure 14: Zero-Shot per-party intra-similarity(in blue) and inter-similarity(in red)

To summarize, the ads generated by Zero-Shot show a trend toward greater uniformity, and this probably highlights a limitation in the LLM's ability to produce diverse and party-specific content without few-shot examples or specific prompts. This result again underscores the importance of properly structuring prompting when we use large language models, in this case having the goal of achieving a comprehensive and differentiated strategic political campaign.

6.1.4. Clustering

Prior to the implementation of classifiers, it was advantageous to conduct preliminary evaluations using prominent clustering techniques, such as t-SNE and K-Means, across the various classes of advertisements.

The clustering results emerged in line with the findings from the similarity analysis section. In fact, generated ads, both Few-Shot and Zero-Shot showed a greater inclination to form distinct and easily identifiable clusters compared to the Real Ads. This enhanced clustering performance for generated ads is evident in Figures 15, 16, 17, and 18, where the generated ads form tighter and more separable groups.

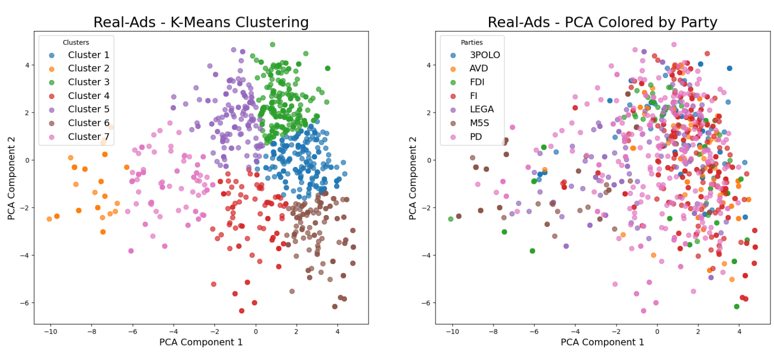

Figure 15: Real-Ads Clustering (K-Means with PCA)

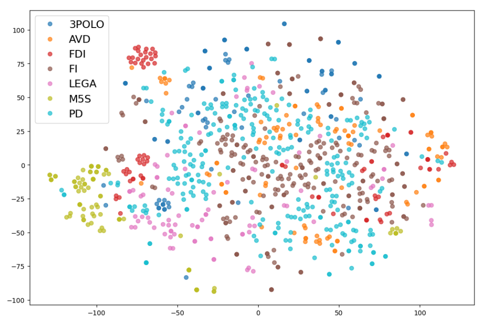

Figure 16: Real-Ads Clustering (t-SNE)

Real-Ads clustering, as shown in Figures 15 and 16, reveals a more dispersed arrangement, indicating greater diversity and less distinct grouping within real political advertisements.
In contrast, Zero-Shot generated ads, depicted in Figures 17 and 18, demonstrate a higher degree of clustering tightness and separation between different political parties. Few-Shot generated ads also show improved clustering compared to Real Ads, even if not pronounced as the Zero-Shot case, results are available in Appendix II.

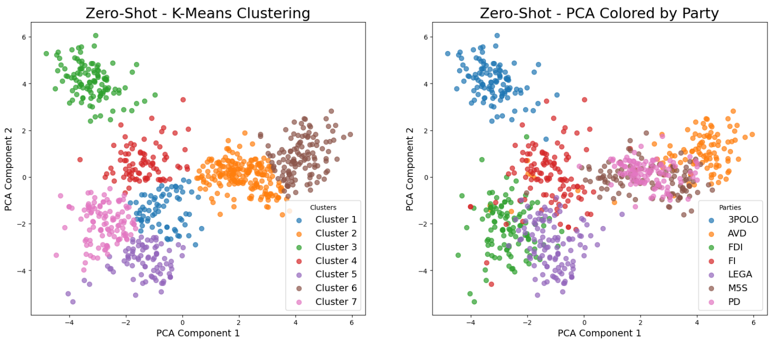

Figure 17: Zero-Shot Clustering (K-Means with PCA)

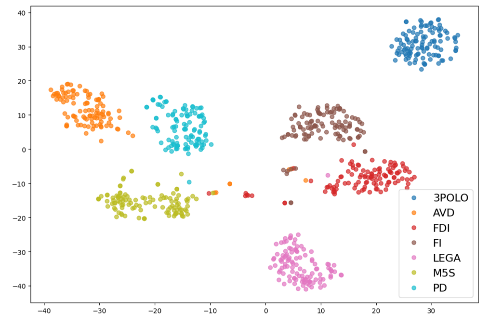

Figure 18: Zero-Shot Clustering (t-SNE)

6.1.5. Findings & Remarks

The similarity and clustering analyses demonstrated that synthetic ads generated through Few-Shot prompting closely align with their respective political parties, exhibiting high internal similarity and distinct inter-party differences, while Zero-Shot ads tend to be more generic yet still maintain clear party-based clustering, highlighting the LLMs' capability to generate targeted and coherent political messaging.
Building on these findings, the subsequent classification performance evaluation examines how effectively machine learning models can categorize these advertisements based on their content and origin.

6.2. Ad Classification

This section presents the classification performance of the developed machine learning models: Siamese Neural Networks, PartyClassifiers, and VsClassifiers. Each model was evaluated on both the entire dataset and the subclasses by considering only the ads of a corresponding party using metrics such as Accuracy, F1-score, and AUC. The results are discussed in detail below, accompanied by relevant figures.

Before we begin comparing the various models, it is important to analyze the structure of our real ads dataset. As exhibited in Table 3, the number of unique real ads available for the various parties is very unbalanced. This brought the need and interest to also test a dataset in which ads are capped at a certain threshold per party to ensure more equitable and less disadvantageous performance for underrepresented parties. This version of the dataset will be called `real-ads-cap`.

Party	Number of Ads
3POLO	54
AVD	65
FDI	47
FI	144
LEGA	71
M5S	33
PD	221
Total Ads	**635**

Table 3: Number of unique ads per party

It was observed that models trained on the Zero-Shot dataset obtained really high performances when tested on same class zero-shot ads but showed reduced generalization capabilities when tested on other unseen data, for example Few-Shot ads. This indicates a limitation in their ability to generalize predictions across different types of generated advertisements and by virtue of that, it was decided not to dwell on these models in this Section. Results for models trained on the Zero-Shot dataset are available in Appendix III.

6.2.1. Siamese Neural Network

The first family of ML models that has been developed and tested is Siamese Neural Network (SNN).
As described in depth in Section 5.4.1, the goal of these Neural Networks is given a pair of ads, determine if they originate from the same party. The performance has been evaluated using F1-Score and Accuracy as metrics.

The Siamese Neural Network demonstrated decent performance on the Real-Ads dataset. Due to the imbalance in the dataset, when trained on initial real-ads dataset certain subclasses exhibited poor results. However, performance significantly balanced and improved for underrepresented parties when trained using the `real-ads-cap` version. The results are shown in the two side-by-side plots in Figure 19.

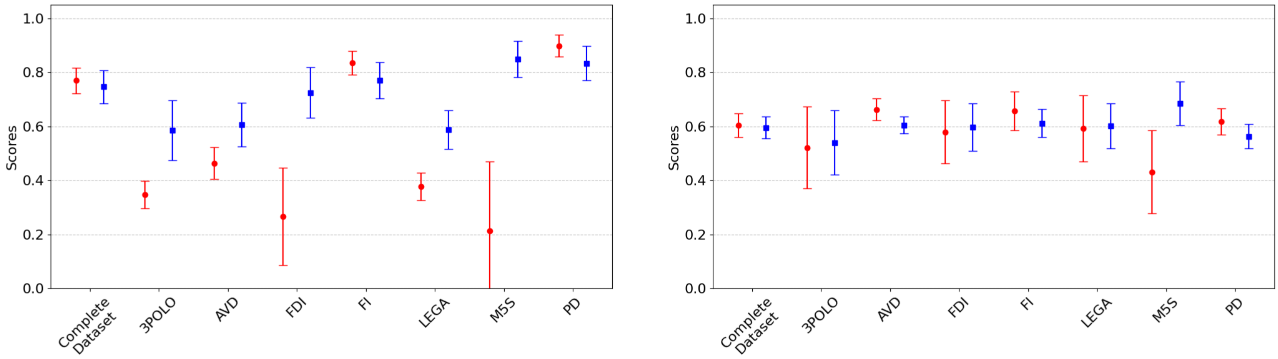

Figure 19: Siamese NN Real-Ads (All and Cap) - F1-score (red) and Accuracy (blue) per party

When trained on generated ads, the SNN achieved really good F1-scores and Accuracy across all parties, as illustrated in Figure 20 for the Few-Shot ads. This indicates that the SNN is highly capable of accurately identifying and classifying LLM generated and, in counterinstance, to again certify the ability of LLMs – Llama 3.1 in this case – to produce party-differentiated political ads.

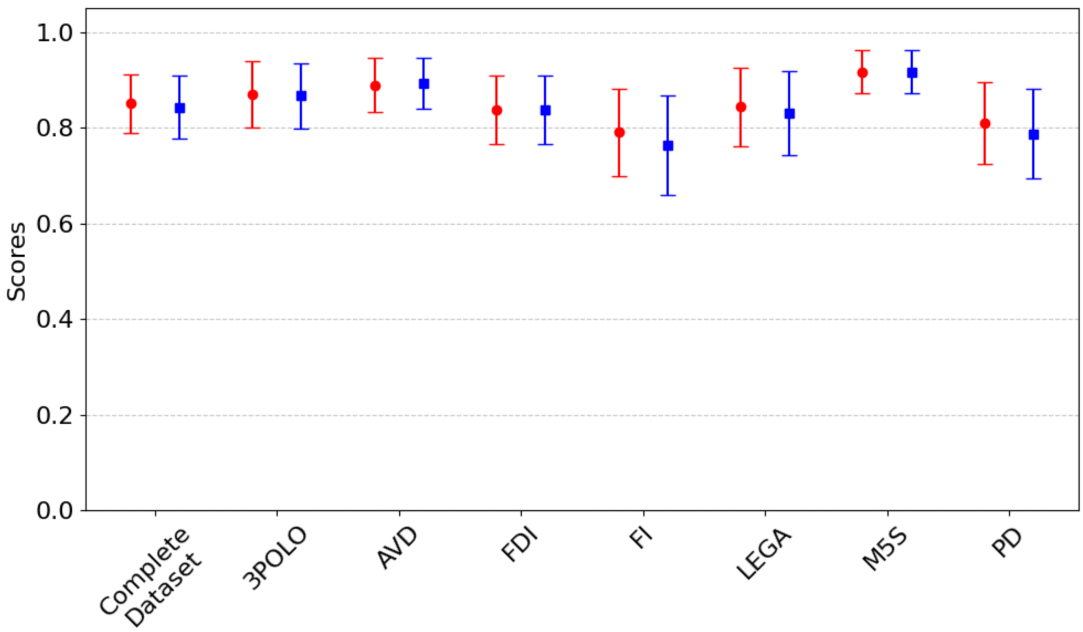

Figure 20: Siamese NN Few-Shot - F1-score (red) and Accuracy (blue) per party

6.2.2. PartyClassifiers

The second family of ML models that has been designed is PartyClassifiers, explained in Section 5.4.2.
With a similar task to the abovementioned, they were employed to predict the political party associated with a given advertisement utilizing both Random Forest (RFC) and XGBoost (XGB) classifiers. The performance of these two classifiers has been evaluated using F1-Score as metric.

As previously, they demonstrated decent performing when trained on Real-Ads dataset in his *complete* version, especially XGBoost-based version. When trained with the *cap* version, performance slightly improved and balanced across the different parties, improving the results especially for the models that had Random Forest as a base, as illustrated in Figure 21.

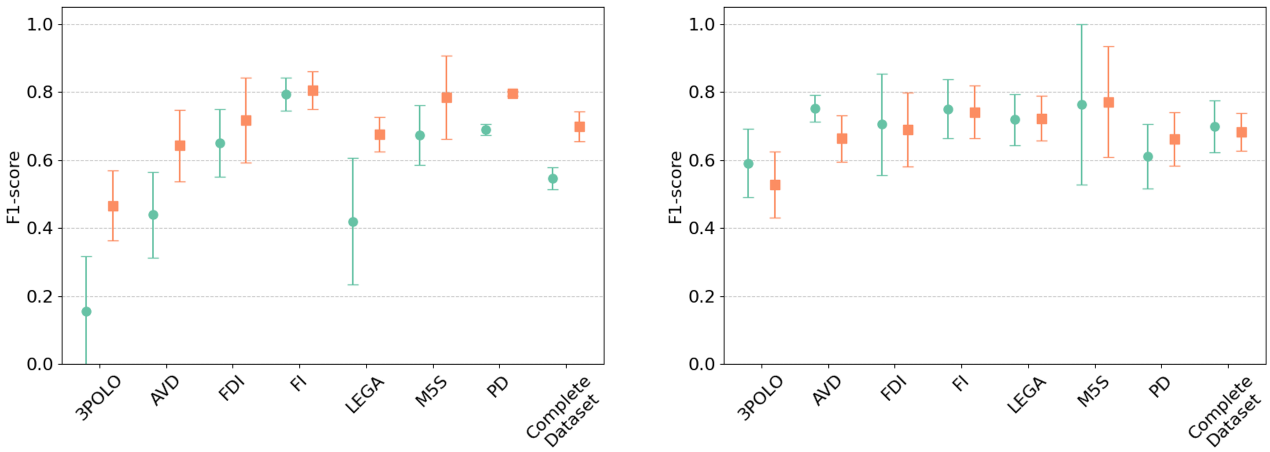

Figure 21: PartyClassifier Real-Ads (All and Cap) - F1-score per party
(Green=RandomForest, Orange=XGBoost)

Regarding LLM generated ads – as visible in Figure 22 for few-shot case – both classifiers exhibit high F1-scores across all parties. Relative to party subclasses, some achieve higher results than others, likely given by greater distinctiveness of language in the campaigns of some parties, to better understand this analysis we can take as an example the difference in terms of political dictionary between the *5 Star Movement* and *Forza Italia*.

The Random Forest and XGBoost classifiers showed really comparable performance, indicating as ambivalent the final choice on the use of one of them. Detailed ROC Curves and AUC metrics for these models are available in Appendix III, providing further insights into their classification capabilities.

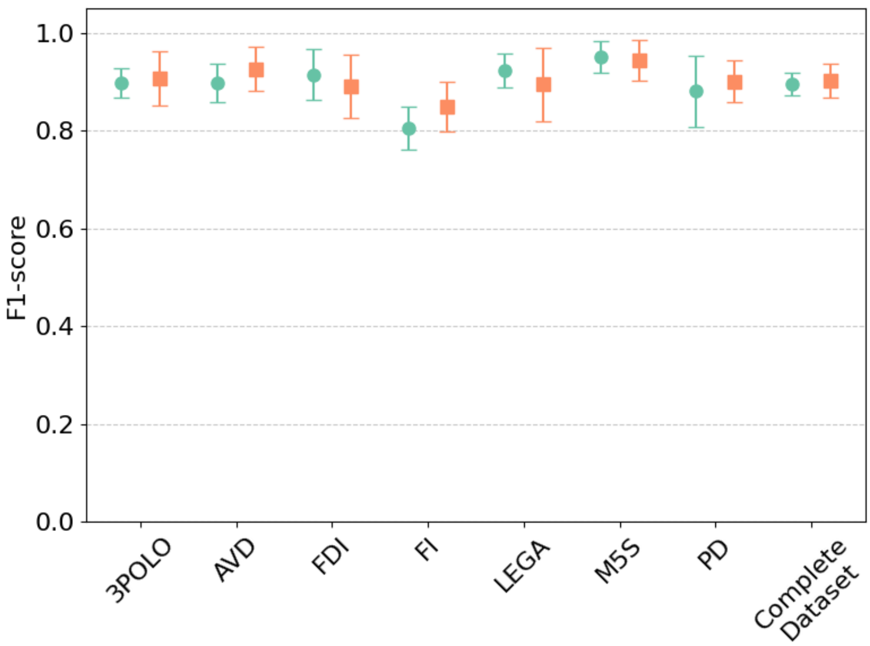

Figure 22: PartyClassifier Few-Shot - F1-score per party
(Green=RandomForest, Orange=XGBoost)

6.2.3. VsClassifiers

As the last stop on our journey among ML classifiers, it is time to assess the results of the VsClassifier. They were previously described, along with their different training data, in Section 5.4.3.
Their task is different than the others model families, in fact VsClassifiers were designed to distinguish between real and generated advertisements. As for PartyClassifiers, Random Forest and XGBoost classifiers were used and their performance has been evaluated using F1-Score as metric.

The results, presented in Figures 23 and 25, indicate that the classifiers performed exceptionally well across all scenarios, achieving high F1-scores in distinguishing between real and generated ads. In addition, analysis of the party subclasses shows that the models behave consistently, demonstrating their ability to extract patterns that are valid across all parties rather than being confined to specific topics or language styles associated with only some of them.

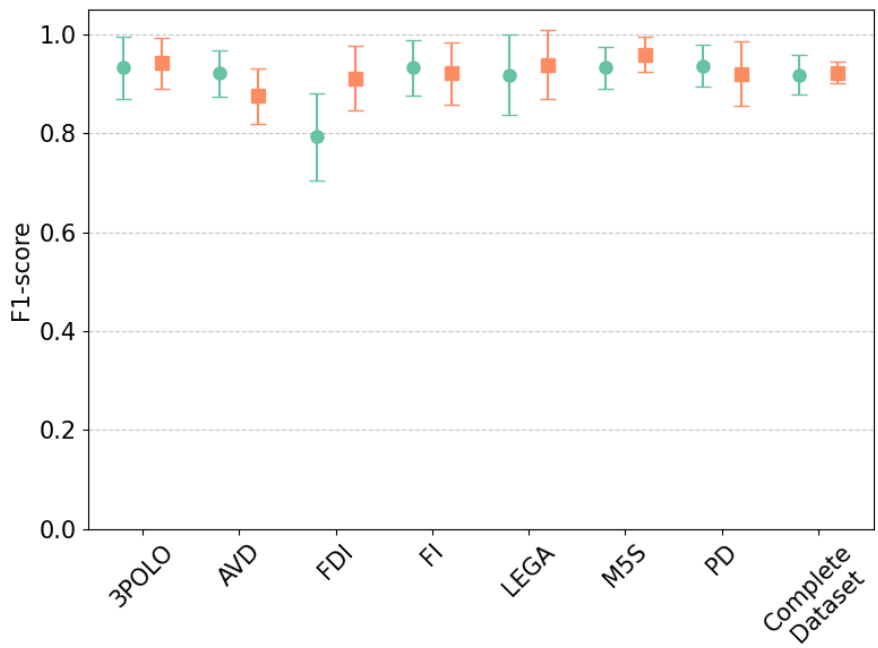

Figure 23: VsClassifier Real-Ads vs Few-Shot - F1-score per party (Green=RandomForest, Orange=XGBoost)

Again, the performance of the Random Forest and XGBoost classifiers is comparable, perhaps with a slight preference toward XGBoost, but no clear winner emerges. This confirms how correct the approach used was for this task, rather than the use or fine-tuning of a specific ML model.

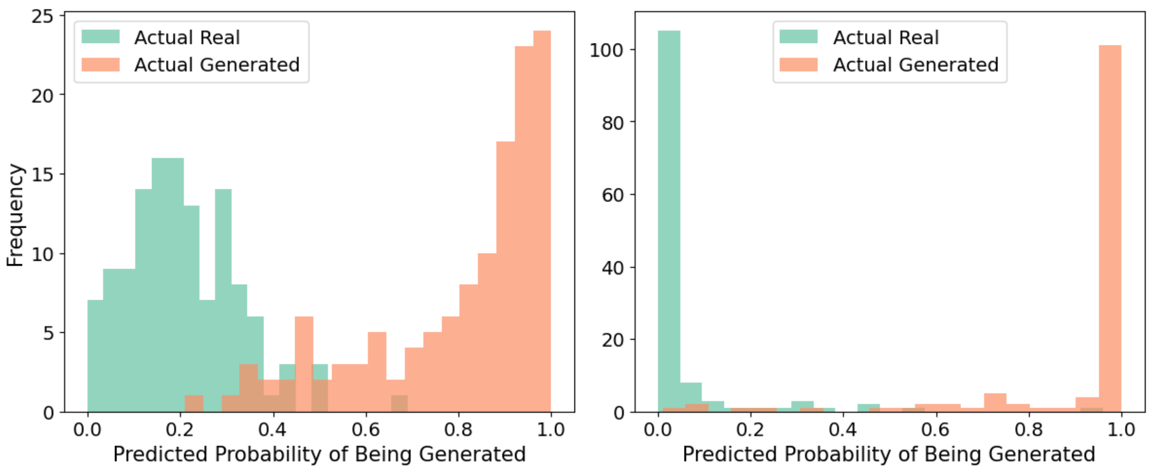

Figure 24: VsClassifier Real-Ads vs Generated-Ads - Probability Distribution (left=RandomForest, right=XGBoost)

To better understand the internal functioning and reasoning of the VsClassifiers, Figure 24 presents the probability distributions for real and generated ads as predicted by the Random Forest and XGBoost classifiers. The plot highlights key differences in how the two models operate: the Random Forest classifier exhibits a reliable distribution for real ads but tends to produce more continuous probability values, resulting in less binary final predictions. In contrast, the XGBoost classifier generates probability scores that are predominantly close to zero or one, reflecting a more decisive and binary prediction approach.

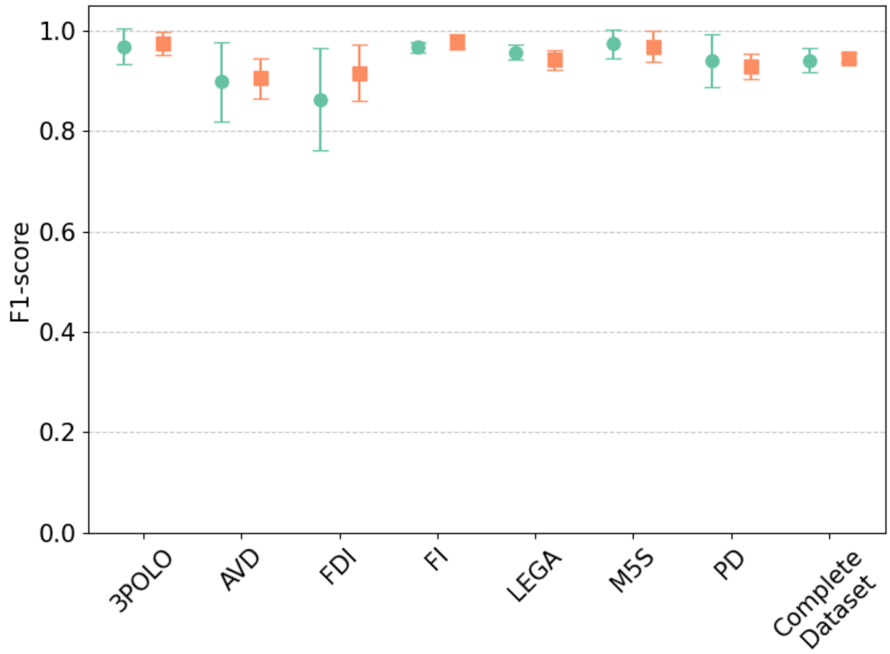

Figure 25: VsClassifier Real-Ads vs Generated-Ads - F1-score per party (Green=RandomForest, Orange=XGBoost)

6.2.4. Human Evaluation

To further validate the performance of the machine learning classifiers, a simple test experiment involving human evaluation was conducted. In this experiment, fifty advertisements were randomly selected, comprising 25 real ads and 25 synthetic ads generated through Few-Shot prompting. Participants were provided with the text of each ad along with the affiliated political party and were tasked with determining whether the ad was real or AI-generated.
The results are summarized in Table 4, comparing the Mean F1-Score and Standard Deviation of the Random Forest Classifier, XGBoost Classifier, and human evaluators.

Model/Human	Mean F1-Score	Std F1-Score
RandomForestClassifier	0.91	0.04
XGBClassifier	0.92	0.02
Human	0.45	0.16

Table 4: Comparison of Machine Learning Models and Human Evaluation

As evident from the results, human judgment emerges as little more skillful than flipping a coin. This stark contrast underscores two key concepts: the reliability and efficiency of machine learning models in classification tasks compared to human judgment, and the difficulty of the human in determining whether an ad is AI-Generated, confirming the effectiveness of using LLM in the policy arena.

6.2.5. Findings & Remarks

The classification experiments demonstrated that machine learning classifiers, including Siamese Neural Networks, PartyClassifiers, and VsClassifiers, achieved high accuracy and F1-scores in predicting political party affiliations – particularly among synthetic ad – and distinguishing between real and LLM-generated advertisements, significantly outperforming human evaluators and highlighting the robustness and effectiveness of ML models in analyzing political advertising.

The comprehensive analysis of similarity and classification performance sets the foundation for the concluding discussions and recommendations presented in the final chapter. These conclusions not only summarize the key findings but also explore the broader implications of using large language models in political campaign strategies.

7. Conclusion

This thesis has comprehensively examined the meeting points between large language models (LLMs) and political communication, with a specific focus on their application in generating and analyzing political advertisements using the 2022 Italian general election as a case study. The primary objective was to assess the capability of LLMs, particularly Meta's `Llama 3.1`, in producing persuasive and party-specific advertisements that mirror the strategies employed in real-world campaigns. This involved the generation of synthetic ads that seek to reflect the ideological stances and communication styles of the various Italian political parties through various prompting strategies, with the aim of creating a solid dataset for comparative analysis.

Through the generation of artificial ads, the study demonstrated that LLMs can effectively replicate the persuasive elements and thematic nuances of authentic political advertisements. The similarity analysis, utilizing Cosine Similarity and BERTScore metrics, revealed that while generated ads exhibit high internal similarity within each party, they also maintain distinctiveness across different political entities. Clustering techniques such as t-SNE and K-Means further underscored the intrinsic grouping of ads by political party, particularly in synthetic dataset. These findings indicate that LLMs have the deep understanding required to craft targeted and ideologically aligned content, closely reproducing the strategic communication observed in actual political campaigns.

The development and evaluation of machine learning classifiers – namely Siamese Neural Networks, PartyClassifiers, and VsClassifiers – further validated the effectiveness of the generated ads. The classifiers achieved high accuracy in predicting party affiliations, confirming that LLMs possess the capability to generate ads that are precisely targeted and tailored for specific political parties. Secondly, the classifiers demonstrated a significantly higher ability to distinguish between real and generated advertisements compared to human evaluators. This result points out the possibility of applying these machine learning techniques in order to automate the analysis and verification of political content, hence making their role crucial in monitoring and safeguarding the integrity of political milieu, particularly in identifying AI-generated disinformation or attempts to manipulate public opinion.

These results bring to light that the use of LLM in the political scenario – including the Italian one – is already realistic, and it is able to both augment and challenge traditional political communication strategies.
Considering the truthfulness that synthetic ads and the seemingly endless potential for personalization and targeting that Large Language Models offer, we are in front of a radical change in the way we have been analyzing political campaigns so far. De facto, this ensures that anyone, with meager expense and required know-how, can create a professional and widely diversified political advertising strategy.

7.1. Limitations and Future Work

This work provides valuable insights into the application of LLMs in political advertising, nevertheless, several limitations must be acknowledged. First, the analysis was confined to Facebook advertisements, potentially overlooking strategies employed on other social media platforms. Second, the synthetic ads were generated based on existing datasets, which may limit the diversity of the generated content. Additionally, no large-scale, rigorous experiments were conducted with human participants to assess the persuasiveness of generated ads, which is essential for validating their real-world impact.

Given these limitations, future research can build upon the foundations laid by this study in several key areas. First, following what was suggested in studies [33] and [58], it would be valuable to directly test Large Language Models on classification tasks including both party classification and real/generated binary classification. This could enhance our understanding of their capabilities and limitations in complex political contexts.
Additionally, expanding the scope to include real-world experiments involving an appropriate number of human participants would provide better empirical validation of the models' effectiveness. Conducting experiments that directly compare real versus generated Italian ads with human evaluators would offer deeper insights into the truthfulness and authenticity of AI-generated content. Furthermore, evaluate the persuasiveness of these ads through controlled experiments would help quantify the impact of LLMs on voter behavior also in the peculiar Italian political landscape, helping bridging the gap between theoretical capabilities and practical applications.

7.2. Ethical Considerations

With great power comes great responsibility. Historically, the correlation between intelligence, education, and the propensity to cause catastrophic harm has served as a natural safeguard for humanity. Smart and educated individuals, who possess the knowledge and skills to effect significant change, rarely engage in actions that could lead to widespread devastation. This *natural safeguard* has protected societies for centuries by ensuring that those with the capability to cause harm are also those most likely to act responsibly.

However, the advent of AI and, more specifically, Large Language Models disrupts this longstanding correlation. Dario Amodei, the CEO of *Anthropic* – the company behind *Claude AI* and arguably one of the most influential AI companies globally – highlights in his recent interview with Lex Fridman [21, 22] that AI possesses the potential to democratize access to powerful capabilities that were once restricted to a select group of highly educated and responsible individuals. This shift means that dangerous technologies and persuasive tools are now accessible to a much larger audience, including those who may not wish to share the ethical standards. Consequently, AI can be exploited to provoke large-scale malicious behavior, particularly within sensitive domains such as political advertising.

The ability of LLMs to produce highly convincing and targeted content raises significant ethical concerns. The misuse of these models in creating disinformation or evil-intentioned applications can undermine our democratic processes, erode public trust and further exacerbate political polarization. As already demonstrated in [62], LLMs can generate disinformation content that is often indistinguishable from human-written one. This underscores the urgent need for responsible AI development, regulatory frameworks, and ethical guidelines to mitigate the risks associated with AI-driven political campaigns.

7.3. Code and Data Availability

To facilitate transparency and reproducibility, all data, code, and prompts utilized in this research are made publicly available on GitHub at `https://github.com/menca-lsn/llm-it-elections-2022`.

References

[1] M. Aggarwal, J. Allen, A. Coppock, D. Frankowski, S. Messing, K. Zhang, J. Barnes, A. Beasley, H. Hantman, and S. Zheng. A 2 million-person, campaign-wide field experiment shows how digital advertising affects voter turnout. *Nature Human Behaviour*, 7(3):332–341, Mar 2023.

[2] I. Ajzen. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211, 1991.

[3] H. Allcott and M. Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36, May 2017.

[4] S. Balietti, L. Getoor, D. G. Goldstein, and D. J. Watts. Reducing opinion polarization: Effects of exposure to similar people with differing political views, Dec. 2021.

[5] D. Banerjee, P. Singh, A. Avadhanam, and S. Srivastava. Benchmarking LLM powered Chatbots: Methods and Metrics, Aug. 2023. arXiv:2308.04624 [cs].

[6] S. M. Breum, D. V. Egdal, V. G. Mortensen, A. G. Møller, and L. M. Aiello. The persuasive power of large language models. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, pages 152–163, 2024.

[7] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language models are few-shot learners, 2020.

[8] M. Burtell and T. Woodside. Artificial influence: An analysis of ai-driven persuasion, 2023.

[9] D. Bär, F. Pierri, G. De Francisci Morales, and S. Feuerriegel. Systematic discrepancies in the delivery of political ads on Facebook and Instagram. *PNAS Nexus*, 3(7):pgae247, 06 2024.

[10] S. Chaiken. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(5):752–766, 1980.

[11] K. Chen, Z. He, J. Yan, T. Shi, and K. Lerman. How susceptible are large language models to ideological manipulation?, 2024.

[12] T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM.

[13] D. Chicco. *Siamese Neural Networks: An Overview*, pages 73–94. Springer US, New York, NY, 2021.

[14] X. Chu, R. Vliegenthart, L. Otto, S. Lecheler, C. de Vreese, and S. Kruikemeier. Do online ads sway voters? understanding the persuasiveness of online political ads. *Political Communication*, 41(2):290–314, 2024.

[15] J. Clusmann, F. R. Kolbinger, H. S. Muti, Z. I. Carrero, J.-N. Eckardt, N. G. Laleh, C. M. L. Löffler, S.-C. Schwarzkopf, M. Unger, G. P. Veldhuizen, S. J. Wagner, and J. N. Kather. The future landscape of large language models in medicine. *Communications Medicine*, 3(1):1–8, Oct. 2023.

[16] M. Dainton. Explaining theories of persuasion.

[17] A. Deshmukh and A. Raut. Enhanced Resume Screening for Smart Hiring Using Sentence-Bidirectional Encoder Representations from Transformers (S-BERT). *International Journal of Advanced Computer Science and Applications (IJACSA)*, 15(8), July 2024.

[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[19] H. Ding, Y. Li, J. Wang, and H. Chen. Large Language Model Agent in Financial Trading: A Survey, July 2024. arXiv:2408.06361 [cs, q-fin] version: 1.

[20] E. Murgolo. Quanto conta negli USA la mobilitazione degli elettori?, 2024. Last accessed 2024-08-18.

[21] L. Fridman. Dario amodei: Anthropic CEO on claude, AGI & the future of AI & humanity | lex fridman podcast #452.

[22] L. Fridman. Transcript for Dario Amodei: Anthropic CEO on Claude, AGI & the Future of AI & Humanity | Lex Fridman Podcast #452, Nov. 2024.

[23] K. M. Goldstein and T. N. Ridout. The politics of participation: Mobilization and turnout over time. *Political Behavior*, 24(1):3–29, 2002.

[24] K. Hackenburg and H. Margetts. Evaluating the persuasive influence of political microtargeting with large language models. *Proceedings of the National Academy of Sciences*, 121(24):e2403116121, 2024.

[25] K. Hackenburg, B. M. Tappin, P. Röttger, S. Hale, J. Bright, and H. Margetts. Evidence of a log scaling law for political persuasion with large language models. *arXiv preprint arXiv:2406.14508*, 2024.

[26] K. Haenschen. The conditional effects of microtargeted facebook advertisements on voter turnout. *Political Behavior*, 45(4):1661–1681, Dec 2023.

[27] M. Hanna and O. Bojar. A fine-grained analysis of BERTScore. In L. Barrault, O. Bojar, F. Bougares, R. Chatterjee, M. R. Costa-jussa, C. Federmann, M. Fishel, A. Fraser, M. Freitag, Y. Graham, R. Grundkiewicz, P. Guzman, B. Haddow, M. Huck, A. J. Yepes, P. Koehn, T. Kocmi, A. Martins, M. Morishita, and C. Monz, editors, *Proceedings of the Sixth Conference on Machine Translation*, pages 507–517, Online, Nov. 2021. Association for Computational Linguistics.

[28] Hugging Face. Llama 3.1 - 405B, 70B & 8B with multilinguality and long context.

[29] J. Isaak and M. J. Hanna. User data privacy: Facebook, cambridge analytica, and privacy protection. *Computer*, 51(8):56–59, 2018.

[30] J. Jindal, S. Bedi, A. Swaminathan, M. Wornow, J. Fries, A. Chaurasia, and N. Shah. Large Language Models in Healthcare: Are We There Yet?, May 2024.

[31] D. Jurafsky and J. H. Martin. Speech and Language Processing.

[32] M. Linegar, R. Kocielnik, and R. M. Alvarez. Large language models and political science. *Frontiers in Political Science*, 5, 2023.

[33] M. Liu and G. Shi. Poliprompt: A high-performance cost-effective llm-based text classification framework for political science, 2024.

[34] Z. Liu, Y. Li, P. Shu, A. Zhong, L. Yang, C. Ju, Z. Wu, C. Ma, J. Luo, C. Chen, S. Kim, J. Hu, H. Dai, L. Zhao, D. Zhu, J. Liu, W. Liu, D. Shen, T. Liu, Q. Li, and X. Li. Radiology-llama2: Best-in-class large language model for radiology, 2023.

[35] S. C. Matz, J. D. Teeny, S. S. Vaid, H. Peters, G. M. Harari, and M. Cerf. The potential of generative ai for personalized persuasion at scale. *Scientific Reports*, 14(1):4692, Feb 2024.

[36] Meta. Ad Targeting dataset.

[37] Meta. Llama 3.1 - model cards & prompt formats.

[38] Ministero dell'Interno. Elezioni politiche 2022. `https://www.interno.gov.it/it/speciali/elezioni-politiche-2022`.

[39] B. Nguyen. Large Language Models in Financial Services | KMS Solutions, July 2024.

[40] OpenAI. OpenAI's GPT-4o model.

[41] E. Panel®. Council Post: Successful Real-World Use Cases For LLMs (And Lessons They Teach).

[42] G. Park and J. Kim. Improving the Preformance of Judicial Precedent Search by Fine-Tuning S-BERT. In J. S. Park, L. T. Yang, Y. Pan, and J. H. Park, editors, *Advances in Computer Science and Ubiquitous Computing*, pages 291–298, Singapore, 2023. Springer Nature.

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[44] R. E. Petty and J. T. Cacioppo. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. Springer-Verlag, 1986.

[45] F. Pierri. GitHub - frapierri/ita-election-2022: A repository of social media posts related to the Italian 2022 general election. — github.com. `https://github.com/frapierri/ita-election-2022`. [Accessed 08-11-2024].

[46] F. Pierri. Political advertisement on facebook and instagram in the run up to 2022 italian general election. In *Proceedings of the 15th ACM Web Science Conference 2023*, pages 13–22, 2023.

[47] 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, Oct. 2023. Association for Computing Machinery.

[48] O. Politici. I social media nella comunicazione politica italiana, May 2020.

[49] N. Reimers. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.

[50] M. Sherif and C. I. Hovland. *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*. Yale University Press, 1961.

[51] A. Simchon, M. Edwards, and S. Lewandowsky. The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS Nexus*, 3(2):pgae035, 01 2024.

[52] K. Singhal, T. Tu, J. Gottweis, R. Sayres, E. Wulczyn, L. Hou, K. Clark, S. Pfohl, H. Cole-Lewis, D. Neal, M. Schaekermann, A. Wang, M. Amin, S. Lachgar, P. Mansfield, S. Prakash, B. Green, E. Dominowska, B. A. y Arcas, N. Tomasev, Y. Liu, R. Wong, C. Semturs, S. S. Mahdavi, J. Barral, D. Webster, G. S. Corrado, Y. Matias, S. Azizi, A. Karthikesalingam, and V. Natarajan. Towards expert-level medical question answering with large language models, 2023.

[53] N. Stiennon, L. Ouyang, J. Wu, D. M. Ziegler, R. Lowe, C. Voss, A. Radford, D. Amodei, and P. Christiano. Learning to summarize from human feedback, 2022.

[54] D. Talbot. How obama really did it. *MIT Technology Review*, 111(5):78–83, 2008.

[55] B. M. Tappin, C. Wittenberg, L. B. Hewitt, A. J. Berinsky, and D. G. Rand. Quantifying the potential persuasive returns to political microtargeting, June 2023.

[56] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, D. Bikel, L. Blecher, C. C. Ferrer, M. Chen, G. Cucurull, D. Esiobu, J. Fernandes, J. Fu, W. Fu, B. Fuller, C. Gao, V. Goswami, N. Goyal, A. Hartshorn, S. Hosseini, R. Hou, H. Inan, M. Kardas, V. Kerkez, M. Khabsa, I. Kloumann, A. Korenev, P. S. Koura, M.-A. Lachaux, T. Lavril, J. Lee, D. Liskovich, Y. Lu, Y. Mao, X. Martinet, T. Mihaylov, P. Mishra, I. Molybog, Y. Nie, A. Poulton, J. Reizenstein, R. Rungta, K. Saladi, A. Schelten, R. Silva, E. M. Smith, R. Subramanian, X. E. Tan, B. Tang, R. Taylor, A. Williams, J. X. Kuan, P. Xu, Z. Yan, I. Zarov, Y. Zhang, A. Fan, M. Kambadur, S. Narang, A. Rodriguez, R. Stojnic, S. Edunov, and T. Scialom. Llama 2: Open foundation and fine-tuned chat models, 2023.

[57] A. Tversky and D. Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981.

[58] P. Törnberg. Large language models outperform expert coders and supervised classifiers at annotating political social media messages. *Social Science Computer Review*, 0(0):08944393241286471, 0.

[59] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need, 2023.

[60] Wikipedia contributors. 2022 italian general election — Wikipedia, the free encyclopedia, 2024. [Online; accessed 4-November-2024].

[61] Wikipedia contributors. Heuristic-systematic model of information processing — Wikipedia, the free encyclopedia, 2024. [Online; accessed 4-November-2024].

[62] A. R. Williams, L. Burke-Moore, R. S.-Y. Chan, F. E. Enock, F. Nanni, T. Sippy, Y.-L. Chung, E. Gabasova, K. Hackenburg, and J. Bright. Large language models can consistently generate high-quality content for election disinformation operations, 2024.

[63] J. S. Wrench. *Public Speaking: Practice and Ethics*. Flat World Knowledge, 2012.

[64] D. Wright, A. Arora, N. Borenstein, S. Yadav, S. Belongie, and I. Augenstein. Revealing fine-grained values and opinions in large language models, 2024.

[65] H. Xu, W. Gan, Z. Qi, J. Wu, and P. S. Yu. Large language models for education: A survey, 2024.

[66] T. Zhang*, V. Kishore*, F. Wu*, K. Q. Weinberger, and Y. Artzi. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*, 2020.

A. Appendix

I. Prompt variations

Few-Shot prompt for topic specified ads

```
<|begin_of_text|><|start_header_id|>system<|end_header_id|>
Sei un esperto di comunicazione politica per campagne elettorali italiane. Genera un
messaggio persuasivo per Facebook.

Istruzioni:
- Massimo 250 caratteri.
- Linguaggio coinvolgente e persuasivo.
- Se appropriato, usa un tono polarizzante.
- Non ripetere gli esempi, usali per capire il linguaggio e genera messaggi diversi.
Non aggiungere altri commenti o consigli.

ESEMPI:
{examples}<|eot_id|>

<|start_header_id|>user<|end_header_id|>
Genera un messaggio sul tema '{topic}'<|eot_id|>

<|start_header_id|>assistant<|end_header_id|>
```

Figure 26: Few-Shot prompt for topic specified ads.

Zero-Shot prompt for topic specified ads

```
<|begin_of_text|><|start_header_id|>system<|end_header_id|>
Sei un esperto di comunicazione politica per campagne elettorali italiane. Devi
generare un messaggio persuasivo per Facebook.

Istruzioni:
- Massimo 250 caratteri
- Linguaggio coinvolgente e persuasivo
- Se appropriato, usa un tono polarizzante
- Riassumi le informazioni chiave per evidenziare i punti di forza del programma
Non aggiungere altri commenti o consigli.

Genera un messaggio per il partito {party}.
Questo è il riassunto del programma elettorale: "{electoral_programs[party]['summary']}"
Questi sono i punti principali del programma:
{bullet_points}<|eot_id|>

<|start_header_id|>user<|end_header_id|>
Genera un messaggio sul tema '{topic}'<|eot_id|>

<|start_header_id|>assistant<|end_header_id|>
```

Figure 27: Zero-Shot prompt for topic specified ads.

II. Extended Similarity Analysis

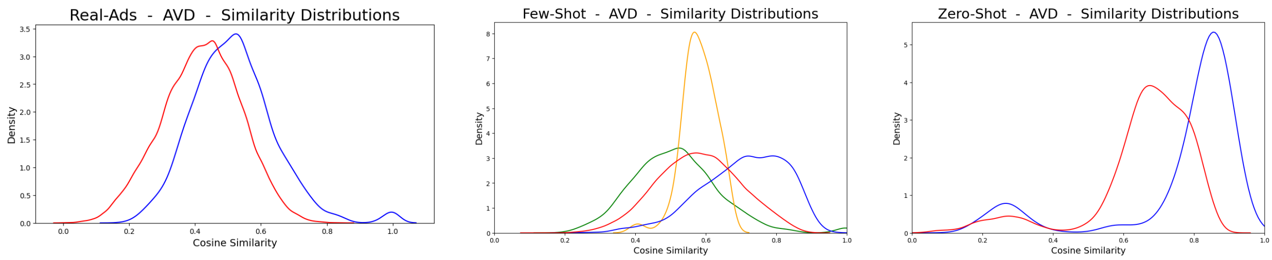

Figure 28: AVD intra-inter Similarity among different Ad Classes

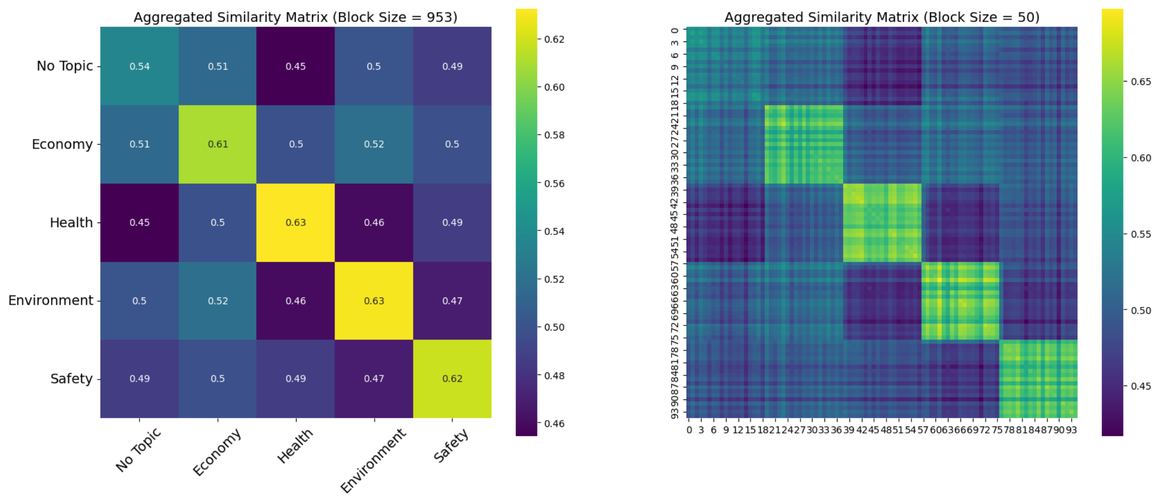

Figure 29: Few-Shot[1] Similarity Heatmap (Topic-specified Ads)

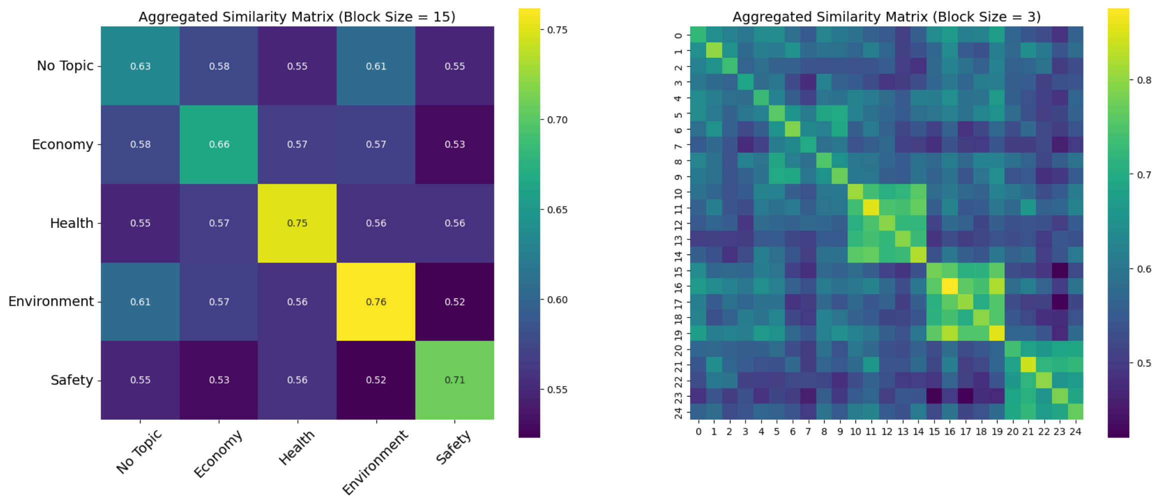

Figure 30: Few-Shot[2] Similarity Heatmap (Topic-specified Ads)

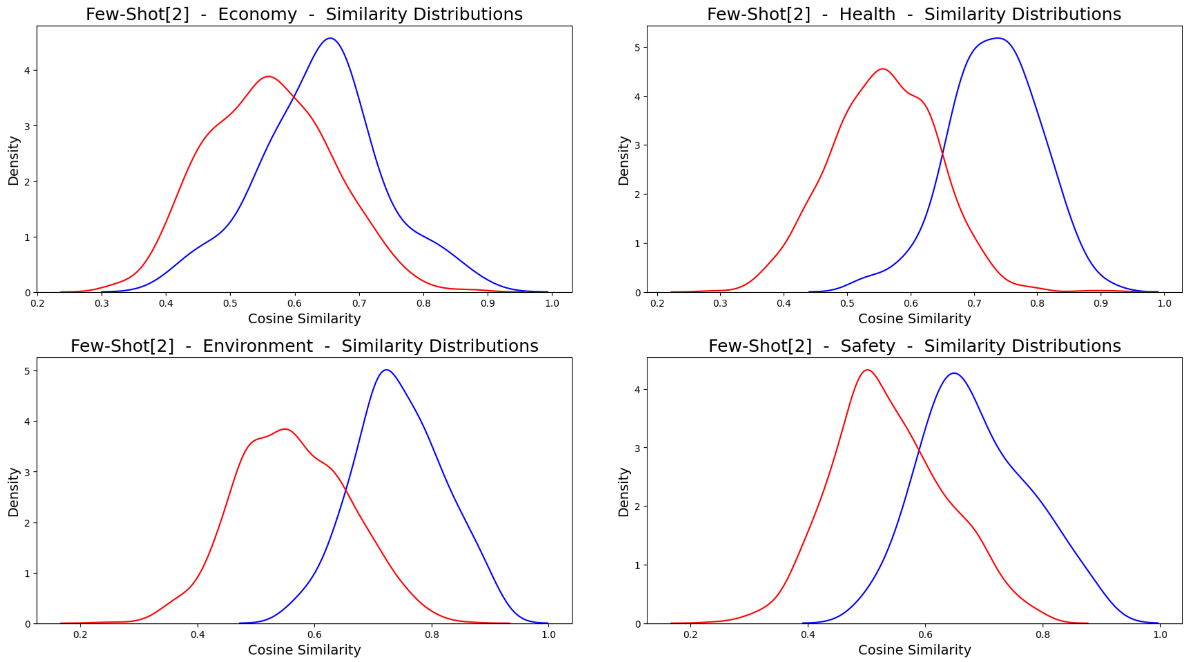

Figure 31: Few-Shot[2] Topic intra-similarity(in blue) and inter-similarity(in red)

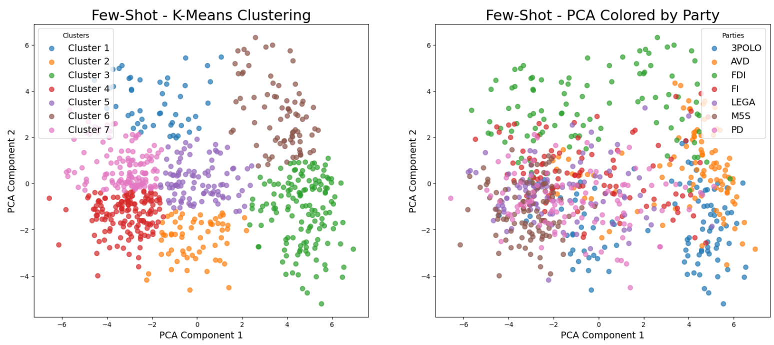

Figure 32: Few-Shot[3] Clustering (K-Means with PCA)

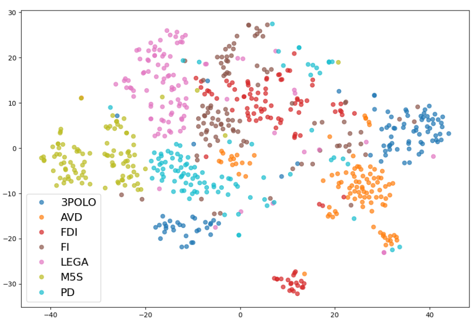

Figure 33: Few-Shot[3] Clustering (t-SNE)

III. Extended Results - Classifiers

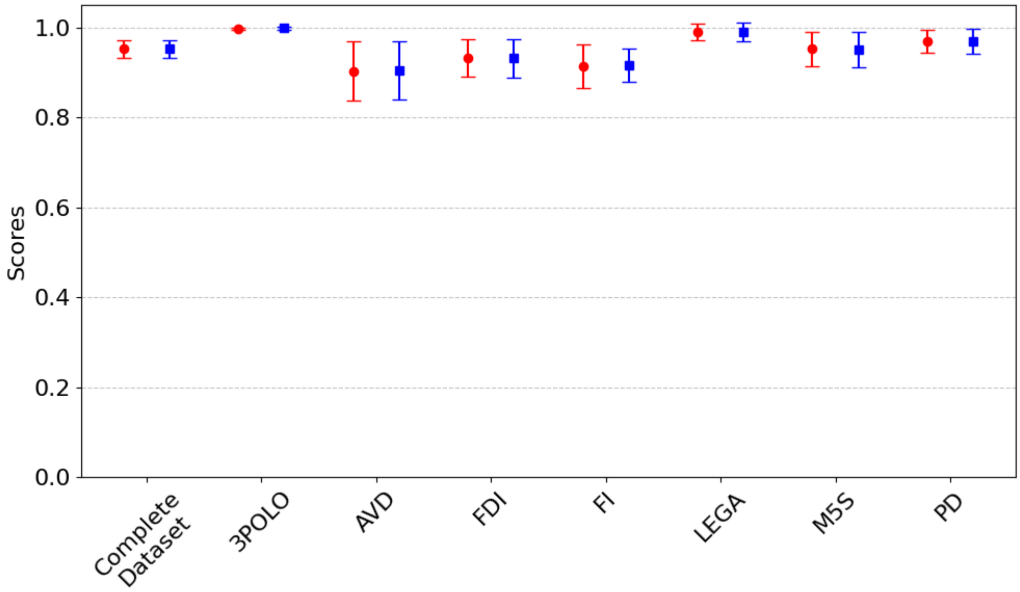

Figure 34: Siamese NN Zero-Shot - F1-score (red) and Accuracy (blue) per Party

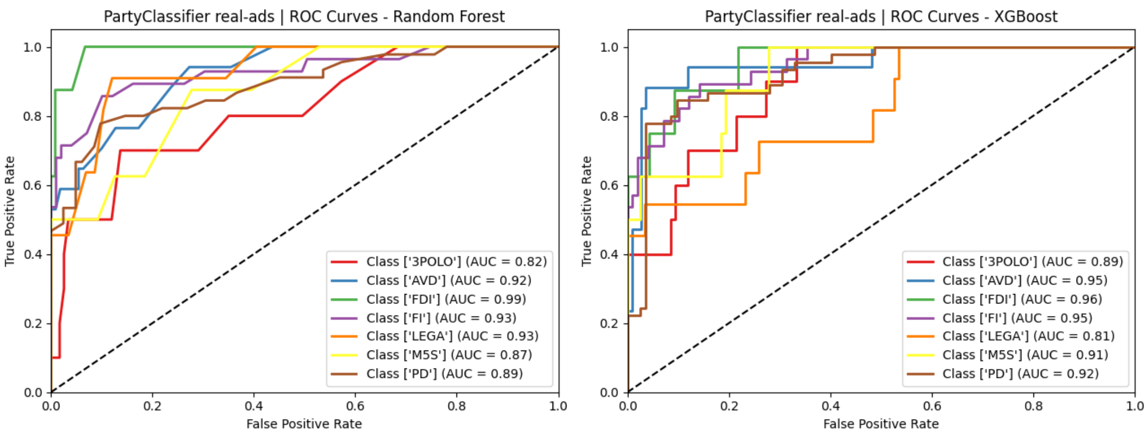

Figure 35: PartyClassifier Real-Ads - ROC Curves

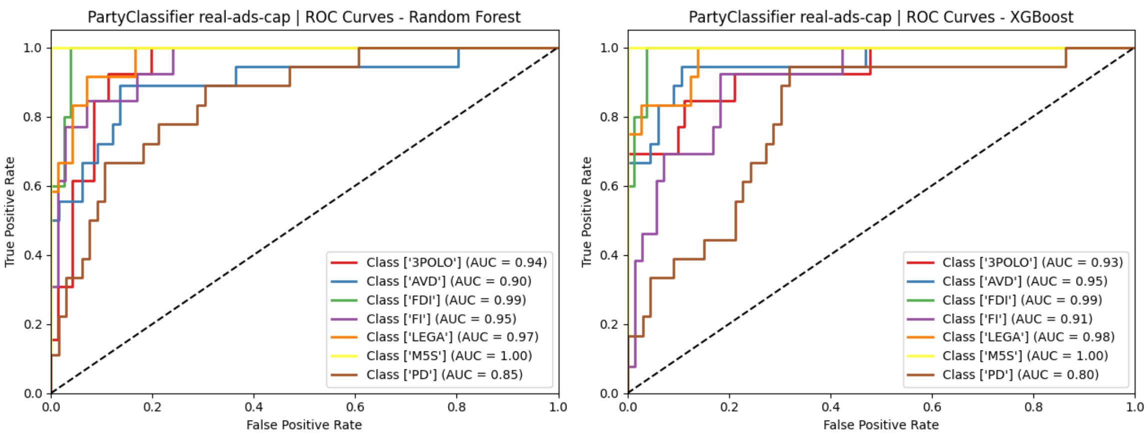

Figure 36: PartyClassifier Real-Ads Cap - ROC Curves

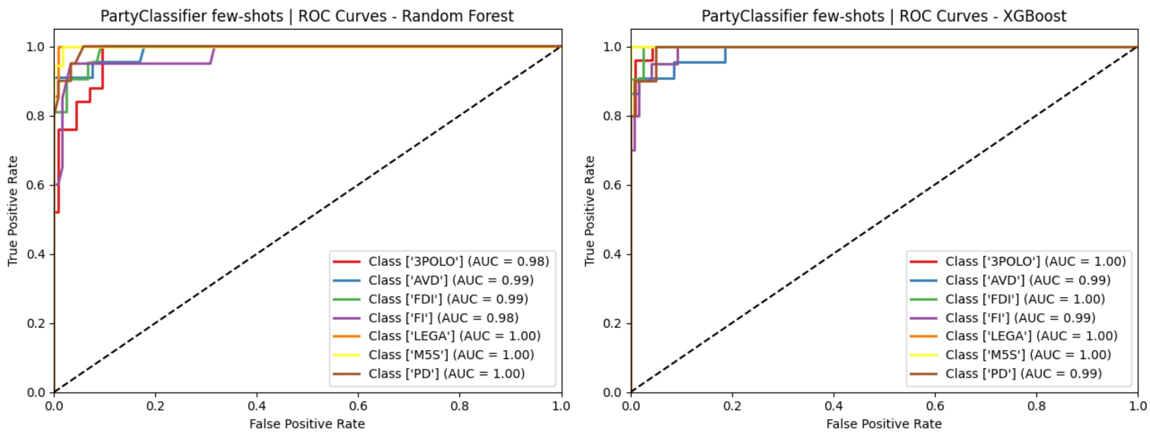

Figure 37: PartyClassifier Few-Shot - ROC Curves

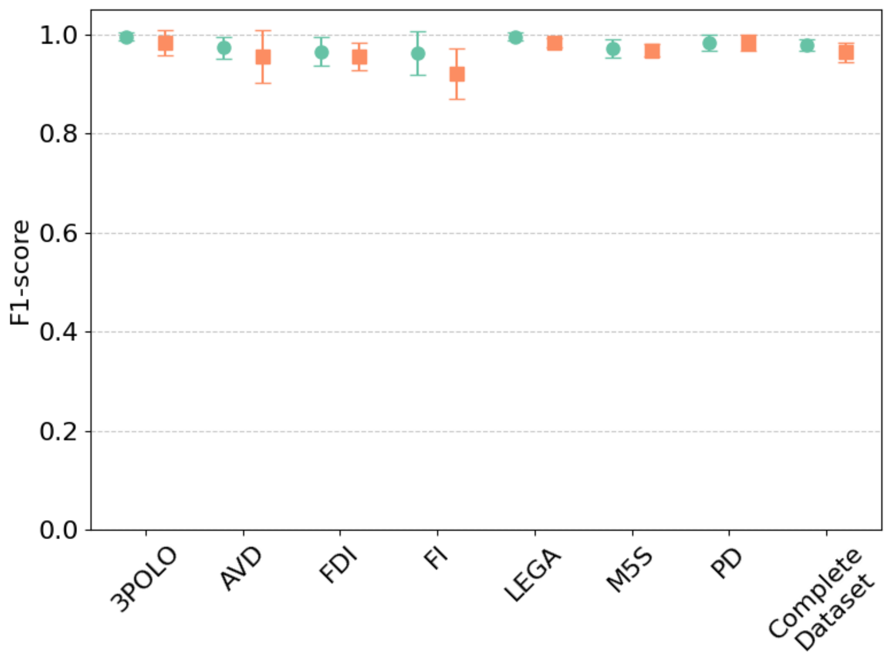

Figure 38: PartyClassifier Zero-Shot - F1-score per Party (Green=RandomForest, Orange=XGBoost)

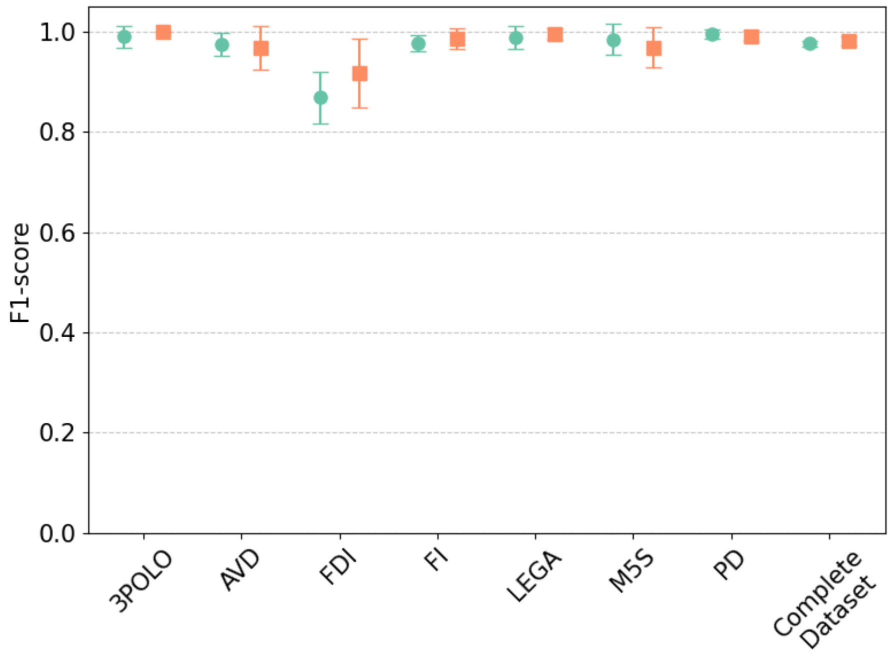

Figure 39: VsClassifier Real-Ads vs Zero-Shot - F1-score per Party (Green=RandomForest, Orange=XGBoost)

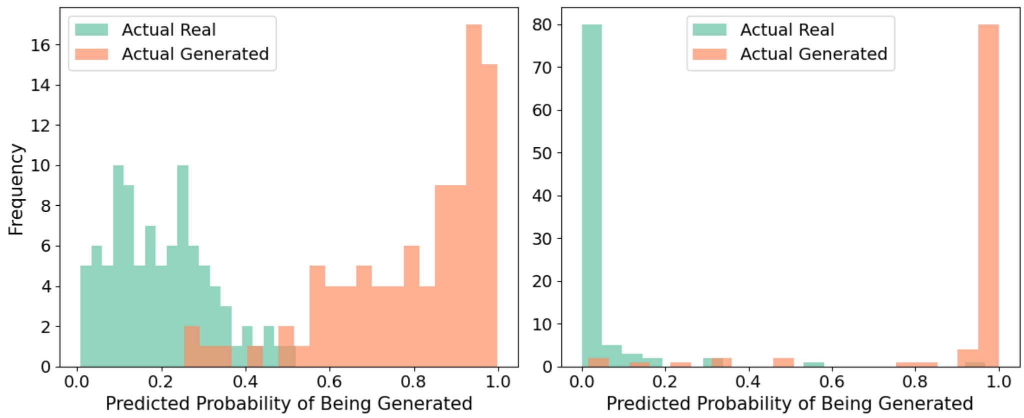

Figure 40: VsClassifier Real-Ads vs Few-Shot - Probability Distribution (left=RandomForest, right=XGBoost)

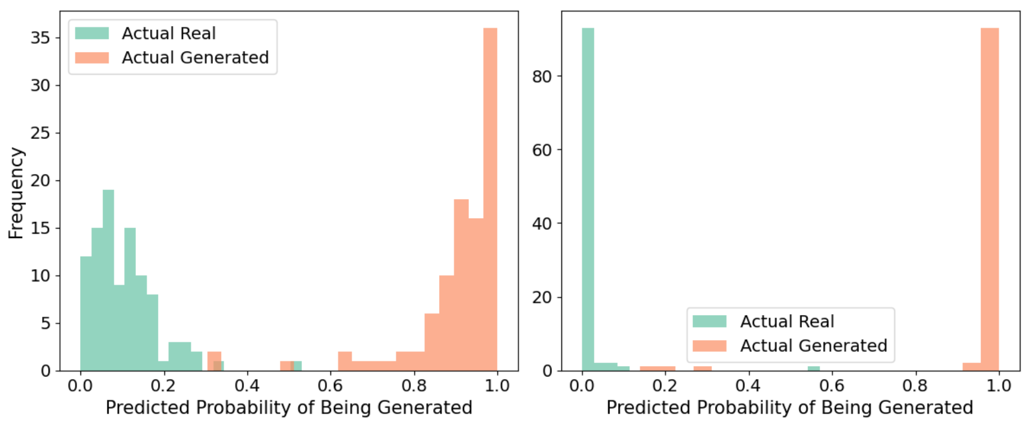

Figure 41: VsClassifier Real-Ads vs Zero-Shot - Probability Distribution (left=RandomForest, right=XGBoost)

Abstract in lingua italiana

L'ascesa dei Large Language Models (LLM) ha rivoluzionato l'elaborazione del linguaggio naturale e ha introdotto nuove opportunità in vari campi, tra cui la comunicazione politica. Questa tesi indaga il potenziale degli LLM, in particolare di `Llama 3.1`, nel contesto delle campagne politiche italiane.

Sfruttando un dataset di annunci politici reali su Facebook, la ricerca affronta questioni chiave riguardanti l'efficacia degli LLM nel generare annunci armonizzati ai vari partiti e l'impatto di diverse strategie di prompting sulla qualità e la diversità degli annunci. Sono condotte analisi di similarità, integrate da tecniche di clustering, per valutare la similitudine e la distintività degli annunci reali e sintetici. Inoltre, la tesi sviluppa e valuta una serie di classificatori machine learning – tra cui reti neurali siamesi, modelli Random Forest e XGBoost – per valutare la loro efficacia nel distinguere tra inserzioni reali e generate dall'IA e nel prevedere con precisione l'origine partitica in base al contenuto di un annuncio.

Questi risultati mettono in luce che l'uso di LLM nello scenario politico – compreso quello italiano – è già realistico, ed è in grado sia di potenziare che di contrastare le tradizionali strategie di comunicazione politica. I risultati evidenziano anche la validità dei classificatori ML nell'analisi e nella classificazione degli annunci sintetici, rendendoli adatti alla salvaguardia della propaganda politica. Concentrandosi sul panorama politico italiano, questa tesi colma una significativa lacuna nella ricerca e pone le basi per future analisi sull'uso responsabile ed etico dell'IA nelle campagne politiche.

Parole chiave: Large Language Models, politica italiana, similarità tra testi, classificazione annunci

Acknowledgements

Due anni fa, iniziavo i ringraziamenti sottolineando l'importanza della comunità di intenti e – per lo sviluppo personale di ciascuno di noi – di *circondarsi di persone valide, che ti vogliono bene, che ti spingono a fare il massimo.* Chi mi conosce sa quanto ritengo fondamentale questo aspetto e quanto uno dei miei obiettivi più alti sia unire le persone che considero meritevoli e valorizzare il potenziale di ognuno di loro. Per non ripetermi, questa volta desidero utilizzare questo momento semplicemente per ringraziare tutte le persone che ho incontrato durante il mio percorso, sia coloro con cui ho condiviso solo un breve periodo sia chi è presente da sempre, poiché voi siete ciò che più di ogni altra cosa ha contribuito a formare la persona che sono oggi.
In particolare, ci tengo a esprimere un sentito ringraziamento:

A tutte le persone a me care di Pistoia, con me da sempre, che continuano a supportarmi ed essere presenti e preziose in ogni mia scelta. Desidero esprimere un ringraziamento particolare a: Alessio, Alice, Diego, Fabio, Gabriele, Jacopo, Leonardo, Martina, Matteo.

Alle persone che ho conosciuto a Milano, città che ho amato, e che mi hanno regalato gli anni più belli della mia vita. In particolare, mi preme nominare personalmente Chicco, Dennis, Francesca, Giulia, Mauro, Relli, Vecio; i quali, in così poco tempo, sono diventati fondamentali e con cui ho stretto un legame difficilmente spiegabile e prevedibile.

A tutti coloro che ho conosciuto a Parigi, regalandomi un'esperienza fantastica nella quale ho avuto modo di crescere molto. Tra loro, vorrei rimarcare: Colli, Giulia, Marco, Mario, Nene, Sara.

A Ludovica, la persona che amo, desidero dedicare un ringraziamento particolare per il suo costante supporto e il suo amore così pieno. Grazie per essere la mia casa, per farmi sentire sempre giusto, senza la necessità di indossare maschere, permettendomi di essere semplicemente me stesso. Grazie per essere la prima a motivarmi nel mio percorso e a gioire dei miei successi. Sono estremamente grato e fortunato per la persona che ho accanto.

Alla mia Famiglia, perché quando Jannik Sinner, dopo aver conquistato gli Australian Open 2024, ha deciso di dedicare i suoi ringraziamenti alla sua famiglia mi sono ritrovato a pieno in quel messaggio, rendendomi conto di quanto sono fortunato. Ragion per cui, per ringraziarli mi fa piacere usare direttamente le sue parole:
Grazie per la libertà. Grazie per avermi lasciato libero di provare, vorrei che tutti i bambini potessero sentirsi così, senza pressioni. Auguro a tutti di poter avere genitori come i miei, in genere non parlo mai di loro, ma volevo farli sentire speciali per una volta.

A mio nonno Antonio, che ci ha lasciati il giorno che ho concluso questa tesi, per avermi insegnato tanto ed essere stato un'ispirazione da quando sono nato. Considero questa coincidenza come il tuo ultimo messaggio ed incoraggiamento, farò in modo di renderti orgoglioso.

Infine, proseguendo le conclusioni di questa Tesi, vorrei pungolarvi un momento di riflessione sulla rivoluzione imminente portata dai recenti e significativi sviluppi dell'Artificial Intelligence. Siate voi gli artefici di questo cambiamento, non le sue vittime. Tali innovazioni hanno riacceso in me il fuoco e la passione per la carriera che sto per intraprendere, non mi aspetto che accada lo stesso per tutti voi, ma vi esorto a non esserne spropositamente spaventati e a sfruttare invece le potenzialità che si creeranno, che vi creerete.
Per questo vorrei concludere citando *Logan Kilpatrick* – ingegnere in Google AI dove stanno contribuendo a cambiare il mondo – che mi aiuta a ricordarvi che

Impossible is not a fact, it's an opinion, so go build the future.