İLETİŞİM ve DİPLOMASİ
Communication and Diplomacy
ISSN: 2147-6772 e-ISSN: 2791-7649 Yıl/Year: Aralık/December 2023 Sayı/Issue: 11

Dissecting Disinformation Dynamics: Insights from a Social Media Environment

Dezenformasyon Dinamiklerini Ayırmak: Sosyal Medya Ortamından İçgörüler

Araştırma Makalesi / Research Article

Yavuz Selim BALCIOĞLU[1], Bülent DOĞAN[2]

Sorumlu yazar/ Corresponding author: Yavuz Selim Balcıoğlu

ORCID: 0000-0001-7138-2972

Geliş tarihi/Received: 12.10.2023

Son revizyon teslimi/Last revision received: 9.12.2023

Kabul tarihi/Accepted: 12.11.2023

Yayın tarihi/Published: 16.12.2023

Atıf/Citation: Balcıoğlu, Y.S. & Doğan, B. (2023). Dissecting disinformation dynamics: Insights from a social media environment, *İletişim ve Diplomasi*, *11*, 107-125.

doi: 10.54722/iletisimvediplomasi.1374744

ABSTRACT

In the digital age, a period we might characterize as a time when societal, economic, and scientific shifts have redefined the trajectory of transformation, leading to the emergence of a networked society; rapid advancements in communication technologies, especially the surge in internet users, increased internet speeds, and enhanced internet and application usability on mobile devices have begun to render traditional media obsolete. This has paved the way for newer digital communication platforms endowed with interactive features. Notably, social media platforms provide users with the means to share information, emotions, thoughts, and ideas more efficiently and swiftly. With advancements in internet technologies, social media platforms have become accessible to vast audiences, resulting in a structure that can be described as a social network society. Consequently, information content and news can be rapidly disseminated and shared with the masses. However, in such communication environments, news is often relayed without verification or is deliberately misconstrued, leading to the emergence of disinformation comprised of fake news and inaccuracies. While controlling this process in a globalized world might pose challenges, the research herein proposes an artificial intelligence-based approach capable of discerning the veracity of news and swiftly verifying it. We extracted data from a platform Facebook and found patterns indicating a significant prevalence

1 Dr., Gebze Teknik Üniversitesi, Kocaeli, Türkiye, ysbalcioglu@gtu.edu.tr

2 Doktora öğrencisi, Gebze Teknik Üniversitesi, Kocaeli, Türkiye, bulentdogan@gtu.edu.tr, ORCID: 0009-0000-7584-5129

of disinformation. Of the 5,000 posts assessed, nearly one-fifth were flagged as misleading. Age, post engagement, and network size, often hypothesized as potential influencers, displayed only weak correlations with the propensity to share or engage with disinformation. The multifaceted nature of disinformation spread underscores the need for an integrated approach, combining technology and user education, to combat its proliferation on digital platforms. These findings demonstrate the pressing need and potential efficacy of AI-driven solutions in countering disinformation in today's digital communication landscape.

Keywords: Disinformation, social media, artificial intelligence, machine learning, media

ÖZ

Dijital çağ, toplumsal, ekonomik ve bilimsel değişikliklerin dönüşümün yörüngesini yeniden tanımladığı bir dönemde; ağa bağlı bir toplumun ortaya çıkmasına neden olmaktadır. İletişim teknolojilerindeki hızlı ilerlemeler, özellikle internet kullanıcılarındaki artış, artan internet hızları ve mobil cihazlarda gelişen internet ve uygulama kullanılabilirliği, geleneksel medyanın kullanımını azaltmaya başlamıştır. Bu durum, etkileşimli özelliklere sahip yeni dijital iletişim platformlarının yolunu açmıştır. Özellikle sosyal medya platformları, kullanıcılara bilgi, duygu, düşünce ve fikirleri daha etkili ve hızlı bir şekilde paylaşma olanağı tanımaktadır. İnternet teknolojilerindeki ilerlemelerle birlikte sosyal medya platformları geniş kitlelere ulaşabilmekte ve bu yapı bir sosyal ağ toplumu olarak tanımlanmaktadır. Sonuç olarak, bilgi içeriği ve haberler hızla yayılarak büyük kitlelerle paylaşılabilmektedir. Ancak, böyle iletişim ortamlarında haberler genellikle ya doğrulanmadan aktarılmakta ya da kasıtlı olarak yanıltıcı bir şekilde sunulmaktadır. Bu da yanıltıcı haberler ve yanlış bilgilerden oluşan bir dezenformasyonun ortaya çıkmasına neden olmaktadır. Küreselleşmiş bir dünyada bu süreci kontrol etmek beraberinde zorluklar da getirmektedir. Bu araştırmada haberin gerçekliğini ayırt edebilen ve hızla doğrulayabilen yapay zekâ tabanlı bir yaklaşım önerilmektedir. Facebook üzerinden elde ettiğimiz bir veri kümesi ile yaptığımız inceleme, önemli dezenformasyon yaygınlığına dair örüntüler ortaya koymaktadır. Değerlendirilen 5 bin gönderinin yaklaşık beşte biri yanıltıcı olarak işaretlenmiştir. Potansiyel etkenler olarak hipotez edilen yaş, gönderi etkileşimi ve ağ büyüklüğü, dezenformasyonla etkileşme veya paylaşma eğilimiyle sadece zayıf korelasyonlar göstermiştir. Dezenformasyonun yayılmasının çok yönlü doğası, dijital platformlarda yayılmasını engellemek için teknoloji ve kullanıcı eğitiminin birleştirilmesi gerektiğini vurgulamaktadır. Bu bulgular, bugünkü dijital iletişim manzarasında dezenformasyona karşı yapay zekâ destekli çözümlerin acil ihtiyacını ve potansiyel etkinliğini göstermektedir.

Anahtar Kelimeler: Dezenformasyon, sosyal medya, yapay zekâ, makine öğrenmesi, medya

INTRODUCTION

Thanks to advancements in internet technologies, new changes have emerged in societies' understanding of reality (Yurdigül & Yıldırım, 2021). Especially, with the internet networks accessible from almost everywhere, social media platforms have become an environment where people can share their emotions, thoughts, information, critiques, and experiences in real-time and receive instant feedback. Now, individuals can also produce content as they desire. In the current state, the pervasive presence and widespread use of social media in every moment of human life have accelerated access to information and made news and information consumption extremely easy. Additionally, the fact that user-generated content, information, or news can reach millions of people in a very short time frame demonstrates how effective social media platforms are in the speed of information dissemination. The ease of access to information and news, which are among the purposes of users using social media platforms, can be cited as the most important aspect of these platforms (Koçyiğit & Koçyiğit, 2023). However, an important issue that arises with this aspect of social media is this: The low awareness and recognition regarding the presence of false news or incorrect information in social media content that can reach vast societal masses causes such content to spread extremely quickly. So, how can this situation be prevented? (Karakoç & Zeybek, 2022). Especially in recent years, the increase in false news and disinformation spread deliberately or inadvertently on social media has heightened sensitivity to the issue, and significant steps are being taken by public authorities to combat false news and disinformation. At the same time, studies on the detection of false news content and disinformation are increasing day by day.

While the developments in the digital world have brought about positive changes in people's lives, the transformation of the perception of reality into fake realities has caused confusion. Individuals now have manufactured realities in their lives that did not exist in earlier times, differing from the past. New technological advancements have made it difficult for individuals to differentiate between the real and the fake (Karakoç & Zeybek, 2022). Especially with the development of artificial intelligence technology, fake content designed to alter people's perception of reality can produce rapid and persuasive results on individuals, thanks to the reach of social media platforms to millions. Moreover, ethically in journalism, confirming the accuracy of information sources and refining them before publishing is an ethical and moral rule (Toğaçar, 2022). This rule can sometimes be violated when motivations like the urge to break news quickly, ambition, vested interests, etc., take precedence. With artificial intelligence technology, machines can mimic human-specific abilities such as thinking, decision-making, and logical reasoning, and they execute these actions thanks to deep learning models, a subfield of artificial intelligence (Sun et al., 2020).

Thanks to advancements in artificial intelligence and information and communication technologies, information content has now become manipulatable, and the created content carries significant risks in terms of reality, leading to the emergence of disinformation with societal impacts. For instance, visual disinformation supported by artificial intelligence technologies, primarily defined as Deepfake, and the artificial intelligence-based language model ChatGPT, can be cited as applications that can manipulate and influence the public en masse, threatening democracy, individuals' rights, and privacy (Etike, 2023). With these and similar technologies, the issue of content reliability and reality has arisen, and especially with Deepfake, synthetic videos that resemble real videos have become producible.

As a result of literature research, it has been observed that many studies have recently been conducted on the detection of fake news based on artificial intelligence. Some of these studies are as follows: In their study, Rohit Kumar Kaliyar and his friends developed a deep learning model named FNDNet to classify fake news (Kaliyar et al., 2020). Feyza Altunbey Özbay and Bilal Alataş used various artificial intelligence models in their studies to determine fake news. In their research (Özbay & Alataş, 2020), Deepak Sreekumar and Bhadrachalam Chitturi utilized the Natural Language Processing model in conjunction with the UKSB model for the detection of fake news data (Sreekumar & Chitturi, 2020). Mesut Toğaçar, Kamil Abdullah Eşidir and Burhan Ergen mentioned in their work that the findings obtained regarding the identification of fake news on the internet using an artificial intelligence-based Natural Language Processing approach are promising for future studies (Toğaçar et al., 2022). Elif Karakoç Keskin and Burcu Zeybek evaluated the impact of Deepfake-based manipulative content on communication in politics after understanding how Deepfake technology, which can manipulate visual and auditory content, operates (Karakoç & Zeybek, 2022). In his study, Nurat Kara expressed that information distortion, detailed as disinformation, has begun to influence not only societies but also the public and private sectors (Kara, 2022). He emphasized the need for all stakeholders of society, especially the public sector, to define a digital transformation strategy in collaboration to combat disinformation.

In our study, we conducted an analysis aimed at detecting manipulative writings on untrue news on a social media platform, Facebook. However, it should be noted that the ability to detect fake and unreal content, which can be created especially with artificial intelligence-supported technologies, through artificial intelligence-based technologies and to prevent this content is found to be noteworthy and significant. Furthermore, it is believed that the dilemmas brought by the choices of using the obtained advanced technology for good or bad purposes continue to exist.

Disinformation

In this section of our study, the concept of "disinformation" is defined and explained. According to the Turkish Language Association (TDK, 2023), "disinformation," expressed as "information distortion," is described by the United Nations Organization as the act of transmitting false information to people and deliberately manipulating them (Kara, 2022). The concept of disinformation originated from the Russian word "dezinformacija" in 1949 and is defined as the intentional conveyance of misleading information to people or the targeted audience (Karlova and Fisher, 2013).

According to Fetzer, disinformation is the dissemination of information that is false, incomplete, or far from the truth with the intention of misleading a particular audience about the facts (as cited by Ertem from İnceoğlu and Akıner, 2019). Especially in social media platforms, malicious content photos, fake documents, false or deceitful propaganda, advertisements causing misperceptions, and political propaganda can be cited as examples of disinformation. All these fake and/or incorrect contents have negative physical and emotional impacts (Fallis, 2014).

In today's world, fake news, untrue information, and content produced for deceitful purposes are now conveyed to vast audiences through digital platforms, particularly social media. The production of information containing disinformation can be done in various ways. Producing content that is directly false, mixing true information with false, conveying accurate information incompletely, and adding unnecessary information elements are types of disinformation. Today, entities professionally producing disinformation content can include public authorities, private sector organizations, and media entities. Even in the realm of international relations, states have now positioned themselves to conduct disinformation activities against other countries, either for their own interests or as preventive measures.

An example of using disinformation to manage information and perception towards masses can be drawn from recent history: before the U.S. invasion of Iraq, the assertion that Saddam possessed weapons of mass destruction was presented to the public through both official channels and all kinds of media outlets. The American public, initially resistant to sending their troops to war, and the entire world were convinced through the constant news presented to them, persuaded that something needed to be done. Although the confession that no weapons of mass destruction were found came post-war, the disinformation had achieved its goal.

Another example is the disinformation based on a 15-year-old Kuwaiti girl, which was used by the U.S. administration seeking support from the United Nations Organization during the First Gulf War, yielding results.

The Nayirah Incident

“On October 10th, during a testimony in front of the Congressional Human Rights Caucus, a 15-year-old Kuwaiti girl named Nayirah tearfully recounted how armed Iraqi soldiers took newborn babies out of incubators in a hospital and left them to die on cold concrete. The story of the babies became an NBC TV news headline that same night and was watched by 35 million people. The story was rebroadcast on other TV channels for days.

On November 27th, a show was presented to the Security Council, displaying pictures showing Iraqi soldiers torturing Kuwaiti prisoners. Alleged eyewitnesses described the torture. Two days later, the Council set a deadline of January 15, 1991, for Iraq to withdraw from Kuwait.

While the U.S. Congress was still debating military action, a new research report brought the drama of the children left to die on the hospital floor to the Congressional Foreign Relations Committee. Amnesty International (AI) also confirmed the baby story. After intense deliberations, Congress granted President Bush war powers.

After the war, ABC TV reporter John Martin conducted investigations in the hospital in Kuwait. Kuwaiti doctors reported that the babies died from neglect due to the chaos of war and a shortage of nurses, and not a single baby was removed from an incubator by Iraqi soldiers. Researchers sent by AI to the hospital received the same response from Kuwaiti officials, leading Amnesty International to retract its initial findings with an apology.

The protagonist of the fabricated scenario, Nayirah, was in fact the daughter of Kuwait’s Washington Ambassador Saud Al-Nasser Al-Sabah. The girl, who made millions cry on TV, was in the U.S. throughout the war. The Kuwaitis supposedly tortured by Iraqis were fake. It was revealed that the State of Kuwait paid $11.5 million to a public relations firm, Hill and Knowlton (H and K), which scripted and executed the entire scenario to encourage military action against Iraq.” (Uluç, 2003).

In today’s world, disinformation activities are carried out almost everywhere and on every topic to manipulate or convince the masses. The areas of disinformation use have diversified and spread. From disinformation to justify war, to propaganda-based disinformation by politicians to gather votes, to unethical disinformation aimed at tarnishing a rival company with false accusations, these activities are carried out in many areas and even on many platforms. Even if the truth of widely spread news is published later, it is often impossible for the truth to reach as many people as the falsehood did. Subsequent corrections are not as attention-grabbing as the false news. Ideological distortions aimed at individuals/institutions/communities are also common disinformation targets. Even if it’s later discovered to be untrue, the irreparable harm remains.

For disinformation purposes, various contents like photos and videos are used. These contents are shared with the target audience along with the false information added to them. The source of information can also be used as a source of disinformation. Disinformation content, which anyone with computer and internet access can easily create and remain anonymous, can also be produced (Ertem, 2019).

Disinformation sources aren't just limited to TV or the internet. Friends or acquaintances who have received disinformation can also be influential. Also, even if disinformation is produced by an organization or individual, sometimes external situations arise. News channels or newspapers can produce false news by accepting information produced by someone else as a reference. The media may not be the primary source of disinformation, but it serves as a carrier and distributor because they trust the source without questioning (Turan, 2015). The media can be used by commercial and financial institutions to convey disinformation. For example, sudden death of a country leader, resignation of a company CEO, or sensational news being withheld from the public for commercial reasons can result in significant financial gain or loss based on the public's reaction. Public administrations can also use disinformation to quell public reaction or divert attention. Examples include presenting inflation data with the most reduced products or a sensational event erupting amidst heavy price hike news.

In e-commerce, which has started to hold a significant place in commercial activities, sites influence purchase decisions by spreading disinformation. For example, some products that are desired to be sold, rather than those actually sold, can be listed under "best sellers".

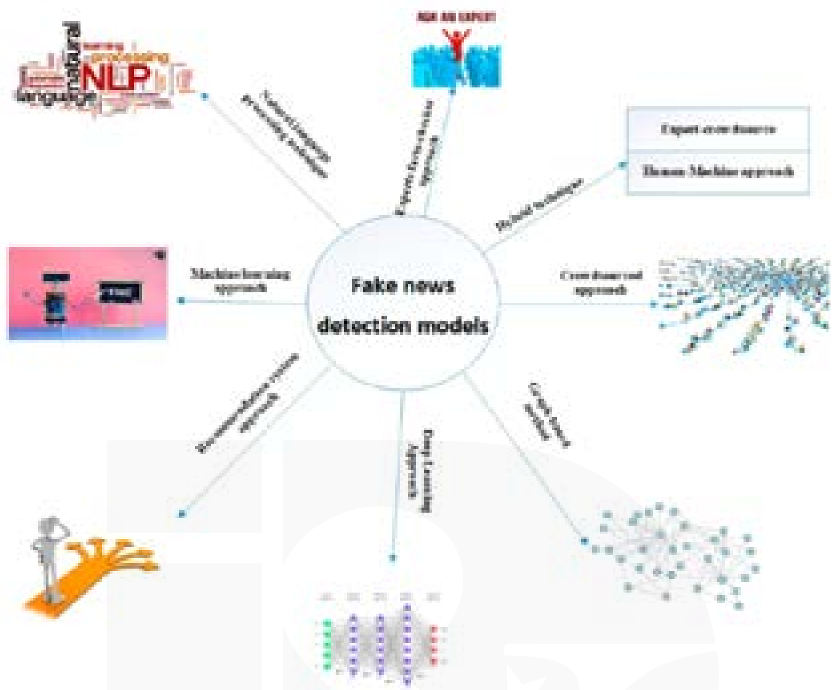
Wardle has divided information pollution into three different types based on the harm they cause and their degree of falsehood, providing a conceptual framework for the definition and scope of disinformation: misinformation, disinformation, and malinformation (Wardle and Derakhshan, 2017).

- Disinformation; the deliberate sharing of false information with the intent to harm,
- Misinformation; sharing incorrect information without intent,
- Malinformation; sharing truthful information with the intent to harm an individual or organization.

Detection of Disinformation on Social Media Platforms

Artificial intelligence-based models have been proposed by some researchers to detect disinformation content produced on social media platforms (Ertem, 2019). The AI-based models used for the detection of disinformation are shown in the figure below.

Figure 1. Disinformation Detection Models



Source: Collins et. al, 2021.

As shown in Figure 1, the models developed to detect fake news can be expressed as follows:

Expert Verification Approach: It is the denial or confirmation of specific content by experts.

Crowdsourced Approach: An approach where the collective effort of groups or individuals replaces the intellectual capacity of any individual.

Machine Learning Approach: Especially eye-catching headlines are used for the spread of content produced as disinformation. In such cases, it is necessary to extract the text and language features of the contents for the machine learning approach.

Natural Language Processing (NLP) Technique: Real and unreal content is classified in pairs, and the contents are analyzed in terms of words and meaning.

Hybrid Technique: It includes techniques in the content-based model as well as social context-based techniques using auxiliary information from different perspectives.

Graph-Based Method: It ensures that users connected to a graphic network have the same features and view the same news content or similar content.

Deep Learning Approach: It uses neural networks to determine the authenticity of the produced content.

Recommendation System Approach: It tries to verify some news content believed to be unique and then recommend these news articles for consumption (Collins et al., 2021).

Method

Data Generation

To explore the dynamics of disinformation on a platform analogous to Facebook, we gathered the data from Facebook. This dataset is representative of user demographics, posting behavior, and the nature of disinformation content. While the data is artificial, the structure and distributions are designed to mirror potential real-world scenarios, making the insights derived from it valuable for preliminary investigations or simulations.

User Demographics

We gathered data from 1,000 users, capturing attributes such as user ID, name, age, location, and the number of friends. Ages were identified drawn from a uniform distribution ranging from 18 to 70 years. The users were assumed to be from 20 distinct locations, and the number of friends per user was modeled to vary between 50 and 5,000.

Posting Behavior

A total of 5,000 posts were gathered from Facebook, with each post randomly attributed to a user. The posts were distributed over the span of a year. Each post contains metadata including the post ID, the ID of the posting user, date and time of the post, content, number of likes, and number of shares. The content of the post in this dataset is placeholder text.

Disinformation Flagging

A subset of these posts was flagged as containing disinformation. The flagging was binary, indicating the presence or absence of disinformation. Approximately 20% of the posts were labeled as disinformation based on a random sampling. For each disinformation post, a confidence level was assigned, ranging from 0.5 to 1, indicating our system's confidence in the disinformation label. Further, each disinformation post was categorized into types such as "Political", "Medical", or "Financial".

Data Analysis

Following the data generation, a series of exploratory and inferential analyses were planned to uncover patterns, correlations, and key metrics. These include basic descriptive statistics, disinformation-centric analysis, location-based trends, time-series evaluations, and correlation studies. We employed Python (McKinney and Team, 2015), a versatile and powerful programming language, to analyze the dataset sourced from Facebook. Python, known for its extensive libraries and tools tailored for data analysis, offered a robust platform to dissect, process, and interpret the large volume of data in our possession. By leveraging libraries such as Pandas for data manipulation, NumPy for numerical computations, and Matplotlib and Seaborn for visualizations, we were able to seamlessly navigate through the dataset, ensuring precise and accurate results. Moreover, the use of Python allowed us to implement advanced statistical techniques to determine correlations and patterns, providing a comprehensive insight into the prevalence and nuances of disinformation within the Facebook ecosystem. The decision to utilize Python was grounded in its efficiency, scalability, and its widespread recognition within the academic and research community for data-driven studies.

Table 1. User age statistics

Statistic	Value
Count	1000.0
Mean	43.49
Standart Deviation	15.07
Minimum	18.0
25th Percentile	30.75
Median	44.0
75th Percentile	57.0
Maximum	69.0

Table 2. Number of friends statistics

Statistic	Value
Mean	2546.56
Standart Deviation	1439.54
Minimum	51.0
25th Percentile	1317.75
Median	2596.0
75th Percentile	3797.75
Maximum	4990.0

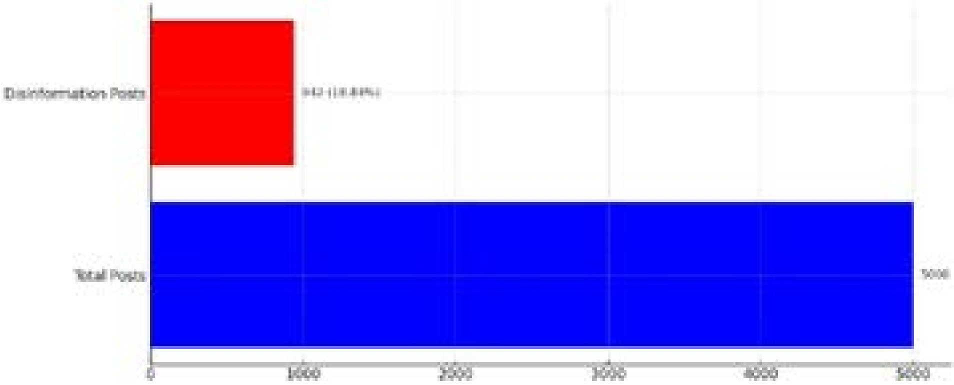
Table 3. Post likes statistics	
Statistic	**Value**
Count	5000.0
Mean	246.73
Standart Deviation	142.34
Minimum	0.0
25th Percentile	125.0
Median	243.0
75th Percentile	370.0
Maximum	499.0

Table 4. Post shares statistics	
Statistic	**Value**
Mean	49.71
Standart Deviation	28.84
Minimum	0.0
25th Percentile	25.0
Median	50.0
75th Percentile	74.0
Maximum	99.0

In our dataset, representing a platform analogous to Facebook, we profiled a total of 1,000 users. The age distribution of these users presented a mean age of 43.49 years with a standard deviation of 15.07 years, indicating a broad range of user demographics. The youngest user was 18 years old, while the oldest was 69, with the middle 50% of the users falling between the ages of 30.75 and 57 years. On the aspect of social connections, users, on average, had approximately 2,546.56 friends. However, the extent of their network varied significantly, with a standard deviation of 1,439.54 friends. While some users had a modest circle with as few as 51 friends, others boasted networks nearing 5,000 friends. When examining user engagement metrics, our dataset encompassed a total of 5,000 posts. These posts, on average, garnered 246.73 likes, with the engagement ranging from no likes to a high of 499 likes. The spread of the data, as indicated by a standard deviation of 142.34 likes, suggests diverse user engagement levels. Sharing behavior followed a similar pattern, with posts being shared an average of 49.71 times. The distribution of shares was slightly more concentrated, given the standard deviation of 28.84 shares.

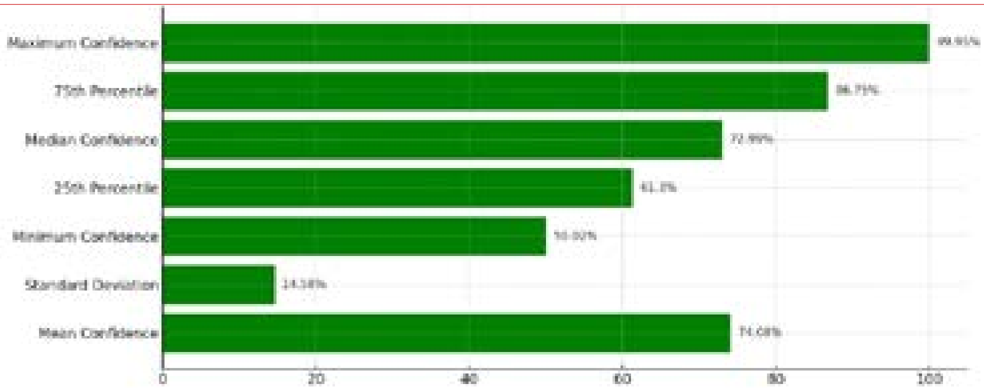
Disinformation Analysis

Figure 2. Disinformation Prevalence



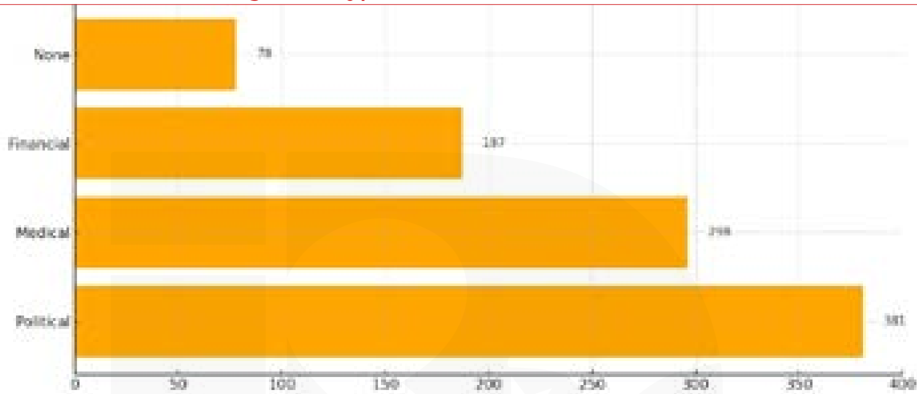
An in-depth examination of a dataset representative of Facebook's post dynamics revealed alarming trends. Out of a sample of 5,000 posts analyzed, a staggering 942 posts (figure 2), constituting 18.84% of the total, were flagged as disinformation. This demonstrates a substantial presence of misleading content, further highlighting the challenge faced by both users and platform administrators in discerning genuine content from fabricated or manipulated narratives. The aforementioned statistics underscore a pressing concern in today's digital age where information dissemination is swift and vast. A nearly 20% prevalence rate of disinformation suggests that users are frequently exposed to misinformation, potentially skewing perceptions and influencing decisions based on false premises. The reliability of content on Facebook, and similar platforms, becomes questionable, necessitating the development of more sophisticated content verification mechanisms and fostering a more discerning user base. In the future, ensuring the integrity of content shared will be paramount for maintaining the trust and safety of the platform's vast user community.

Figure 3. Confidence levels for disinformation flags



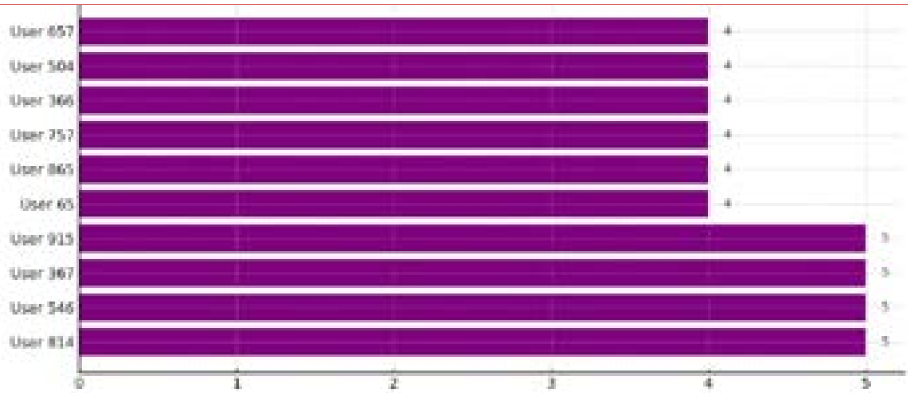
Understanding the accuracy of disinformation detection is pivotal, especially in a digital era swamped with misleading content. The bar graph dedicated to the confidence levels of flagged posts provides an insightful overview of the certainty associated with the identification of each disinformation post in figure 3. A higher confidence level indicates a stronger likelihood that the post is genuinely misleading, while a lower level suggests ambiguity. By analyzing these levels, stakeholders can gain a clearer perspective on the reliability of the disinformation detection mechanism in place and make informed decisions about potential interventions or further verification procedures.

Figure 4. Types of disinformation



The complex nature of disinformation makes it essential to categorize and understand its different forms. By breaking down the flagged posts according to the theme of disinformation, the graph offers a structured view into the prevalent narratives or tactics leveraged by misinformers shown in figure 4. Some themes might be politically driven, others might be based on societal rumors, while some could be commercially motivated falsehoods. Identifying the most common types of disinformation not only helps in targeting specific countermeasures but also aids in educating the user base about prevalent misleading themes, thereby empowering them to be more vigilant.

Figure 5. Users with the highest number of disinformation posts



In our assessment of disinformation dynamics within the dataset, several key patterns emerged. Out of the 5,000 posts profiled (figure 5), 942 were flagged as disinformation, accounting for 18.84% of the content. The system's confidence in these disinformation flags varied, with an average confidence of 74.08%. The spread of confidence levels was substantial, as indicated by a standard deviation of 14.58%, with certain flags having confidence as low as 50.02% and others nearing certainty at 99.95%. Examining the nature of the disinformation, political themes were predominant, making up 381 of the flagged posts. This was followed by medical and financial themes, which accounted for 296 and 187 posts, respectively. Intriguingly, 78 posts were categorized under "None", suggesting potential false positives or categorization challenges in the flagging process.

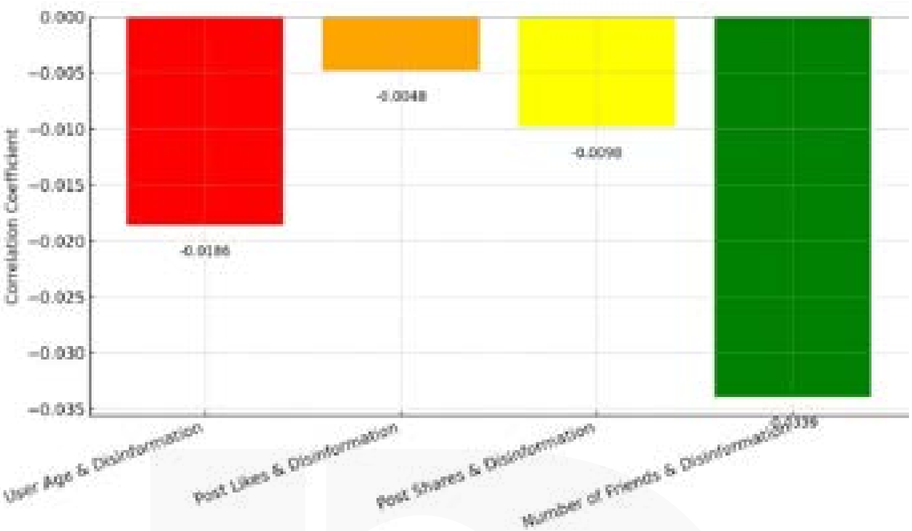
Further scrutiny into user behavior revealed that certain users were recurrent propagators of disinformation. Specifically, users 814, 546, 367, and 915 each had 5 posts flagged as disinformation, making them the top contributors of such content in this environment.

Correlation Analysis

In the pursuit of understanding the dynamics and potential drivers behind the spread of disinformation, a series of correlation analyses were conducted on our dataset. Preliminary findings, although intriguing, highlighted the complexity of the disinformation landscape.

The relationship between a user's age and their propensity to share disinformation yielded a correlation coefficient of −0.0186−0.0186, indicating a very weak inverse relationship. In simpler terms, older users were marginally less inclined to share disinformation, though the effect was virtually negligible. When investigating the interplay between post engagement metrics (likes and shares) and disinformation, the results mirrored the aforementioned patterns: both likes (−0.0048−0.0048) and shares (−0.0098−0.0098) demonstrated weak negative correlations with disinformation. This subtle trend suggests that posts with higher engagement were slightly less likely to contain disinformation. Lastly, analyzing the number of friends against disinformation sharing tendencies produced a correlation of −0.0339−0.0339. While this indicates that users with larger networks were somewhat less prone to share disinformation, the correlation remained weak. These findings collectively underscore that while certain factors may influence disinformation dynamics, they do so subtly, emphasizing the multifaceted nature of the issue.

Figure 6. Correlation analysis between various factors and disinformation



In a recent examination of the factors contributing to the prevalence of disinformation, a bar graph was employed to depict the correlation analysis between various determinants and the manifestation of misleading content in figure 6. With the y-axis quantifying the correlation coefficient, it is imperative to note that coefficients proximate to 0 symbolize weak correlations. In contrast, coefficients nearing -1 or 1 signify robust negative or positive correlations, respectively. Upon analysis, the data elucidated that the correlation between 'User Age' and 'Disinformation' exhibited the most prominent negative relationship, even though it remains within a weak correlation range. Similarly, factors such as 'Post Likes', 'Post Shares', and the 'Number of Friends' also presented weak negative correlations with disinformation. This analytical representation accentuates the nuanced relationships between various user attributes and their potential predispositions towards disseminating or encountering disinformation.

RESULTS

Our exploration into the dataset, crafted to resemble a social media platform akin to Facebook, revealed intriguing patterns about the prevalence and nature of disinformation. Out of the 5,000 posts assessed, a non-trivial 18.84% were flagged as disinformation. Delving deeper into the nature of these flagged posts, political themes emerged as the predominant source of disinformation, accounting for 381 of the flagged content. Medical and financial themes followed, with 296 and 187 posts, respectively. Notably, 78 posts were ambiguously categorized under "None", suggesting potential challenges in the categorization mechanism, potentially pointing towards false positives.

The dataset also shed light on the confidence levels associated with these disinformation flags. The average confidence stood at 74.08%, with a fairly wide distribution, as indicated by a standard deviation of 14.58%. This range, spanning from a minimum of 50.02% to a near-certain 99.95%, underscores the variability in the system's assurance in flagging content as disinformation.

An ancillary objective was to discern any discernible patterns between user demographics or behaviors and the propensity to share or engage with disinformation. Age, often hypothesized as a potential factor, displayed a weak negative correlation of −0.0186−0.0186 with disinformation propensity. This suggests that while older users were marginally less inclined to share misleading content, the relationship was almost negligible. The post engagement metrics, specifically likes (−0.0048−0.0048) and shares (−0.0098−0.0098), also demonstrated weak negative correlations with disinformation. These subtle correlations hint that higher engagement does not necessarily correlate with a higher likelihood of a post being misleading. Furthermore, the size of a user's network, represented by their number of friends, exhibited a weak negative correlation of −0.0339−0.0339 with disinformation sharing tendencies.

In summation, while the dataset highlighted a significant presence of disinformation, the correlations with age, engagement metrics, and network size were weak. This underscores the multifaceted nature of disinformation propagation, suggesting that simple demographics or engagement metrics might not be substantial drivers. The nuances and complexities unveiled in this analysis emphasize the need for a multifactorial approach to understanding and combating disinformation on digital platforms.

CONCLUSION

The proliferation of disinformation on digital platforms presents an exigent challenge in today's information age. Our analysis of a dataset, mirroring the dynamics of a platform analogous to Facebook, offers salient insights into the intricacies of this modern quandary. While nearly one-fifth of the posts in our dataset were flagged as disinformation, the factors traditionally hypothesized to influence such dissemination—such as age, post engagement, and network size—showed only weak correlations. This suggests that the roots of disinformation spread are more complex than often assumed, and simplistic demographic or engagement-based models might be insufficient to capture its full breadth.

Indeed, the findings underscore the insidious nature of disinformation. Its prevalence, coupled with the weak correlations with user attributes, alludes to the possibility of deeper, perhaps more covert, mechanisms at play. The digital landscape, with its vastness and the anonymity it can afford, provides fertile ground for the propagation of misleading narratives. This research accentuates the necessity for a multi-pronged approach to combat disinformation. While algorithms and flagging mechanisms are vital, they must be complemented by user education, critical thinking promotion, and perhaps even digital literacy programs.

In closing, the challenge of disinformation is multifaceted, and its impact on societal discourse, decision-making, and even democratic processes cannot be understated. While this study provides a snapshot of the problem within a simulated environment, it is emblematic of a larger, real-world issue. As we move forward in the digital age, a concerted effort from technologists, policymakers, educators, and users will be paramount in mitigating the effects of disinformation and ensuring the veracity of information on digital platforms.

REFERENCES

Collins, B., Hoang, D. T., Nguyen, N. T., Hwang, D. (2021). Trends in Combating Fake News on Social Media–A Survey, *Journal of Information and Telecommunication*, 5(2), 247-266.

Ertem Y. E. (2019), Sosyal Medyada Dezenformasyon, Yüksek Lisans Tezi, Marmara Üniversitesi.

Etike, Ş. (2023). ChatGPT ile Baş Etmek: Emek ve Eşitlik Odaklı Bir Çerçevenin Gerekliliği, *Emek Araştırma Dergisi (GEAD)*, Cilt 14, Sayı 23, Haziran 2023, 115-132.

Fallis, D. (2014). A Functional Analysis of Disinformation. *iConference*, Berlin, 4-7 March, 621-627.

Kara, N. (2022). Mezenformasyon ve Dezenformasyon Faaliyetleri, Sektörel Riskler ve İletişim Teknolojileri, *Denetişim Dergisi*, 26, 44-51, 2022.

Karakoç, E. & Zeybek, B., (2022). Görmek İnanmaya Yeter Mi? Görsel Dezenformasyonun Ayırt Edici Biçimi Olarak Siyasi Deepfake İçerikler, *Marmara Üniversitesi Öneri Dergisi*, Cilt 17, Sayı 57.

Karlova, N. A. & Fisher, K. E. (2013). Plz RT: A Social Diffusion Model of 75 Misinformation and Disinformation for Understanding Human Information Behaviour. *Information Research*, 18, 1-17.

Kırık, A. M. & Özkoçak, V. (2023). Medya ve İletişim Bağlamında Yapay Zekâ Tarihi ve Teknolojisi: Chatgpt ve Deepfake İle Gelen Dijital Dönüşüm, *Karadeniz Uluslararası Bilimsel Dergi,* (58), 73-99.

Koçyiğit, A. & Koçyiğit, M. (2023). Dijital Çağda Sosyal Medyada Dezenformasyonla Mücadele, *Dijital Çağda Medya Araştırmaları*, (177-207).

Kurnaz, A. (2022). Dijital Siyasetin Yükselişi ve Yapay Zekâ (The Rise of Digital Politics and Artificial Intelligence), (December 27, 2022). Siyaset, *Kamu Yönetimi ve Uluslararası İlişkiler Bağlamında Yapay Zeka Tartışmaları*, Ekin Yayınevi.

Long, S. H. & Hamzah, M. P. B. (2021). Fake News Detection. *In Computational Science and Technology: 7th ICCST 2020*, Pattaya, Thailand, 29–30 August, 2020 (pp. 295-303). Springer Singapore.

McKinney, W. & Team, P. D. (2015). Pa*ndas-Powerful python data analysis toolkit. Pandas—Powerful Python Data Analysis Toolkit*, 1625.

Özbay, F. A. & Alataş, B. (2020). Çevrimiçi Sosyal Medyada Sahte Haber Tespiti, *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 11(1), 91-103.

Özkoçak, V. & Kırık, M. (2023). Seçim ve Propaganda Süreçlerinde Yapay Zekâ, Büyük Veri ve Algoritmaların Etkisi: 14 Mayıs 2023 Türkiye Genel Seçimleri Örneği, *Social Sciences Research Journal*, 12 (3), 412-428.

Sahoo, S. R. & Gupta, B. B. (2021). Multiple Features Based Approach for Automatic Fake News Detection on Social Networks Using Deep Learning, *Applied Soft Computing*, 100, 106983.

Shu, K. (2022). Combating disinformation on social media: A computational perspective. *BenchCouncil Transactions on Benchmarks*, Standards and Evaluations, 2(1), 100035.

Sreekumar D. & Chitturi B., (2020). Deep Neural Approach to Fake-News Identification. *Procedia Computer Science*, 167: 2236–43, https://doi.org/https://doi.org/10.1016/j.procs.2020.03.27 6.

Taşkın, S. G., Küçüksille, E. U., Topal, K. (2021). Twitter Üzerinde Türkçe Sahte Haber Tespiti, *Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 23(1), 151-172.

Toğaçar, M., Eşidir, K., A. Ergen, B. (2022). Yapay Zekâ Tabanlı Doğal Dil İşleme Yaklaşımını Kullanarak İnternet Ortamında Yayınlanmış Sahte Haberlerin Tespiti, *Zeki Sistemler Teori ve Uygulamaları Dergisi*, 5(1) (2022) 1-8.

Tok, İ. (2020). Hakikat ötesi (post-truth) çağda yeni medyada yalan/sahte haberle mücadele. (Eds. Ş. Sağıroğlu, H. İbrahim Bülbül, A. Kılıç, M. Küçükali), *Dijital Okuryazarlık: Araçlar, Metodolojiler, Uygulamalar ve Öneriler*, Ankara: Nobel Yayınevi.

Turan, C. (2015). Açıklığın Yanılsaması: Dezenformasyon Çağımızın Kitle İmha Silahı mı, *Akademik Bilişim Konferansı*, Yayın. 109-115.

Uluç, D. (2003). İlk Körfez Savaşı'nı Nayirah Başlattı", Hürriyet Gazetesi, 23 Mart 2003.

Wardle, C. & Derakhshan, H. (2017). Information Disorder: Toward an İnterdisciplinary Framework For Research And Policymaking, Council Of Europe Report DGI.

Yurdigül, Y. & Yıldırım, A. (2021). Gerçeklik Algısına Bir Müdahale Aracı Olarak Sentetik Medya Teknolojileri, *İletişim ve Diplomasi*, 5, 105-121.

Yazar katkı düzeyi/Author contributions:

Makale Tasarımı: Y.S. Balcıoğlu. Literatür Taraması: B. Doğan. Veri Toplama ve Analiz: Y.S. Balcıoğlu. Sonuç: Y.S. Balcıoğlu. Son Okuma, Kontrol ve Sorumluluk: Y.S. Balcıoğlu

Design of Article: Y.S. Balcıoğlu. Literature review: Y.S. Balcıoğlu. Data acquisition and analysis: Y.S. Balcıoğlu. Final reading, checking and approval: Y.S. Balcıoğlu.

Hakem değerlendirmesi/Peer review:

Dış bağımsız/Externally peer reviewed

Çıkar çatışması/Conflict of interest:

Yazarlar çıkar çatışması bildirmemiştir/The authors have no conflict of interest to declare

Finansal destek/Grant support:

Yazarlar bu makalede finansal destek almadığını beyan etmiştir/The authors declared that this article has received no financial support.