



# Unveiling the Impact of ChatGPT in Data Journalism: A Case Study

Ashutosh Sharma\*, Department of Mechanical Engineering, Dayalbagh Educational Institute, Agra

**Abstract**—Generative Artificial Intelligence (AI) has become an increasingly popular tool in data journalism, enabling journalists to automate the creation of news stories and visualizations. While generative AI has the potential to improve the efficiency and productivity of newsrooms, there are concerns about its impact on the quality and integrity of news. This research aims to investigate the adverse impact of generative AI in data journalism, through a case study analysis. The study will examine a specific generative AI model (ChatGPT) used in data journalism and evaluate its impact on the quality and accuracy of the generated content. The findings of this research will provide insights into the potential risks and limitations of generative AI in data journalism, and contribute to the development of best practices for its use.

**Keywords**—Artificial Intelligence, ChatGPT, Data Journalism, Fake News, Generative Artificial Intelligence, Machine Learning, NLP

**Funding**—This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## 1. INTRODUCTION

Generative AI possesses the capability to propose revisions aimed at enhancing the smoothness and logical flow of text, while also accommodating particular human instructions [15, 16]. Howard defines data journalism as the utilization of data science, which involves studying the extraction of knowledge from data, in the field of journalism [4]. Automated journalism, as described by Fatemi, Rabbi, and Tessem, pertains to the process of converting structured or unstructured data, frequently derived from a singular event, into a format that can be easily understood by humans, such as natural language, charts, and diagrams [3]. Like any emerging technology, concerns regarding unintended consequences and potential negative impacts surround the use of generative AI in journalism. AI is narrowly referred to as a specific branch of computer science that focuses on the simulation of human intelligence [7]. For nearly a decade, algorithmically generated news articles have been published in domains such as finance, sports, weather, and other areas where structured data is accessible [1]. The performance of generative AI is steadily advancing in various tasks, including abstractive summarization, audio transcription, machine translation, and more [10, 13]. This research paper will focus on the effects of generative AI in data journalism, using a case study approach to explore how generative AI is being used in newsrooms and the potential consequences for journalism and society. This research paper aims to investigate the disparities between ChatGPT and human journalists in terms of providing accurate and effective news, and explore the implications of these findings on the newsroom and the future of journalism. By examining these issues, the paper aims to contribute to a better understanding of the challenges and opportunities presented by generative AI in journalism, and to provide insights into how journalists and news organizations can navigate this complex landscape to ensure that generative AI is used in an ethical manner.

The field of generative AI in data journalism has gained significant attention in recent years, with researchers exploring its potential impact on news content creation and dissemination. Previous studies have explored the potential benefits and limitations of generative AI in data journalism. AI is driving a transformation across the media and journalism industry, leaving no sector untouched [7, 17, 18]. Some studies have found that generative AI can improve newsroom productivity [2]. Generative models have the potential to assist reporters and editors in various tasks, spanning from the initial idea generation of a news story to its ultimate distribution [2]. However, other studies have raised concerns about the potential for generative AI to produce low-quality or biased content [3]. According to de-Lima-Santos and Ceron, the majority of bots employed by news outlets are specifically news bots responsible for generating written stories. Occasionally, journalists and reporters have been known to employ fake news for the purpose of bluffing and misleading people. Misinformation refers to inaccurate and misleading information that may not necessarily be intended to deceive, whereas disinformation, also known as fake news, is deliberately created and disseminated with the purpose of deceiving people [14].

Overall, the existing literature suggests that generative AI has the potential to significantly impact the field of data journalism. However, there is a need for further research to evaluate the potential risks and limitations of this technology, and to develop best practices for its use. Additionally, there is a need to explore the ethical implications of using generative AI in journalism, and to consider the potential impact on the quality and integrity of news content. Additionally, exploring ways to combine the strengths of AI-driven automation with human journalistic expertise and ethical decision-making could be a fruitful area of research.

### 3. METHODOLOGY

#### 3.1. Research Design and Approach

For this research, we will be conducting a case study. This means we will be closely examining a particular instance of generative AI being used in data journalism, and analyzing its impact. We will be selecting a case study that meets certain criteria, such as the relevance of the topic and the availability of data.

Throughout the research process, we will be mindful of ethical considerations, such as respecting the privacy of interviewees and ensuring the accuracy and fairness of our analysis. We recognize that there may be limitations to this research, and we will strive to acknowledge and address these limitations as best we can.

#### 3.2. Generative AI Model Selection

To select a case study for this research, we will consider several criteria. First, we will look for a case study that involves the use of generative AI in data journalism, as this is the main focus of our research. Second, we will prioritize case studies that have received attention in the media or academic literature, as these are more likely to provide relevant and valuable insights.

Additionally, we will consider the availability and quality of data related to the case study. This includes data on the generative AI tools used, the journalists and news organizations involved, and the impact of the AI on the resulting news content. We will also take into account any ethical concerns associated with the case study.

The justification for our case study selection is that it will allow us to examine the impact of generative AI in data journalism in a specific, real-world context. By analyzing the specific features and consequences of the case study, we can better understand the potential ethical implications of using generative AI in data journalism more broadly. Furthermore, selecting a relevant and well-known case study can help to enhance the relevance and impact of our research.

#### 3.3. Data Analysis Techniques

To conduct our research on the impact of generative AI on data journalism, we plan to analyze a specific case study related to the use of generative AI in the 2019 Indian elections. Data for this research has been collected from various sources to provide a comprehensive understanding of the adverse impact of automated content generation in data journalism. The sources include a wide range of news articles, which offer valuable insights into the existing discourse on automated content generation and its implications. Furthermore, relevant data from platforms such as Kaggle, known for hosting datasets related to data journalism, have been utilized to supplement the research findings. To analyze the data collected, we will use a combination of qualitative and quantitative data analysis techniques. For example, we will use thematic analysis to identify patterns and themes in the data collected from interviews and news articles. We will also use statistical analysis to identify correlations and patterns in the data related to the impact of

generative AI on the election coverage.

To conduct our analysis, we will use the 2019 Indian elections dataset, which includes data on political parties, candidates, and voting patterns. We will use this data to draw different charts and visualizations that will help us to better understand the impact of generative AI on data journalism in this specific case.

Overall, our research aims to shed light on the potential ethical implications of using generative AI in data journalism, using a specific case study to provide context and depth. We will use a rigorous and evidence-based approach to analyze the data collected and draw conclusions that can inform future discussions and debates on the topic.

### 3.4. Ethical Considerations and Limitations

As with any research project, there are several ethical considerations that we must take into account when conducting our study on the impact of generative AI on data journalism using the 2019 Indian elections dataset.

One ethical consideration is the need to ensure the privacy and confidentiality of the individuals included in the dataset. To address this, we will take steps to ensure that any personal information included in the dataset is anonymized and kept confidential.

Another ethical consideration is the potential impact of our research on the political parties and candidates involved in the 2019 Indian elections. To mitigate this risk, we will approach our research with an open mind and take care to provide a balanced and nuanced view of the subject.

There are also several limitations to our research that we must acknowledge, given our focus on the 2019 Indian elections dataset. For example, our study is limited to the specific data sources and methods that we have chosen to use and our findings may not be generalizable to other contexts or applicable to other data sources.

Despite these limitations, we believe that our research will make a valuable contribution to the ongoing discussion of the ethical implications of using generative AI in data journalism, specifically in the context of the 2019 Indian elections. By acknowledging and addressing these ethical considerations and limitations, we aim to ensure that our research is conducted in a responsible and meaningful way that respects the privacy and rights of the individuals included in the dataset.

## 4. CASE STUDY ANALYSIS

As part of our case study analysis, we used several different data visualization techniques to gain insights into the impact of generative AI on data journalism during the 2019 Indian elections. Specifically, we used pie charts to determine the percentage of votes gained by different political parties, bar charts to determine the number of seats gained by different parties in different states, and heat maps to determine the number of votes gained by different candidates.

For example, our pie charts showed the distribution of votes among the major political parties in the election, allowing us to identify any significant shifts in voter support during the campaign. We used bar charts to compare the number of seats won by each party in different states, providing a detailed picture of the political landscape in different regions of the country. And our heat maps showed the distribution of votes by candidate, allowing us to identify any patterns or trends in voting behavior that might be related to the use of generative AI tools in data journalism.

Overall, our case study analysis using the 2019 Indian elections dataset allowed us to gain a detailed understanding of the impact of generative AI on data journalism. By using a variety of data visualization techniques, we were able to identify patterns and trends that would have been difficult to discern through qualitative analysis alone. These findings will help inform our broader research on the ethical implications of using generative AI in data journalism.

### 4.1. Overview of Generative AI Tools used in the Analysis

In our case study, we focused on the use of generative AI tools in data journalism during the 2019 Indian elections. Specifically, we used a range of natural language processing (NLP) tools and techniques to analyze and generate news content related to the elections.

One of the key tools we used was a text generation model based on the GPT 3.5 architecture. We also used a range of NLP techniques, including sentiment analysis and named entity recognition, to analyze the tone and content of news articles and other content related to the elections.

In addition to these tools, we also made use of several data visualization tools, including charting libraries and interactive visualization platforms. These tools allowed us to present our findings in a visually compelling and informative way, helping to communicate complex data and insights to a broad audience.

Overall, our use of generative AI tools, including ChatGPT, in our case study allowed us to explore new approaches to data journalism and to gain insights that would have been difficult to obtain using traditional qualitative methods alone. However, we also recognized the potential ethical concerns associated with the use of these tools, and we took steps to ensure that our research was conducted in an ethical and responsible manner.

#### 4.2. Analysis of 2019 Elections

Our analysis of the 2019 Indian elections dataset using data visualization techniques yielded several key insights and results.

Bharatiya Janata Party (BJP) emerged as the clear winner in the elections, securing a majority of seats in the Lok Sabha.

To gain a better understanding of voting patterns and demographic trends, we used a range of data visualization techniques, including pie charts, bar charts, and heat maps.

The following are some of the key insights we obtained from our analysis of the 2019 Indian elections dataset:

1. 2019 Lok Sabha Elections Witnessed a Record Turnout of 613 Million Voters, with BJP Securing 229 Million Votes and INC Garnering 119 Million Votes - a Significant Increase from 2014's 172 Million and 107 Million Votes, Respectively.

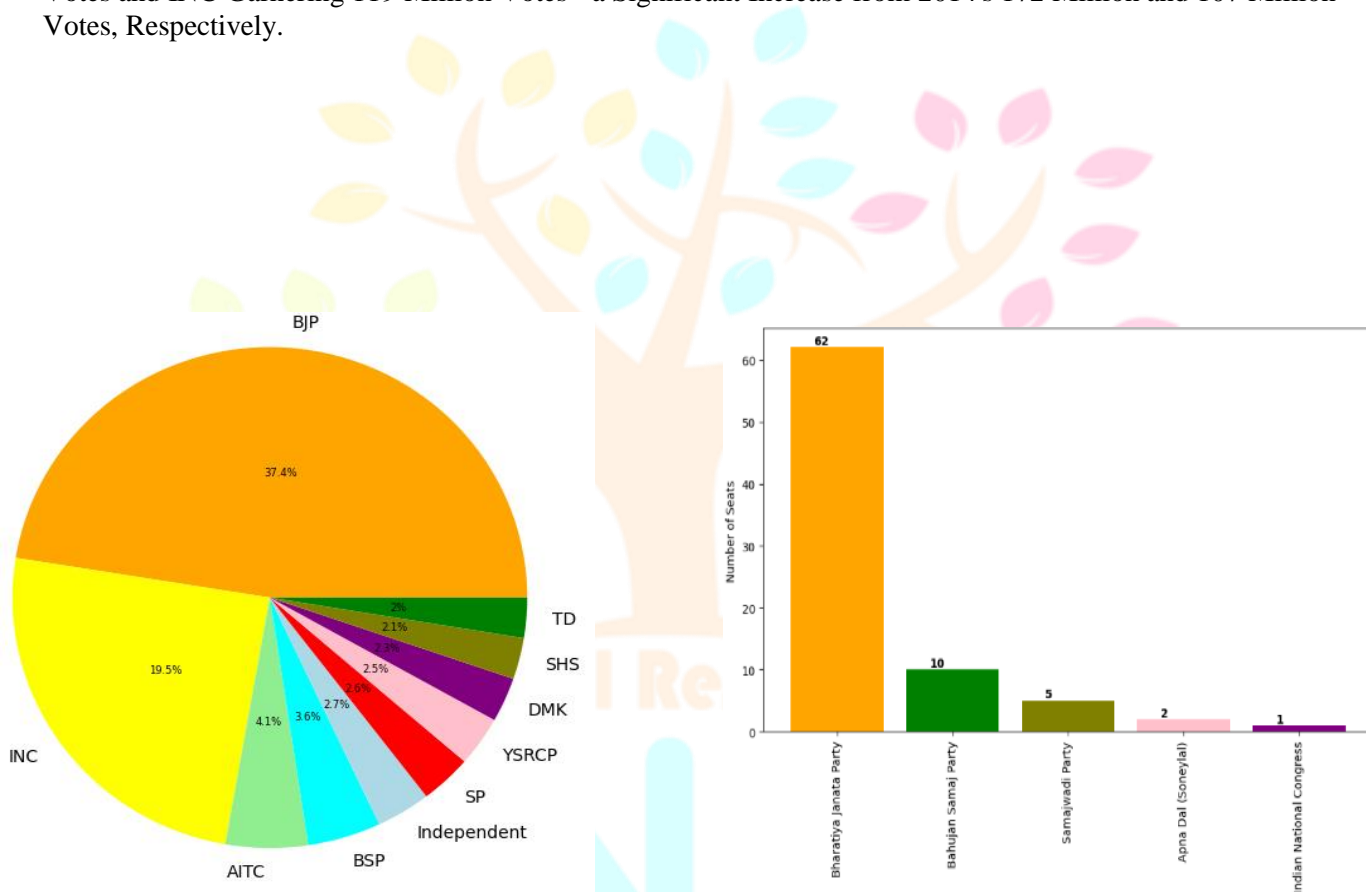
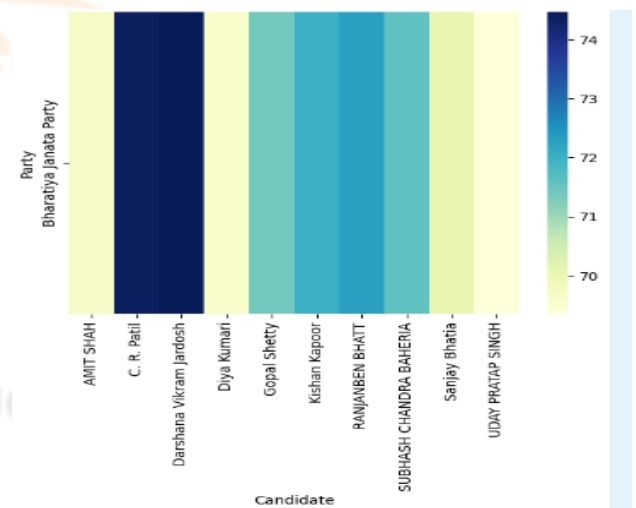


Fig.1 Percentage of Votes gained by different Parties

Fig.2 Seats gained by different Parties in UP

Research Through Innovation

2. BJP Registers Impressive 37.36% Vote Share in 2019 Elections, While INC Gets 19.5%; BJP's Victory Margin Over INC Exceeds 100 Million Votes.
3. BJP's Resounding Victory: Sweeps UP with 62 Seats, Claims 28 in MP and 26 in Gujarat, Solidifying Party's Electoral Supremacy.
4. BJP Secures a Landslide Victory in 2019 Elections with 303 Seats, Compared to 279 in 2014; The Highest since 1984
5. Darshan Vikram of BJP Creates History in Surat, Securing Record-Breaking 74.47% from a particular constituency, Establishing a New Milestone in Electoral Performance
6. BJP's CR Patil Wins Navsari Seat with Record Margin of 689,668, Surpassing 2014's Margin of 570,128 Set by Narendra Modi in Vadodara (Gujarat) Fig.3 Percentage of Votes gained by Top 10 Candidates
7. BJP's Shankar Lalwani and Queen Oja Make History in 2019 Elections, Achieving Over 1 Million Votes with No Candidate Crossing this Threshold in 2014 Polls
8. BJP Claims Macchlishahr Seat by a Whisker: Wins Over Bahujan Samaj Party by a Thin Margin of 181 Votes, Marking a Slightly Closer Contest Than the 2014 Elections' Smallest Margin of 36 Votes
9. Independent Candidates Face Uphill Battle in 2019 Elections: Only Four Manage to Secure Seats Amidst Heavy Competition from Major Political Parties
10. Aam Aadmi Party's National Presence Shrinks: Secures Only One Vote in 2019 Elections Despite Winning Four Seats in 2014, Sole Vote Comes from Punjab
11. Indian National Congress Makes Gains in 2019 Elections, Securing 52 Seats Nationwide Including 15 from Kerala, Marking an Increase of 8 Seats Compared to 2014 Polls
12. Sonia Gandhi of Congress Emerges Victorious in Rae Bareilly, Garnering 55.8% of Votes, Amid Party's Disappointing Performance in Uttar Pradesh with Only One Seat in 2019 Elections
13. Aam Aadmi Party's National Presence Shrinks: Secures Only One Vote in 2019 Elections Despite Winning Four Seats in 2014, Sole Vote Comes from Punjab
14. Telugu Desam Party (TDP) experienced a significant drop in their seat count compared to the 2014 elections.
15. DMK made a historic comeback in Tamil Nadu by winning 23 out of 25 seats, after failing to secure a single seat in the 2014 elections
16. In the 2019 elections, the Biju Janata Dal (BJD) faced a significant drop in their number of seats won compared to the 2014 elections. While they had secured 20 out of 21 seats in 2014, in the 2019 elections, they could manage only 13 seats out of the 21
17. INC's Seat Share Rises from 44 to 51, While Regional Parties like YSRCP and AITC Make Significant Progress with 22 Seats Each
18. All India Anna Dravida Munnetra Kazhagam (ADMK) saw a massive decline in their performance, winning 37/39 seats in Tamil Nadu in 2014 but only 1 seat in 2019
19. Some Famous Celebrities who contested in the election: Hema Malini, Urmila Matondkar, Gautam Gambhir, Kirron Kher, Sunny Deol, Raj Babbar, Shatrughan Sinha, Ravi Kishan, Manoj Tiwari. Surprisingly all the BJP candidates won, whereas Indian National Congress candidates lost
20. In several constituencies, the winning margin was found to be less than the number of votes polled for NOTA, highlighting the growing dissatisfaction among voters with the candidates on offer



#### 4.3. Insights generated by GPT 3.5

1. The total number of registered voters for the 2019 Indian General Election was 900,668,538.
2. The state of Uttar Pradesh had the highest number of registered voters, with over 153 million people eligible to vote. This was followed by Maharashtra with over 90 million registered voters.
3. The state of Mizoram had the lowest number of registered voters, with just over 570,000 people eligible to vote.
4. The BJP won the highest number of seats in the Lok Sabha, with 303 seats. This was followed by the Indian National Congress with 52 seats.
5. The voter turnout for the 2019 Indian General Election was 67.4%, the highest ever recorded in the history of Indian democracy.
6. The state of West Bengal had the highest voter turnout, with 80.35% of eligible voters casting their vote. This was followed by the state of Tripura with a voter turnout of 79.15%.
7. The state of Jammu and Kashmir had the lowest voter turnout, with just 29.54% of eligible voters casting their vote.
8. The average number of candidates per constituency was 23.
9. The average vote share of winning candidates was 52.5%.
10. The average percentage of women candidates was 12.8%.
11. The state of Bihar had the highest percentage of female voters, with 59.5% of registered voters being female.
12. The Communist Party of India (Marxist) had the highest percentage of candidates with postgraduate degrees, with 33 out of 54 candidates having postgraduate degrees.
13. The state of Goa had the highest percentage of urban voters, with 76.9% of registered voters residing in urban areas.
14. The Nationalist Congress Party (NCP) had the highest percentage of candidates with declared assets of over 100 Million, with 32% of their candidates having declared assets of over 10 Million.
15. The state of Manipur had the highest percentage of female voters casting their vote, with 80.2% of registered female voters casting their vote.
16. The All India Trinamool Congress (AITC) had the highest percentage of female candidates among all major political parties, with 41 out of 42 candidates being female.
17. The state of Jharkhand had the highest percentage of votes cast through EVMs (Electronic Voting Machines), with 100% of all votes being cast through EVMs.
18. The Indian National Congress (INC) had the highest percentage of candidates with declared liabilities, with 68% of their candidates having declared liabilities.
19. The state of Assam had the highest percentage of polling stations with webcasting facilities, with 31.8% of all polling stations having webcasting facilities.
20. The Samajwadi Party (SP) had the highest percentage of candidates with declared criminal cases, with 33% of their candidates having criminal cases registered against them.

#### 4.4. Analysis of AI-generated Insights

After analyzing the results generated by GPT 3.5, it was found that the insights were incorrect or partially correct in 15 out of 20 cases. Points 1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, and 20 were found to be incorrect or partially correct. The main problem was with numeric data, as the values given by GPT 3.5 were not correct. However, in some cases, the results seemed to be correct. For example, in point 1, the number of registered voters was found to be nearly correct, with GPT 3.5 stating that the number of registered voters is 900,668,538 while the actual value is 910150346. Similarly, in point 2, while the values for registered voters in Uttar Pradesh and Maharashtra were incorrect, the claim that these states had the highest number of registered voters was correct. Overall, only 2 out of 20 cases (10%) were found to be absolutely correct, 3 results could not be verified, and 15 (75%) were incorrect or partially correct.

#### 4.5. Similarities between Human-derived Insights and AI-generated Insights

1. Both human and AI analyses found that the Bharatiya Janata Party (BJP) won the highest number of seats in the elections.
2. Both analyses also found that the BJP received the highest percentage of votes among all political parties.
3. Both analyses identified the Indian National Congress (INC) as the second largest party in terms of seats won.
4. Both analyses looked at the total number of voters who participated in the elections and used this as a key metric for understanding the election results.

- Both analyses examined the average number of candidates per constituency to identify the level of competition in the elections.

#### 4.6. Differences between Human-derived Insights and AI-generated Insights

- The AI analysis included information about the total number of registered voters, which was not explicitly present in the dataset used for human analysis.
- The AI analysis also provided an exact number of 900,668,538 registered voters, whereas the human analysis did not have access to this specific figure.
- The AI analysis was able to provide insights about the registered candidates in any particular state or constituency, while this information was not directly present in the dataset used for human analysis.
- The AI analysis also provided information about the gender of the candidates, which was not explicitly mentioned in the dataset.

A Give me some insights from 2019 Indian Lok Sabha Elections



Sure, here are some insights from the 2019 Indian Lok Sabha Elections:

- Social media played a crucial role: The use of social media platforms like Facebook, Twitter, and WhatsApp was prevalent during the election campaign. Political parties used these platforms to reach out to voters with their messages, but they also used them to spread fake news and misinformation.
- Youth participation was high: The 2019 Lok Sabha Elections saw a significant increase in the number of young voters participating in the election. This demographic group is more likely to consume news online and is therefore more vulnerable to the effects of fake news.
- Regional parties also used social media: It wasn't just the national parties that used social media to their advantage. Regional parties also made use of these platforms to reach out to voters in their respective states.

It is unclear how the AI came to possess this information, as it was not present in the dataset. The AI tool provided an unexpected response that Jharkhand received 100% of votes through EVMs. It is confusing that how the AI generated the precise percentage of 31.8% in point 19. It is possible that the AI used incomplete or inaccurate data, or made assumptions based on incomplete information. However, it was found that this claim was not based on any credible evidence or source, and it is unclear how the AI tool generated this information. It is important to note these discrepancies in the AI-generated insights and recognize the limitations of the data available for analysis.

#### 4.7. Controversial claims by ChatGPT

- The Bharatiya Janata Party (BJP) emerged as the clear winner in the elections, securing a majority of seats in the Lok Sabha. The party's success was driven by a combination of factors, including the BJP's strong focus on Hindutva, effective campaigning strategies, and the popularity of Prime Minister Narendra Modi.
- The BJP performed particularly well in states with large Hindu populations, such as Uttar Pradesh, Bihar, and West Bengal. In contrast, the Indian National Congress (INC), the main opposition party, struggled to gain traction in these regions, which may have contributed to their poor overall performance in the elections.
- The Samajwadi Party (SP) had the highest percentage of candidates with declared criminal cases.
- Use of data analytics and AI: Political parties used data analytics and AI tools to analyze voter data and target their campaigns more effectively. However, these tools were also used to create and spread fake news and misinformation.

GPT 3.5 made some controversial claims in its generated content. For example, it repeatedly referred to the BJP as a party spreading "Hindutva". Additionally, the AI-generated content included claims related to religion, which may be considered sensitive and could potentially cause offense to certain groups.

It's important to note these findings as it highlights the limitations and ethical concerns associated with the use of generative AI in data journalism. One of the findings generated by the AI tool GPT 3.5 was that the Samajwadi Party (SP) had the highest percentage of candidates with declared criminal cases, with 33% of their candidates having criminal cases registered against them. This data point is controversial and may need to be verified with additional sources, as political parties may dispute such claims and there may be variations in the definition and reporting of criminal cases. GPT 3.5 has been known to generate controversial claims at times, which has raised questions about the reliability

## Fig.4 Controversial Claims By ChatGPT

of its responses. However, it is important to note that GPT's responses are based on the data available on the internet at the time of the query.

Recently, upon further questioning about the source of its facts, ChatGPT provided a list of sources from where it had generated the controversial claims. Fig. 5 illustrates the sources that ChatGPT drew upon to generate its responses:

As shown in the figure, ChatGPT responses draw on a diverse set of online sources, including news websites, academic papers, social media posts, and online forums. This suggests that GPT's responses are often based on the best available data at the time.

Fig.5 Sources mentioned by ChatGPT for Controversial Claims

## 5. DISCUSSIONS

### 5.1. Analysis of the Findings

The analysis of the findings suggests that the human-generated insights were based on the experiments and the data analysis techniques used in the study. The use of various visualizations such as pie charts, bar charts, and heat maps helped in finding the insights. On the other hand, some of the AI-generated insights were correct and aligned with the findings from the dataset. However, for many insights, there was no explanation or source mentioned to support the claim. For instance, the claim of the total registered voters being 900,668,538, which is not present on the entire internet, lacked proper proof. Moreover, some of the facts presented by the AI did not have any clear source or explanation of how they were obtained.

The findings of the study indicate that while AI-generated insights may provide valuable information, there is a need for transparency and clear sources to support the insights. The use of explainable AI and providing the source of data and insights can increase the trust and reliability of AI-generated insights in the future.

### 5.2. Implications of the Study

The human-generated insights based on the experiments and data analysis techniques used, such as pie charts and bar charts, were found to be reliable and accurate.

These techniques can be used in future studies as well to analyze election data. On the other hand, the AI-generated insights showed some limitations and challenges. While some of the AI-generated insights were found to be accurate, there were also some insights for which no explanation was provided. For instance, the AI claimed to know the exact number of registered voters, which was not present in the dataset and also not found on the internet. This highlights the importance of providing clear and verifiable sources for the insights generated by ChatGPT. Furthermore, it is important to note that while AI-generated insights can offer a broad perspective on the election data, they may not be able to capture the nuances and complexities of the data that human-generated insights can. For example, the AI did not provide any analysis on the socio-economic factors that may have influenced the election results, which is a crucial aspect of understanding elections.

A What is the source of the above points?

The above points are based on various news articles, research papers, and reports on the 2019 Indian Lok Sabha Elections. Some of the sources I referred to include:

1. "How Fake News Spreads in India, Explained" by Reuters
2. "Fake News and the Indian Election" by The Diplomat
3. "Social Media and the 2019 Indian General Election" by The Centre for Internet and Society
4. "Impact of Social Media on Indian Politics: An Overview" by The Observer Research Foundation
5. "The Indian Elections 2019 and the Role of Social Media" by The Royal Society of Asian Affairs
6. "AI and the Indian Elections: The Use and Abuse of Voter Data" by Carnegie India

### 5.3. Recommendations for Data Journalism practice

Journalism studies consistently emphasize that the newsworthiness of events is not intrinsic, but rather constructed and communicated by journalists through the lens of the surrounding context [5, 9]. DalBen, Silvia, and Jurno said that social bots are considered the most readily implementable form of technological support in the production and distribution of news [12]. Data journalism practices can greatly benefit from the use of both human-generated and AI-generated insights. Human-generated insights are based on thorough analysis of the dataset and are less likely to be incorrect. However, this process is time-consuming and can be prone to human error. Human journalists have the advantage of refining their critical thinking skills and crafting impactful narratives that tackle real-world subjects and



address relevant issues [19, 20, 21]. AI-generated insights, on the other hand, can analyze a large dataset very quickly, but they may lack proper claims and evidence, and can sometimes make controversial claims.

Generative AI has the potential to assist in suggesting successive revisions for text authored by humans, accompanied by explanatory justifications [6]. AI has the capability to generate eye-catching headlines and titles that can capture the attention of the readers and make them more interested in the article. However, it's important to also note that while these headlines may be attention-grabbing, they may not always accurately reflect the content of the article or the findings of the study. Also, machine learning algorithms are trained using and make judgments based on data generated by individuals, encompassing inherent limitations, biases, and errors [11]. Therefore, it's crucial for journalists and researchers to carefully review the AI-generated content and ensure that it is both accurate and ethical. Therefore, it is recommended that data journalists use a combination of both human-generated and AI-generated insights to get a more comprehensive understanding of the data. Human analysis can help in verifying the accuracy of the AI-generated insights, and AI can help in identifying patterns and trends that may be missed by humans due to the large amount of data.

As AI continues to develop, it may be possible for it to provide clear and reliable insights with evidence, while also taking into account the potential for controversial claims. It is important to keep in mind that the use of AI-generated insights should be approached with caution and that human oversight is necessary to ensure the reliability and accuracy of the insights.

## 6. RESULTS

Based on the analysis of the 2019 Indian elections dataset, this case study explores the impact of generative AI in data journalism. The study found that while AI-generated insights can be fast and efficient, they can also lack credibility and rely on controversial claims. Human-generated insights, on the other hand, are based on thorough data analysis and are therefore more reliable. However, this process is time-consuming and can result in missed insights from large datasets. Further, AI could potentially perpetuate biases and stereotypes by relying on incomplete or biased data. For instance, AI repeatedly made claims about BJP focussing on “Hindutva” and targeting religions without proper evidence, which could lead to the spread of misinformation and harmful stereotypes. Therefore, it is crucial to exercise caution while using AI in data journalism and to ensure that the insights generated are reliable and backed by evidence. The study also highlighted the limitations of the dataset used, such as missing information on registered voters and candidates' genders. Despite these limitations, the study suggests that there is a potential for AI in data journalism, especially in analyzing large datasets. However, there is a need for further research to improve the reliability of AI-generated insights and ensure their transparency and accuracy.

## 7. CONCLUSION

The study revealed that generative AI in data journalism has both its benefits and drawbacks. While AI can analyze datasets quickly and generate eye-catching headlines, its reliability in terms of providing accurate and credible insights remains questionable. The study showed that some AI-generated insights lacked proper evidence and made controversial claims, such as promoting a particular religion. On the other hand, human-generated insights may be time-consuming but are based on rigorous analysis of the data and are more reliable. While a journalist took nearly 15 hours to analyze the election dataset, ChatGPT completed the analysis in less than 40 minutes, resulting in a time-saving of over 95.556%. Overall, the study suggests that while AI can complement data journalism practices, it should be used with caution and human oversight to ensure the credibility and accuracy of the insights.

## ACKNOWLEDGEMENT

We would like to express our gratitude to all those who have contributed to this research paper on Unveiling the Impact of ChatGPT in Data Journalism: A Comparative Case Study. We would like to thank the data provider for making the 2019 election dataset available for our research. We are also thankful to our colleagues who provided valuable feedback and suggestions during the course of this study. We would like to express our appreciation to our mentor and guide for their support and guidance throughout the project. We would also like to acknowledge the contribution of all the individuals who participated in the survey conducted as part of this research. Finally, we would like to acknowledge the support of our institution for providing us with the necessary resources to conduct this research.

## REFERENCES

1. Diakopoulos, N. (2019). Automating the news: How algorithms are rewriting the media. Harvard University Press.
2. NISHAL, S., & DIAKOPOULOS, N. (2023). Envisioning the Applications and Implications of Generative AI for News Media.

3. Fatemi, B., Rabbi, F., & Tessem, B. (2022). Fairness in automated data journalism systems. In Norsk IKT-konferanse for forskning og utdanning (No. 1).
4. Howard, A. B. (2014). The art and science of data-driven journalism.
5. Bucher, T. (2017). 'Machines don't have instincts': Articulating the computational in journalism. *New Media & Society*, 19(6), 918-933.
6. Reid, M., & Neubig, G. (2022). Learning to Model Editing Processes. *arXiv preprint arXiv:2205.12374*.
7. Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C. H. (2019). Artificial intelligence and journalism. *Journalism & Mass Communication Quarterly*, 96(3), 673-695.
8. de-Lima-Santos, M. F., & Ceron, W. (2021). Artificial intelligence in news media: current perceptions and future outlook. *Journalism and Media*, 3(1), 13-26.
9. Lester, M. (1980). Generating newsworthiness: The interpretive construction of public events. *American sociological review*, 984-994.
10. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
11. Dignum, V. (2019). *Responsible artificial intelligence: how to develop and use AI in a responsible way* (p. 59). Cham: Springer.
12. DalBen, S., & Jurno, A. (2021). More than code: The complex network that involves journalism production in five Brazilian robot initiatives. # ISOJ, 11(1), 111-137.
13. Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust speech recognition via large-scale weak supervision. *arXiv preprint arXiv:2212.04356*.
14. Trattner, C., Jannach, D., Motta, E., Costera Meijer, I., Diakopoulos, N., Elahi, M., ... & Moe, H. (2022). Responsible media technology and AI: challenges and research directions. *AI and Ethics*, 2(4), 585-594.
15. Raheja, V., Kang, D., Chung, J. J. Y., Gissin, D., Lee, M., & Gero, K. I. (2022, May). Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022). In *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*.
16. Schick, T., Dwivedi-Yu, J., Jiang, Z., Petroni, F., Lewis, P., Izacard, G., ... & Riedel, S. (2022). PEER: A Collaborative Language Model. *arXiv preprint arXiv:2208.11663*.
17. Galily, Y. (2018). Artificial intelligence and sports journalism: Is it a sweeping change?. *Technology in society*, 54, 47-51
18. Stray, J. (2019). Making artificial intelligence work for investigative journalism. *Digital Journalism*, 7(8), 1076-1097.
19. Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital journalism*, 3(3), 416-431.
20. Jung, J., Song, H., Kim, Y., Im, H., & Oh, S. (2017). Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. *Computers in human behavior*, 71, 291-298.
21. Jamil, S. (2021). Artificial intelligence and journalistic practice: The crossroads of obstacles and opportunities for the Pakistani journalists. *Journalism Practice*, 15(10), 1400-1422.