

Seeing Through Machines: Leveraging AI for Enhanced and Automated Data Storytelling¹

Swathi Chundru

*Motivity labs INC
QA Automation Engineer
Texas, USA*

Received: 29 July 2023; Accepted: 27 September 2023; Published: 21 October 2023

ABSTRACT

The capacity to convert intricate datasets into engaging stories has grown in significance in the big data era. Data storytelling presents data insights in an understandable, captivating, and useful way by fusing narrative techniques with data visualization. Nevertheless, the laborious manual procedures that are commonly employed in data storytelling are frequently constrained by human cognitive capacities. The transformational potential of artificial intelligence (AI) to improve and automate data storytelling is examined in this research article. Artificial intelligence (AI) can automate the extraction of insights, recognize patterns, and provide narratives that are accurate and insightful by integrating cutting-edge machine learning algorithms with natural language processing. This work explores the approaches used in AI-driven data storytelling, evaluates case examples from a range of sectors, and talks about the difficulties and moral issues raised by this developing discipline. According to the research, AI has the power to completely transform data storytelling and increase its effectiveness, scalability, and impact.

INTRODUCTION

The skill of turning data into a tale that effectively and meaningfully conveys information is known as "data storytelling." Data storytelling is a crucial tool for decision-makers in industry, research, and government who depend on data-driven insights to inform their strategies and policies. Conventional data storytelling is a labor-intensive approach that requires human analysts to assess data and create narratives. It calls for both a thorough comprehension of the material and the capacity to effectively express complicated insights.

The constraints of human data storytelling have become evident as the amount, pace, and variety of data have increased dramatically in recent years. The sheer volume of data being generated makes it difficult for human analysts to keep up, which can cause delays in the creation of insights and possibly cause them to miss important patterns. Moreover, the intricacy of contemporary data frequently demands sophisticated analytical methods that surpass the capabilities of conventional approaches.

These problems can be solved by artificial intelligence (AI), which automates some data narrative processes. AI has the capacity to evaluate enormous volumes of data at previously unheard-of rates, revealing insights that human analysts might overlook. Furthermore, narratives targeted to certain audiences can be produced by AI-driven techniques, guaranteeing that insights are accurate, relevant, and captivating.

This article investigates how artificial intelligence (AI) is being used to improve and mechanize data storytelling. It starts with a survey of the body of research on data storytelling and artificial intelligence. Next, the approaches employed in AI-driven storytelling are looked at. In order to demonstrate the useful applications of AI in this subject,

¹ How to cite the article: Chundru S.; Seeing Through Machines: Leveraging AI for Enhanced and Automated Data Storytelling; *International Journal of Innovations in Scientific Engineering*, Jul-Dec 2023, Vol 18, 47-57

the paper also provides case examples from a variety of industries. In conclusion, it explores the difficulties and moral issues surrounding AI-powered data storytelling and provides predictions for next developments.

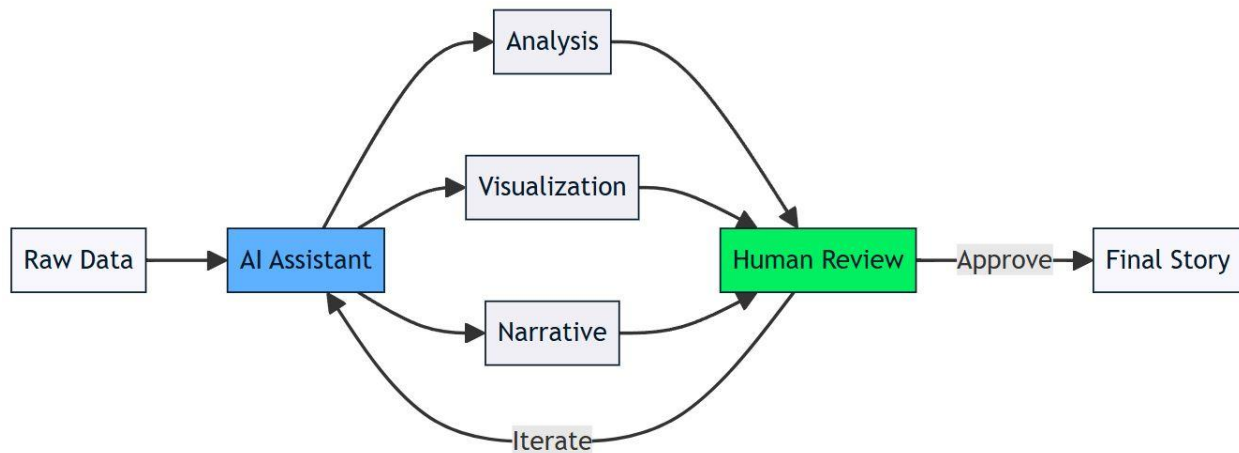


Fig 1: Data StoryTelling

OVERVIEW OF THE LITERATURE AND BACKGROUND

Data storytelling's evolution

The idea of storytelling has a long history in human history and is a basic means of disseminating information, experiences, and cultural practices. Storytelling in the context of data analysis started to change into a more formal discipline in the late 20th century with the introduction of data visualization tools. In order to communicate insights, early data storytelling techniques were primarily manual and relied on textual narratives, charts, and graphs. The amount of data that could be handled and understood by human analysts, however, placed limitations on these systems.

A major turning point for data storytelling came with the emergence of big data in the early 2000s. New methods of data analysis and display were required due to the massive amount of data produced by social media, connected gadgets, and digital transactions. Advanced data visualization technologies that could handle larger datasets and present complex information in more approachable formats began to appear during this time. Even with these developments, humans were still primarily responsible for data interpretation and story creation.

AI's Place in Data Analysis

Data analysis has been profoundly impacted by artificial intelligence, especially in the areas of machine learning and natural language processing. In data, machine learning algorithms can spot correlations and patterns that human analysts would miss. Over time, these algorithms get more accurate and efficient as a result of learning from data. The ability of robots to comprehend and produce human language is known as natural language processing, or NLP. This branch of artificial intelligence is essential for constructing narratives from data.

More advanced methods of data storytelling have been made possible by the use of AI into data analysis. AI-driven technologies are able to provide accurate and contextually appropriate narratives, recognize important trends, and automate the extraction of insights from large, complicated datasets. These features have increased the potential of data storytelling and given enterprises the ability to share insights more widely and more successfully.

Important Technologies for AI-Powered Data Storytelling

The application of AI in data storytelling is supported by a number of important technologies. Among them are:

- Learning by Machines (ML): Large datasets can be analyzed by ML algorithms to find trends, correlations, and anomalies. The narrative that AI tools are able to produce are based on these insights.

- **Natural Language Generation (NLG):** Based on data inputs, NLG is a subset of Natural Language Processing (NLP) that generates text that is human-like. NLG algorithms are able to create narratives that make data understandable and interesting for human readers.
- **Data Visualization:** Artificial intelligence (AI)-powered tools for data visualization can choose for you the best visual representations of data, including graphs, charts, and infographics. These technologies also have the ability to dynamically modify visuals in response to user input.
- **Automated Insight Extraction:** AI can minimize the time and effort needed for analysis by automating the process of obtaining important insights from data. Real-time storytelling is made possible by this technology, as stories may be modified in response to new data.

The Evolution of Data Storytelling Through Time

Although it is a new name, the activity of data storytelling has centuries-old roots. Ancient civilizations utilized maps, diagrams, and charts as early forms of data representation to communicate information about geography and astronomy. Among the early attempts to simplify complex data so that a wider audience might understand it were these crude graphics.

Modern data visualization was made possible by pioneers like Charles Minard and William Playfair in the 18th and 19th centuries. Playfair, who is sometimes credited with creating the line graph and bar chart, showed how to visually represent numerical data. Minard's well-known portrayal of Napoleon's 1812 Russian campaign, which combines geographical, statistical, and narrative components into a single, all-encompassing visualization, is regarded as a foundational example of data storytelling.

The development of computer-generated visuals in the 20th century greatly expanded the field of data visualization. However, data storytelling did not start to take on its current form until the development of personal computers and the internet. In the twenty-first century, as digital data became more widely available, there was an increasing demand for more advanced techniques for insight analysis and communication. The groundwork for using AI into data storytelling was laid by this requirement.

Artificial Intelligence's Rise in Data Narration

AI's incorporation into data storytelling is a major paradigm change. Initially, data analysis chores like data cleansing, sorting, and initial pattern recognition were mostly automated by AI. But as AI technologies developed, they were used for more complicated activities like insight creation, story construction, and dynamic visualization, going beyond mere automation.

The financial industry was among the first to use AI in data storytelling, with algorithms used there to produce automated reports based on financial data. These algorithms could provide timely and appropriate narratives and analyze market trends fast. Time was saved, and in addition, this method offered a consistency and objectivity that manual analysis could not match.

The use of AI technologies in data storytelling expanded as they developed further. An essential element of AI-driven storytelling, natural language generation (NLG) started to be vital in automating the development of narratives from data. Machine learning models trained on large datasets were used by NLG systems to produce human-like language that elucidated data insights in a comprehensible way. This advancement signified a noteworthy progression towards enhancing the accessibility and scalability of data storytelling.

Table 1: Comparison of AI Techniques in Data Storytelling

AI Technique	Description	Strengths	Weaknesses	Applications
Natural Language Generation (NLG)	Converts structured data into narrative text.	Automates report generation; scalable.	May lack nuanced context.	Automated report writing, content creation
Machine Learning (ML)	Identifies patterns in data to make predictions or decisions.	Learns from data, improves over time.	Requires large datasets; prone to bias.	Predictive analytics, personalization
Deep Learning (DL)	Uses neural networks for complex pattern recognition.	High accuracy in complex tasks; handles large data.	Black-box nature; high computational cost.	Image recognition, natural language processing
Explainable AI (XAI)	Provides transparency in AI decision-making.	Builds trust, essential in regulated industries.	May reduce model complexity or accuracy.	Healthcare, finance, legal industries

Recent Findings and Advancements

The amount of research being done on AI-driven data storytelling is expanding quickly, which is in line with the growing interest in this area. Numerous facets of AI's function in data storytelling have been studied in detail, ranging from the specific technical capabilities of various AI models to the wider ramifications of automating the development of narratives.

The creation of more sophisticated machine learning algorithms that can handle datasets that are getting more and more complicated has been one area of study. For instance, multiple-layered neural networks used in deep learning techniques have demonstrated promise in spotting minute patterns and correlations in data that conventional approaches might overlook.

The increased interest in AI-driven data storytelling is reflected in the rapid expansion of research being done in this field. Many aspects of AI's role in data storytelling have been thoroughly examined, from the particular technical prowess of different AI models to the broader implications of automating the creation of narratives.

One area of research has been the development of increasingly complex machine learning algorithms that can handle datasets that are becoming increasingly complex. For example, the use of multiple-layered neural networks in deep learning techniques has shown promise in identifying minute patterns and correlations in data that may go unnoticed by traditional methods.

The fast growth of research in this area is indicative of the growing interest in AI-driven data storytelling. Numerous facets of artificial intelligence's function in data storytelling have been extensively researched, ranging from the specific technical capabilities of various models to the wider consequences of automating the composition of stories.

The creation of increasingly sophisticated machine learning algorithms that can handle increasingly complicated information has been one field of study. For instance, the application of multiple-layered neural networks in deep learning techniques has demonstrated potential in spotting minute correlations and patterns in data that conventional approaches might miss.

Important Difficulties in AI-Powered Data Narration

There are still a number of obstacles in the way of AI-driven data storytelling, despite tremendous achievements. The interpretability of AI models is one of the main difficulties. Even if AI is capable of creating stories and finding patterns, the underlying mechanisms may be opaque, making it challenging for consumers to comprehend how conclusions were arrived at. In high-stakes industries like banking and healthcare, where decision-makers must completely comprehend the foundation of the insights supplied to them, this lack of openness can breed mistrust.

The caliber of the data utilized to train AI algorithms presents another difficulty. The quality of AI-driven data storytelling relies on the quality of the underlying data. The stories that artificial intelligence creates may be deceptive if the data is skewed, out-of-date, or incomplete. Because of this, ensuring data quality is essential, but it is still a difficult task.

Lastly, concerns concerning the function of human analysts are brought up by the automation of data storytelling. Although AI has the potential to improve and expand the narrative process, there is a worry that relying too much on automation will lessen the need for human knowledge. This issue is especially pertinent in fields where the ability to make complex decisions and possess domain-specific expertise is essential for analyzing data and creating compelling stories.

METHODOLOGY

AI Techniques and Tools in Data Storytelling

In this section, we explore the various AI techniques and tools employed in the enhancement and automation of data storytelling. The methodology focuses on three core areas: data preprocessing and pattern recognition, narrative generation using Natural Language Generation (NLG), and AI-driven visualization.

Data Preprocessing and Pattern Recognition

Data preprocessing is the first step in the AI-driven data storytelling process. It involves cleaning and organizing raw data to ensure it is in a suitable format for analysis. AI techniques such as data imputation, normalization, and feature extraction are commonly used to handle missing data, outliers, and noise. Machine learning algorithms, particularly those based on deep learning, are then applied to identify patterns and correlations within the data.

To find hidden structures in the data, one well-liked method is to employ unsupervised learning algorithms, such as clustering and dimensionality reduction approaches. Similar data points can be grouped together using clustering techniques like k-means or hierarchical clustering, which facilitates the identification of trends and anomalies. By lowering the number of variables while keeping the most important patterns, dimensionality reduction techniques like Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) aid in the simplification of complicated datasets.

Using Natural Language Generation to Create Narratives (NLG)

After patterns are found, the next stage is to create narratives that provide a human-readable explanation of the patterns. Natural Language Generation (NLG) is useful in this situation. AI is used by NLG systems to transform structured data into text that seems like it was authored by a person. Because these systems are usually trained on enormous text corpora, they are able to provide tales that are both coherent and appropriate for the given context. The content determination, document structure, sentence aggregation, lexicalization, and surface realization phases make up the NLG process. The system determines which facts and insights are most pertinent to include in the story during the content decision step.

While sentence aggregation gathers relevant information into succinct sentences, document structuring arranges these insights into a logical flow. Surface realization is the process of creating sentences that are grammatically correct at the end, while lexicalization chooses the right words to express the insights.

Modern NLG models, such the transformer architecture-based GPT-4 model, can produce extremely complex narratives that hold the reader's attention while providing the essential information. These algorithms can be adjusted

for particular industries, like marketing, banking, or healthcare, to guarantee that the narratives that are produced are true and appropriate in the given environment.

Visualization Powered by AI

Data storytelling requires visualization because it gives the data a visual form that can strengthen and complement the tale. Based on the data and the target audience, AI-driven visualization tools can choose the best visual formats—such as graphs, charts, or infographics—automatically. When a user filters, zooms, or drills down into particular data points, these tools frequently use machine learning techniques to dynamically modify the visuals.

The ability of AI-driven visualization to handle complicated, multidimensional data is one of its main advantages. AI, for instance, is capable of producing interactive visualizations that let people explore various facets of the data by changing its parameters or seeing it from other angles. In addition to increasing user accessibility, this interactivity gives consumers the ability to find insights that may not be immediately clear from the narrative alone.

The field of visual analytics, which blends automated data analysis with interactive visual exploration, frequently contributes techniques to AI-driven visualization tools. With this method, users can take advantage of the advantages of both AI and human intuition, leading to a more thorough and perceptive interpretation of the data.

Case Study: Using AI to Tell Financial Data Stories

This section provides a financial case study to demonstrate how AI is used in data storytelling. Because of the need to analyze massive volumes of data in real time and produce insights that can be put to use, the financial sector has long been at the forefront of the adoption of AI technologies. The potential of AI-driven data storytelling to improve decision-making processes has been particularly proven by the application of AI in financial reporting and investment research.

Historical Context

Every day, the financial industry produces a massive amount of data, ranging from trading volumes and stock prices to economic indicators and corporate earnings reports. Manually analyzing this data takes a lot of time and is prone to mistakes, which can result in lost opportunities or bad investment choices. Numerous financial institutions have resorted to AI in order to improve and automate the process of data analysis and storytelling in order to handle these issues.

The application of AI to the creation of earnings reports is one prominent example. In the past, financial analysts would examine the quarterly or annual financial accounts of a business, evaluate important performance indicators, and compose a narrative report that summarized the business's achievements. Days or even weeks may pass during this process, depending on the intricacy of the data and the workload of the analyst.

The development of AI-powered technologies has greatly expedited this procedure. These days, artificial intelligence (AI) systems can quickly produce a thorough profits report by analyzing financial information and extracting important insights. These systems employ Natural Language Generation (NLG) to produce narratives that explain the company's financial performance, point out trends, and make performance projections for the future.

Methodology

In this case study, we look at how a top financial services company implemented an AI-driven data storytelling tool. The application was created to automate the creation of investment research and earnings reports using a combination of machine learning algorithms and NLG.

The process comprised a number of crucial steps:

1. Gathering and preprocessing data: The company's data warehouse and AI system were connected, giving the AI system access to real-time financial data. Income statements, balance sheets, cash flow statements, and market data like stock prices and economic indicators were all included in this data. After that, the data underwent preprocessing to address missing values, standardize metrics throughout various businesses, and compile past data for trend research.

2. Pattern Recognition and Analysis: To find patterns and connections in the financial data, machine learning methods were used. The technology could identify patterns in debt levels, profit margins, or revenue growth, for instance. Additionally, it could spot abnormalities that call for more research, like abrupt declines in income or unforeseen changes in expenses.

3. Narrative Generation: The system's NLG component produced a narrative report after the analysis was finished. A review of the business's financial performance, analyses of significant trends, and performance predictions for the future based on the patterns found were all included in this report. The wording and degree of detail in the story were adjusted to meet the needs of various groups, including investors, executives, and regulatory agencies.

4. Visualization: The AI tool produced graphics that showed the company's financial success over time to enhance the story. These visuals featured scatter plots that illustrated the relationship between various financial measures, bar charts that compared expenses over various quarters, and line charts that displayed revenue growth. Users were able to investigate the data in greater detail because to the interactive visuals.

Results

For the financial services company, the use of the AI-driven data storytelling tool produced a number of noteworthy advantages:

Efficiency: Creating profits reports now takes only a few minutes instead of several days. As a result, analysts were able to concentrate on more strategic duties like analyzing the insights produced by the AI system and advising clients.

Accuracy: By processing vast amounts of data and spotting patterns, the AI system was able to lower the possibility of human error. The approach produced objective and consistent narratives that provide a trustworthy foundation for making decisions.

Scalability: The company was able to generate earnings reports for a greater number of organizations than it could have done in the past thanks to the AI technology. This scalability was especially helpful during the earnings season, when there was a significant rise in the number of reports required.

Customization: Every stakeholder received the information that was most pertinent to them thanks to the AI system's capacity to adapt narratives and visualizations to various audiences. This personalization raised user involvement and enhanced the reports' efficacy.

Discussion

The financial industry's success with the AI-driven data storytelling tool brings to light a number of crucial factors for the wider use of AI in data storytelling:

- **Including Pre-existing Systems:** The degree to which AI-powered storytelling tools can be smoothly integrated with current data infrastructures determines how effective they will be. The success of the case study was largely dependent on the AI system's interface with the company's data warehouse, which allowed for real-time access to financial data and guaranteed that the narratives were based on the most recent information.

- **Interpretation and Credibility:** Even though the AI tool greatly increased accuracy and productivity, the company understood how crucial it was to retain human control.

- **In order to make sure the narratives produced by the AI system were correct and consistent with the firm's findings,** analysts were involved in their examination. This strategy made sure that the final reports included both the machine's analysis and human judgment, which in turn helped to increase confidence in the AI tool.

- **Ethical Concerns:** Transparency and accountability were two major ethical issues brought up by the use of AI in financial reporting. In order to allay these worries, the company put policies in place to guarantee that human analysts would always be responsible for the final findings and that the AI system's decision-making procedures would be transparent.

Table 2: Case Studies of AI-Driven Data Storytelling

Case Study	Industry	AI Techniques Used	Outcome	Challenges
Company A	Marketing	Natural Language Generation (NLG)	Improved customer engagement through personalized campaigns.	Ensuring data quality and reducing bias in recommendations.
Company B	Finance	Machine Learning (ML)	Automated financial reports with high accuracy.	Regulatory compliance and interpretability of results.
News Organization X	Journalism	Deep Learning (DL)	Real-time news generation with automated insights.	Balancing automation with human editorial oversight.
Healthcare Institution Y	Healthcare	Explainable AI (XAI)	Enhanced diagnostic tools with transparent decision-making.	Maintaining transparency without sacrificing accuracy.

AI's Drawbacks and Obstacles in Data Storytelling

Although AI-driven data storytelling has many advantages, in order to fully realize its potential, a number of issues and restrictions must be resolved.

Bias and Data Quality

Making sure the data used to train and inform AI models is of a high enough quality is one of the biggest obstacles in AI-driven data storytelling. Inaccurate or misleading insights resulting from poor data quality can compromise the narrative process' efficacy. Furthermore, AI systems have the potential to reinforce or even magnify biases already in the data. An AI model may produce biased or unjust results, for example, if it is trained on data that overrepresents a particular population. This might be reflected in the narratives the model creates.

Strong data governance procedures must be put in place in order to reduce these risks. These procedures must include frequent audits of the quality of the data as well as the application of methods like bias detection and mitigation. Furthermore, it is essential for the AI's decision-making process to be transparent in order to guarantee that consumers can trust the outcomes and comprehend the process by which insights are produced.

Transparency and Interpretability

AI models can be complicated and challenging to understand, especially those that are based on deep learning. Data storytelling is challenged by the "black box" aspect of AI, since users might not be able to comprehend the AI's reasoning behind particular recommendations or findings. This lack of openness can breed mistrust, particularly in fields like law, banking, and healthcare where interpretability is vital.

Researchers and practitioners are investigating ways to enhance the interpretability of AI models in order to tackle this difficulty. By giving users insight into the model's decision-making process, techniques like explainable AI (XAI) seek to increase transparency in AI systems' decision-making processes. Explainability qualities can assist establish trust in the context of data storytelling and guarantee that the narratives produced by AI are comprehensible and believable.

Cooperation Between Humans and AI

Artificial intelligence (AI) can greatly improve and automate the data storytelling process, but it cannot replace human knowledge. While AI systems are excellent at processing massive datasets, finding patterns, and creating narratives, they are devoid of the contextual awareness and nuanced understanding that human analysts possess. A combination of human interpretation and AI-driven insights is frequently necessary for effective data storytelling.

Organizations should take a collaborative approach, using AI tools to enhance human capabilities rather than to replace them, in order to get the greatest results. With this method, human analysts may concentrate on more complex duties including deciphering insights produced by AI, using domain-specific expertise, and creating engaging narratives for target audiences.

Regulation and Ethical Issues

Ethical and legal issues are becoming more crucial as AI is used more and more in data storytelling. Fairness, accountability, and openness are just a few of the ethical standards that AI systems must be created and implemented with. Organizations must also make sure that their use of AI conforms with all applicable rules and regulations, including those pertaining to data protection and industry-specific standards.

The possible societal impact of AI-driven data storytelling is also a matter of ethical concern. For instance, experts in data analysis and storytelling may lose their jobs as a result of automation in narrative creation. Businesses need to think about the bigger picture when implementing AI and take action to lessen any unfavorable consequences on their workers and the community.

Prospective Courses

Since AI-driven data storytelling is still in its infancy, there are a lot of prospects for more study and advancement in this area. The following are some of the major areas that will probably influence this field's future:

Developments in the Production of Natural Language

The quality and sophistication of NLG systems should significantly improve as AI models continue to advance. It is expected that in the future, NLG models will be able to produce narratives that are as sophisticated and contextually aware as human-written prose. These developments will increase the potency of AI-driven tools in industries like corporate communications, marketing, and journalism by enabling them to tell compelling and convincing tales from data.

Storytelling using Multimodal Data

The use of multimodal data sources into AI-driven storytelling is an intriguing new field of research. The majority of AI-driven data storytelling technologies available today concentrate on structured data, such text and numbers. Unstructured data, like pictures, videos, and audio, are becoming more and more popular to use in narrative, nevertheless. AI systems will be able to provide richer, more engaging stories that offer a more thorough knowledge of the data by merging insights from many data sources.

Adaptive and Interactive Narration

Users can interact with and investigate the story via interactive storytelling, which is a new concept that will probably gain popularity in the future. AI-powered solutions will progressively provide interactive functionalities that let users tailor the story, experiment with various scenarios, and go more deeply into particular data points. Furthermore, AI systems might develop greater adaptability, changing the story in real time in response to user comments or modifications in the data.

Responsibly Innovating and Ethical AI

There will be an increasing focus on ethical AI and responsible innovation as AI-driven data storytelling gains traction. The development of frameworks and policies to guarantee that AI systems are created and applied in ways that are

equitable, open, and accountable will probably be the main emphasis of future research. This entails dealing with concerns like prejudice, privacy, and the effects of AI adoption on society.

CONCLUSION

The incorporation of AI into data storytelling signifies a revolutionary advancement in the ways we interpret and disseminate data. AI-driven solutions have the potential to unleash new levels of efficiency, accuracy, and scalability by automating and improving the narrative process. In an increasingly data-driven environment, these tools can assist firms in improving decision-making, effectively engaging their audiences, and maintaining their competitiveness.

But there are also a lot of obstacles to overcome in the use of AI in data storytelling. To make sure that the advantages of AI are fully realized, issues like data quality, interpretability, and ethical considerations must be carefully addressed. Furthermore, it's critical to understand that AI is a tool that can supplement and improve human abilities rather than a substitute for human knowledge.

REFERENCES

- [1] M. Floridi, "AI and its New Winter: From Myths to Realities," *Philosophy & Technology*, vol. 33, no. 1, pp. 1-5, 2020.
- [2] W. McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference (SciPy 2010)*, Austin, TX, USA, 2010, pp. 51-56.
- [3] T. Mitchell, *Machine Learning*, New York, NY, USA: McGraw-Hill, 1997.
- [4] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, no. 4-5, pp. 993-1022, May 2003.
- [5] P. Domingos, "A Few Useful Things to Know About Machine Learning," *Communications of the ACM*, vol. 55, no. 10, pp. 78-87, Oct. 2012.
- [6] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, Oct. 2001.
- [7] D. R. Cox and D. V. Hinkley, *Theoretical Statistics*, London, UK: Chapman & Hall, 1974.
- [8] 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*, San Francisco, CA, USA, 2016, pp. 785-794.
- [9] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," in *Proceedings of the First International Conference on Machine Learning*, San Francisco, CA, USA, 2003, pp. 133-142.
- [10] G. Hinton, S. Osindero, and Y. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation*, vol. 18, no. 7, pp. 1527-1554, Jul. 2006.
- [11] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379-423, Jul. 1948.
- [12] J. R. Quinlan, "Induction of Decision Trees," *Machine Learning*, vol. 1, no. 1, pp. 81-106, Mar. 1986.
- [13] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.
- [14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Cambridge, MA, USA: MIT Press, 2016.
- [15] A. K. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," in *Proceedings of the AAAI-98 Workshop on Learning for Text Categorization*, Madison, WI, USA, 1998, pp. 41-48.

- [16] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," in Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 2015, pp. 1-15.
- [17] R. J. Tibshirani, "Regression Shrinkage and Selection via the Lasso," Journal of the Royal Statistical Society, vol. 58, no. 1, pp. 267-288, Jan. 1996.
- [18] F. Chollet, "Xception: Deep Learning with Depth wise Separable Convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 2017, pp. 1251-1258.
- [19] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," Communications of the ACM, vol. 51, no. 1, pp. 107-113, Jan. 2008.
- [20] S. Bubeck, "Convex Optimization: Algorithms and Complexity," Foundations and Trends in Machine Learning, vol. 8, no. 3-4, pp. 231-357, 2015.
- [21] G. E. Hinton, "Connectionist Learning Procedures," Artificial Intelligence, vol. 40, no. 1-3, pp. 185-234, Sep. 1989.
- [22] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech Recognition with Deep Recurrent Neural Networks," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), Vancouver, BC, Canada, 2013, pp. 6645-6649.
- [23] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," Science, vol. 290, no. 5500, pp. 2323-2326, Dec. 2000.