Discuss the Applications of AI Technologies in Smart News Generation

XinQi Wei, The Hong Kong Polytechnic University, Hong Kong SAR Cong Qi, The Hong Kong Polytechnic University, Hong Kong SAR

The Asian Conference on Media, Communication & Film 2023 Official Conference Proceedings

Abstract

With the advancement of the Artificial Intelligence (AI) algorithms, the machine-generated news is widely employed in the news media industry. This has brought tremendous opportunities and challenges to the journalism industry as well as the entire society. This paper introduces the current applications of AI technologies in smart news generation. Specifically, we focus on the automation processes of the current text news generation and video news generation by using AI technologies such as machine learning, deep learning, artificial neural network, and natural language processing and generation. It then summarizes the effectiveness of the machine-generated news from the perspectives of timeliness, efficiency, level of objectiveness, reliability and bias avoidance. We conclude the paper by raising up the challenges AI-generated news bring to the news purveyor, to the news readers, and to the news content generation itself. As AI writing is in a progressive era, journalists nowadays need to consider transforming and moving from a functional model of writing simple news stories to more in-depth and creative reporting, and be able to co-exist harmoniously with AI news bots. This paper provides researchers and practitioners from the field of news media with solid technological foundations to understand the current smart news generation mechanisms, and opens the door for further exploration and discussion of machine-generated news in many other fields of news media industry. Further empirical studies on the effectiveness of AI generated news across multiple scenarios and countries are suggested to verify our research results.

Keywords: Machine-Generated News, AI, Machine Learning, Fake News, News Media



The International Academic Forum www.iafor.org

Introduction

There has been notable advancement in the development and implementation of intelligent news media in recent years. Prominent digital technologies, including Artificial Intelligence (AI), have significantly enhanced the automation, timeliness, and overall volume of news production. The field of intelligent news media involves IT applications, such as natural language processing (NLP), natural language generation (NLG), deep learning, data processing, artificial neural networks (ANN), computer vision, and others (Williamson, 2019). These technologies could be used for news generation purposes, as well as for the identification of fraudulent news, particularly in light of the substantial volume of misleading data generated by social media platforms. (Ozbay and Alatas, 2020).

In addition to automatically generated news articles, to cater to the demands of modern news consumers, certain platforms have evolved to incorporate multiple advanced technologies enabling the generation of concise and intelligent news videos. These technology platforms are usually called AI content generators (Dey, 2022), or Machine-Generated Content (MGC) platforms (Textbroker, 2019). The emergence of these new forms of news has captured the attention of the journalism industry, bringing forth unprecedented challenges and opportunities for the news media sector.

This paper endeavors to offer a concise overview of the current applications of smart news media employing AI technologies. Subsequently, it will undertake an evaluation and comprehensive discussion of the effectiveness of AI-enabled news from diverse perspectives. At last, potential future research directions will be discussed.

AI Applications in News Media

Brief Review

News-writing bots have gained extensive usage in the news media, with media outlets worldwide progressively introducing their own news-writing bots since 2001. Notably, in 2015, Xinhua News Agency unveiled the Kuaibixiaoxin robot, and Tencent Finance launched its robot reporter called "Dreamwriter". In 2016, Today's Headlines incorporated the use of the "Xiaomingbot". These writing robots have significantly aided media platforms in augmenting their output during the Rio Olympics. In the US, there are also similar writing bots, such as "Heliograf" from the Washington Post, which won the BIGGIES prize in 2018 (WashPost, 2018); Examples of such applications include "Wordsmith" from the Associated Press and "Heliograf" from the Washington Post. Furthermore, besides the text-based news writing bots, some certain video news content generation platforms like Media Brain MAGIC are garnering significant attention from users, despite the system not being fully matured. In short, these applications can be distinguished into two categories: text-based and video-based, and a representative application of each will be presented next.

Text-Based News: Xiaomingbot

Xiaomingbot, developed by ByteDance and Shanghai Jiao Tong University, is a multilingual and multimodal intelligent newsbot with four key capabilities. These include automatic news generation, multilingual translation, news reading, and animated visual presentation. Utilizing text generation algorithms, Xiaomingbot can generate news content from data, and it employs machine translation algorithms to translate the text into multiple languages. The system produces a visual avatar to read the news, with facial expression and lip motion synchronized with automatically generated voice (Xu, et al., 2020). The animated image of Xiaomingbot shows up in figure 1. The architecture of Xiaomingbot is displayed in figure 2, (Xu, et al., 2020). It demonstrated the capabilities and components of Xiaomingbot, which include text, voice, and animation.



Figure 1: Xiaomingbot animated

Figure 2: Xiaomingbot's system architecture

In 2020, the AI lab department in ByteDance company upgraded its capacity, and introduced a 3D version of Xiaomingbot. Also, the firm reposition Xiaomingbot as a 3D multilingual AI news reporter in 2020.

Video News Production Platform: Media brain·MAGIC

"Media Brain" is the first domestic media AI platform developed independently by Xinhua Zhiyun, which integrates cloud computing, the Internet of Things (IoTs), big data, AI, and other technologies to provide multiple services for media organizations (China Money AI, 2020). It includes lead discovery, material collection, editing and production, distribution and dissemination, and feedback monitoring, which make the applications and services of AI in news scenarios more intelligent. The mission of the Media Brain platform is to help content producers gather, process, and organize news sources better and faster.



Figure 3: Media brain MAGIC platform

AI Technologies in News Generation

Machine Learning

To do fact-checking automatically, we could use machine learning to detect fake news (Khan, 2019). Generally, there are three categories in machine learning: supervised learning, unsupervised learning, and reinforcement learning. In Balaji (2021)'s work, the classification of machine learning algorithms was summarized in Figure 3. In his work, an in-depth study on how to use machine learning to comprehensively analyze social media data via a variety of algorithms was discussed.



Figure 4: Classification of machine learning algorithms, from Balaji, 2021

In these algorithms, according to Ozbay (2020)'s research, a supervised classification method could be applied to detect fake news via online social media effectively (Ozbay, 2020). To examine the efficiency and predictive capacity of these algorithms, Reis et al. (2019) presented a new set of methods in both theory and practice. The research result revealed that although all sets of features have the ability to recognize fake news, there is still some room for further improvement in the performance of all models for fake news detection.

Deep Learning

Besides machine learning, deep learning is also widely used in smart news media to do text classification. In the news media, deep learning is mainly used for multi-label news text classification with an aim to help news consumers accurately filter the right news content. Today's machine learning relies highly on advanced deep learning and predictive analytics. Deep learning is especially helpful in handing a larger amount of unstructured data (e.g., pure text) in the era of big data. In sum, deep learning models rely on large and complex datasets to mimic the learning capabilities of the human brain (Chan-Olmsted, 2019; Hassaballah and Awad, 2020) to do text classification.

In recent years, many studies have been done on how to use deep learning methods to do multi-label news classification. For instance, the label-aware attention and semantic dependency (LAA_SD) modeling based on graph convolutional network (GCN) enhances the combination of text features and label semantics, and thus improves the efficiency of deep learning (Liu, et al., 2022). Another approach uses the FAA-MLSME algorithm and the WTL-Bert model to expand the data based on an attention mechanism (Ding, 2022). The aim of these models is to reduce the proportion of imbalances in multi-label data sets, and improve the learning effect, thus ameliorating the adverse effects of unbalanced data sets in classification. Another autoregressive language model, GPT-3, applied deep learning to study human-like text (Floridi and Chiriatti, 2020). However, there is a short of literature to empirically test the performance of these newly generated deep learning models in reality.

Artificial Neural Network

Similar with general machine learning, artificial neural network is also useful in fake news detection. As real and fake news spreads online, some investigators have found that fake news spreads faster than real news (Vosoughi et al., 2018). Contrary to conventional thinking, news bots accelerate the entire dissemination process. Therefore, fake news detection is essential for smart journalism.

Zhang (2022) proposed a detecting method based on social context and news articles. By using two existing real-world fake news datasets - BuzzFeed and Fakeddit, the combination of content attributes and users' communication with specific news articles are examined. The results of the study show that the method performs better than existing fake news detection methods. Another application is the classification of a larger amount of news data. According to Ruan and Yang (2022), existing convolutional neural networks could be combined with bi-directional, long and short-term memory neural networks (BiLSTM) to improve existing news text classification models with an accuracy of up to 98.96% (Ruan & Yang, 2022).

Natural Language Processing (NLP) and Natural Language Generation (NLG)

NLP, especially NLG are useful in automatic news generation. NLP is a set of computer technologies in which machines provide a deep understanding of natural language based on algorithms set up in advance, and it is used to automatically analyze and represent human languages (Chowdhary and Chowdhary, 2020). NLG is a sub-field of NLP. It is the use of artificial intelligence (AI) programming to produce written or spoken narratives from a data set. The algorithms in NLP could enable machines to search for news faster than humans, and to better structure and edit the content of the text (Semenov, 2022).

In recent years, a variety of deep learning models have been applied to natural language processing (NLP) to improve, accelerate, and automate the text analytics functions and NLP features. Moreover, these methods are offering superior solutions to convert unstructured text into valuable data and insights. In news industry, Deep learning + NLP could be used to learn and analyze news sentiments and simulate texts that could be more easily understood by readers (Vicari, 2021). News sentiment will influence public's perception of news events in a certain way. For example, companies want their customers to be influenced by financial news sentiment that affects value trends and makes profits (Brown, 2016). Deep NLP facilitates linguistic representation and detection of fake news on social networks, which ensures the integrity of the message (De Oliveira, et al., 2021).

Effectiveness Discussion

Timeliness and Productivity

News timeliness is essential to attract news readers. For news providers and recommendation systems, the number of clicks on a story is important as it shows its popularity, and timeliness is the main driver (Xiong, et al., 2021). Machine-generated news possesses a strong capability in the journalism industry due to its ability to gather and process information in a timely manner. Xiong et al. (2021) proposed a deep news click prediction (DNCP) model to test machine-generated news' timeliness and attractiveness, and empirically proved a high correlation between the timeliness and attractiveness of news via an experiment of over 30,000 articles extracted from Today's Headlines dataset.

In terms of productivity, the efficiency of machine-generated news output has been greatly improved thanks to AI. According to Kang Liu (Chinanews, 2018), director of Tencent's AI program and head of Dreamwriter, their robot reporter "Dreamwriter" takes an average of 0.46 seconds to produce a single article, and up to 2,500 articles per day. The video news generator- MAGIC takes only 6 seconds to produce a video news story. This reaction speed allows the media to react as quickly as possible after receiving information to produce news content for news consumers. During the first 13 days of the Olympic games in 2016, a total

of 457 newsletters and event reports on badminton by Xiaomingbot. More than 30 long-form detailed reports were maintained every day (Zhang et al., 2016). It not only covered all the events from the group stage to the final stage but also posted at a speed simultaneously with the live TV broadcast.

Objectivity

News, being a mode of communication, endeavors to present factual or reasonably objective information to the broader public (Weaver & Willnat, 2012). In today's era, while individuals are increasingly drawn to personalized and recreational news owing to the inherent attributes of competitiveness, entertainment, and the utility of social activities, it is important to note that objectivity remains a crucial standard for the news media in certain domains. Carlson (2017)'s research suggested that the selection of algorithms for news has long been perceived by the general public as a technical solution to a professional problem, often lacking in objectivity and threatening to overstep the industry's existing framework of practice. Carlson (2017) believed that human's judgment of algorithms is a new challenge because it requires multiple opinions to make judgments about the objectiveness of the news. Specifically, human beings are inherently subjective and are susceptible to changes influenced by their environment. Although algorithms are inherently objective and based on valid data and information support, the overload of information on the web, especially in the social media realm, affects their subjectivity in terms of information sources. For example, the introduction and popularity of politically biased news sources represent a significant historical shift in the media environment, with important unexplored consequences (Kelly, 2019). During the 2016 election, social media facilitated the widespread dissemination of "fake news," fictionalized reports presented as factual news stories (Allcott & Gentzkow, 2017). These reports lacked objectivity, but their wide dissemination led to the introduction of biased news stories into social opinion, reflecting the importance of detecting fake news.

Alongside the inherent objectivity of algorithms, it has been argued that automatically generated professional news stories themselves possess objectivity (Wu, 2020). According to Wu (2020)'s research, athletic news stories are rated as more objective than other automatically generated news stories. However, some opponents argue that the algorithm is strongly dependent on the data, and if the adopted data set is incorrect or incomplete, the algorithm may produce non-objective results.

Reliability and Bias

Reliability and bias are also closely related to fake news. In the big data era, fake news has always been at the center of discussion in the media field. Floods of misinformation online have become a norm in the age of big data. While misinformation and disinformation have always been a feature of human society, modern technologies have made it easier and faster for actors with malicious intents to spread disinformation anywhere in the world to reach a wider audience than was possible before (Bergstrom, 2019). For example, in Andreas and Mario (2018)'s study, investigators aim to study readers' attitudes toward AI news by changing the article resource intentionally. The study results state that when the stated source of some news changed, readers always gave higher ratings to the human-written articles. In contrast, when the experiment changed the actual source, which means readers are uncertain of the real author, the subjects found the computer-written articles more credible. The study's results showed that different levels of knowledge about different types of news caused various perceptions, which further led to distinctive biases and reliability perceptions of news

articles. Nevertheless, this study demonstrated that using algorithms to assist news-writing might be able to help readers better understand and perceive the news contents.

Similarly, Marinescu et al. (2022) designed a Turing experiment, which included a study designed to test whether participants could distinguish between AI generated news and human-writing news. Before reading those articles, participants were unaware of the inclusion of AI-generated news and were assumed to have no prior exposure to such news. At the end of the experiment, when participants were asked if they had read the automatically generated news articles, 45% doubted their initial response and admitted that they may not have been aware of the AI-authored articles at the time they read the news (Marinescu et al., 2022). The result shows that when the audience does not know the source of the manuscript, there is not much difference in the degree of judgment of the two articles. This indirectly implies that general audience keep an accepting attitude toward the AI news writing manuscript itself, and do not show distinctive preference for the two types of news.

Further Discussions

Concerns encompass the implications of AI for the journalism profession, the manner in which AI technology is employed, and the level of public confidence in news organizations (Ali & Hassoun, 2019).

Content Quality Concerns

Machine-generated news often faces challenges of monotony and shallow content. In today's online news landscape, artificial intelligence heavily relies on algorithms and news templates to automatically generate content, leading to repetitive and monotonous news coverage. Furthermore, since news-generating algorithms do not yet possess the capability to consider social, political, and economic factors, the resulting news articles tend to lack depth compared to manually curated reporting, such as that found in The Economist. However, machine-generated news brings a unique advantage with its unparalleled ability to search for data, which greatly contributes to the breadth of news coverage. Additionally, advancements in information retrieval technology hold the promise of enhancing the ability to delve into more comprehensive and in-depth news reporting.

The second issue pertains to the continuous mention of fake news in previous discussions. Despite various pieces of evidence demonstrating the widespread dissemination of online misinformation as a prevailing trend, it is crucial to recognize that distinguishing news from other forms of entertainment requires the reinforcement of retrieval techniques and legal regulations pertaining to fake news. Additionally, it is important to address the ongoing controversy surrounding the evaluation of news objectivity and credibility (Kelly, 2019).

Challenges to Journalists

As algorithms become increasingly integrated, the proliferation of machine-generated news stories presents numerous challenges for journalists, compelling them to reflect on and consider how to distinguish their skills. Specifically, while machines excel in generating news, journalists place greater emphasis on possessing in-depth analytical skills, creativity, and the ability to personalize news writing (Van et al., 2012). Consequently, journalists must undergo a transformation, shifting from a functional model of simply reporting news to engaging in more comprehensive and in-depth journalism. According to Carlson (2017),

journalists will experience heightened pressure as the number of articles generated by AI continues to increase. In summary, the role of the "news gatekeeper" remains incomplete in its automation, as the news articles produced by AI may lack depth, necessitating additional review and revision work by journalists with specialized knowledge in specific subject areas.

Trust Issues

The trust controversy surrounding machine-generated news stems from concerns regarding audience privacy. When news is automatically generated by AI, there is a potential risk of the algorithm capturing the audience's private information and sharing it with other news consumers. An example of this can be seen in the user privacy permission agreements on the Tencent News mobile app, which explicitly state that the app collects and records users' device information, browsing habits, and clicks. This data is stored as log files and used for personalized information delivery. The audience may be unaware that the news app has access to their personal information. When audiences become aware of such privacy issues, it can significantly impact their level of trust in AI-generated news. This raises additional concerns that need to be taken into consideration.

Conclusion

This article focuses on the application of AI in the news media industry, initiating a divergent discussion from two representative examples regarding its technological aspects and its output efficiency. The effectiveness of AI technologies utilized in news is discussed from the perspectives of timeliness, objectivity, and reliability. Additionally, potential risks are examined, particularly the significant challenges and opportunities posed to the future prospects of machine-generated news due to the emergence of fake news resulting from information overload and other factors.

Furthermore, the academic community currently lacks empirical data to prove or support the aforementioned discussions on the issues previously mentioned. In terms of the news media industry, the majority of machine-generated news remains confined to news-writing robots and text-based content, with limited academic research on video news. These aspects can serve as further directions for research directions.

References

- Ali, W., & Hassoun, M. (2019). Artificial intelligence and automated journalism: Contemporary challenges and new opportunities. *International journal of media*, *journalism, and mass communications*, 5 (1), 40-49.
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, *31*(2), 211-236.
- Andreas G., Mario H., Bastian H., and Hans-Bernd, B. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, 19 (5), 595–610.
- Balaji, T. K., Annavarapu, C. S. R., & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. *Computer Science Review*, 40,100395.
- Bergstrom, C. T., & Bak-Coleman, J. B. (2019). Information gerrymandering in social networks skews collective decision-making. *Nature*, *573*(7772), 40-41.
- Brown, E. D. (2014). Analysis of Twitter messages for sentiment and insight for use in stock market decision-making (Doctoral dissertation, Dakota State University).
- Carlson, M. (2017). *Journalistic authority: Legitimating news in the digital era*. Columbia University Press.
- China money AI (2020). Alibaba-backed media AI firm finds Innovative Applications. *China money*. https://www.chinamoneynetwork.com/2020/12/03/alibaba-backed-media-ai-firm-find s-innovative-applications
- Chinanews (2018). "Smart Writing": Producing a story in less than a second. *China news*. https://www.chinanews.com/business/2018/07-09/8560661.shtml
- Chowdhary, K., & Chowdhary, K. R. (2020). Natural language processing. *Fundamentals of artificial intelligence*, 603-649.
- De Oliveira, N. R., Pisa, P. S., Lopez, M. A., de Medeiros, D. S. V., & Mattos, D. M. (2021). Identifying fake news on social networks based on natural language processing: trends and challenges. *Information*, 12(1), 38.
- Dey, V. (2022). Deep Dive: How AI content generators work. *VentureBeat*. https://venturebeat.com/ai/deep-dive-how-ai-content-generators-work/

Ding, H. (2022), Research and implementation of multi-label news text classification system based on deep learning. http://www.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD2023 01&filename=1022785095.nh&uniplatform=OVERSEA&v=lkerPxhPtHduJEJ5BQix Cyoyhj1HNIKppKIeyCh23bjRB8ihuLvQS4BcziEPeS7A

- Dong, C., Li, Y., Gong, H., Chen, M., Li, J., Shen, Y., & Yang, M. (2022). A survey of natural language generation. ACM Computing Surveys, 55(8), 1-38.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*, 681-694.
- Kelly, D. (2019). Evaluating the news: (Mis) perceptions of objectivity and credibility. *Political Behavior*, *41*(2), 445-471.
- Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. *Machine Learning with Applications*, *4*, 100032.
- Latar, N. L. (2015). The robot journalist in the age of social physics: The end of human journalism?. *The new world of transitioned media: Digital realignment and industry transformation*, 65-80.
- Liu, B., Liu, X., Ren, H., Qian, J., & Wang, Y. (2022). Text multi-label learning method based on label-aware attention and semantic dependency. *Multimedia Tools and Applications*, 81(5), 7219-7237.
- Marinescu, V., Fox, B., Roventa-Frumusani, D., Branea, S., & Marinache, R. (2022). News Audience's Perceptions of and Attitudes Towards AI-Generated News. In *Futures of Journalism: Technology-stimulated Evolution in the Audience-News Media Relationship* (pp. 295-311). Cham: Springer International Publishing.
- Ozbay, F. A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: statistical mechanics and its applications*, *540*, 123174.
- Reis, J. C., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). Supervised learning for fake news detection. *IEEE Intelligent Systems*, 34(2), 76-81.
- Ren, Y., Ruan, Y., Tan, X., Qin, T., Zhao, S., Zhao, Z., & Liu, T. Y. (2019). Fastspeech: Fast, robust and controllable text to speech. *Advances in neural information processing* systems, 32.
- Ruan, J. J. & Yang, P., (2022), chinese News Text Classification Based on Att-CN-Bi LSTM Model, ISSN: 1008-7974 (12): 65-70.
- Semenov, V. P., Myalenka, V. Y., Yakovlev, A. I., & Meshcheryakov, D. E. (2022, May). Artificial Intelligence in the Media Industry (on the Minds of AI2Media). In 2022 XXV International Conference on Soft Computing and Measurements (SCM) 229-232. IEEE.
- Textbroker (2019), Machine-Generated Content: Advantages and Disadvantages, *Textbroker*. https://www.textbroker.com/machine-generated-content
- Van Dalen, A. (2012). The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism practice*, *6*(5-6), 648 658.

- Vicari, M., & Gaspari, M. (2021). Analysis of news sentiments using natural language processing and deep learning. *AI & society*, *36*, 931-937.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *science*, *359*(6380), 1146-1151.
- WashPostPR. (2021). The Post's Heliograf and ModBot Technologies take first place in 2018 global biggies awards. *The Washington Post*. https://www.washingtonpost.com/pr/wp/2018/03/23/the-posts-heliograf-and-modbot-t echnologies-take-first-place-in-2018-global-biggies-awards/
- Weaver, D. H., & Willnat, L. (Eds.). (2020). *The global journalist in the 21st century*. Routledge.
- Williamson, B., Potter, J., & Eynon, R. (2019). New research problems and agendas in learning, media and technology: the editors' wishlist. *Learning, Media and Technology*, 44(2), 87-91.
- Wlodarczak, P., Soar, J., & Ally, M. (2015, October). Multimedia data mining using deep learning. In 2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC), 190-196, IEEE.
- Wu, Y. (2020). Is automated journalistic writing less biased? An experimental test of auto-written and human-written news stories. *Journalism practice*, 14(8), 1008-1028.
- Xiong, J., Yu, L., Zhang, D., & Leng, Y. (2021). DNCP: An attention-based deep learning approach enhanced with attractiveness and timeliness of news for online news click prediction. *Information & Management*, 58(2), 103428.
- Xu, R., Cao, J., Wang, M., Chen, J., Zhou, H., Zeng, Y., ... & Li, L. (2020). Xiaomingbot: A multilingual robot news reporter. arXiv preprint arXiv:2007.08005.
- Zhang, J., Yao, J. G., & Wan, X. (2016, August). Towards constructing sports news from live text commentary. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 1361-1371.
- Zhang, M. (2022). A Fake News Detection Method Based on Neural Network, *Microcomputer Applications*. 38(12).

Contact email: cong.qi@polyu.edu.hk