 doi:10.20944/preprints202110.0020.v1

Article

Artificial Intelligence in News Media: Current Perceptions and Future Outlook

Mathias-Felipe de-Lima-Santos [1,*] and Wilson Ceron [2]

[1] School of Communication, University of Navarra; mdelimas@unav.es
[2] Institute of Science and Technology, Federal University of São Paulo; wilson.seron@unifesp.br
* Correspondence: mdelimas@unav.es

Abstract: In recent years, news media have been hugely disrupted by the potential of technological-driven approaches in the creation, production, and distribution of news products and services. Artificial intelligence (AI) has emerged from the realm of science fiction and has become a very real tool that can aid society in addressing many issues, including the challenges faced by the news industry. The ubiquity of computing has become apparent and has shown the different approaches that can be achieved using AI. We analyzed the news industry AI adoption based on the seven subfields emanated from AI: (i) machine learning; (ii) computer vision (CV); (iii) speech recognition; (iv) natural language processing (NLP); (v) planning, scheduling, and optimization; (vi) expert systems; and (vii) robotics. Our findings suggest that three subfields are being more developed in the news media: machine learning, planning, scheduling & optimization, and computer vision. Other areas are still not fully deployed in the journalistic field. Most of the AI news projects rely on funds from tech companies, such as Google. This limits the potential of AI in the news industry to a small number of players. We conclude by providing examples of how these subfields are being developed in journalism and present an agenda for future research.

Keywords: journalism; artificial intelligence; computer science; machine learning; computer vision; NLP.

1. Introduction

We are now living in a world where technology and data adjudicate a large array of decisions in our lives. This means that every industry needs to adapt and embrace these technologies to become sustainable in the future. This is not different for the news industry, "the future of journalism and its business models seem to be bound by the penetration of technological deployment"(de-Lima-Santos and Mesquita 2021, 1416–141).

In recent years, news media have been hugely disrupted by the potential of technological-driven approaches in the creation, production, and distribution of news products and services (Hernandez Serrano, Greenhill, and Graham 2015; Örnebring 2010), which can be seen in novel news products and practices: data journalism (Coddington 2015; Hermida and Young 2019), immersive and drone journalisms (Harvard 2020; Kang et al. 2019), analytics (Nelson and Tandoc 2018; Ferrer-Conill and Tandoc 2018), and automation (Linden 2017b; Lewis, Guzman, and Schmidt 2019).

Along this journey, scholars have expanded their knowledge and experience, confirming that artificial intelligence (AI) capabilities are evolving every year and making its costs lower as well as more affordable computing power. The concept of AI refers more narrowly "to a branch of computer science focused on simulating human intelligence" (Broussard et al. 2019, 673).

AI in news media can certainly make life easier for overburdened resources, without replacing the unique skills of journalists. In addition, artificial intelligence can enhance new forms of participation and leverage new products that could increase news media consumption (Diakopoulos 2020; Jamil 2020). However, a number of challenges tend to delay technological innovations in newsrooms, such as resistance to change, the

institutional landscape, historical competition, insufficient funding, lack of skills, and complementary ambitions (Boczkowski 2004; de-Lima-Santos and Mesquita 2021; Krumsvik et al. 2019; Paulussen 2016).

Despite the practical deployment of AI in news media, the academic literature is still at a nascent stage. In this context, this study aims to expand it by examining case studies collected by JournalismAI, a project from the London School of Economics and Political Science (LSE). The list includes 102 cases that were collected by the institutions as well as examples that were submitted by other researchers and practitioners in the AI field. Although this list is not exhaustive, this sheds light on the level of the development of AI in news and which areas.

In order to make a theoretical contribution to this field, this study draws upon literature on artificial intelligence to sketch an outline of the field and understands where the journalism industry is positioning itself. Thus, this article poses the following questions:

RQ1. How is news media positioning itself in the subfields of artificial intelligence?
RQ2. To what extent, is AI being deployed in the news industry?
RQ3. What are the future avenues for AI in news media?

Therefore, the contribution of this paper is twofold. First, to describe the types of AI being developed in the news media industry that aims to facilitate research on the topic and helps practitioners to understand the possibilities, implications, and responsibilities that come along with AI adoption and use. Second, to provide a comprehensive analysis of the subfields to encourage scholars and professionals to consider further steps to the deployment of AI in the news industry.

2. Theoretical Framework

2.1. Artificial Intelligence in its Current Manifestation

Although the term has gained popularity in recent years, AI is not new. It dates back to 1955, when Stanford University's professor John McCarthy used the term to describe the science and engineering of making intelligent machines (McCarthy 1998). The pervasive nature of information and communications technology (ICT) and the datafication of society have expanded their applicability in a variety of fields, such as journalism (Gelgel 2020).

As technology evolves, new opportunities are available to both consumers and businesses, helping to address some of the challenges of our time (Paulussen 2016). As a result, the field of artificial intelligence has seen significant progress in recent years, led by numerous technological developments that made it more affordable. Although Hollywood and science fiction movies often depict AI as sentiment machines such as robots that can mimic human reasoning and behavior (Broussard 2018), the field is concerned with understanding and building intelligent entities that can compute "how to act effectively and safely in wide variety of novel situations" (Russell and Norvig 2021, 19). Therefore, intelligence means performing human tasks, such as recognizing images or performing repetitive tasks (Broussard et al. 2019). In this sense, some authors understand intelligence in AI as rationality, which can be loosely understood as doing the right things (Russell and Norvig 2021). However, different definitions have been developed by scholars and experts over the years. In simple terms, AI can be defined as the process of "creating computing machines and systems that perform operations analogous to human learning and decision-making" (Castro and New 2016, 2). Therefore, AI is a step-by-step procedure for solving problems.

The academic scholarship of AI has been part of computer science for decades, but the automation of cognitive tasks became possible only after the developments relating to data, sensors, and advances in technology (Chan-Olmsted 2019). The ubiquity of computing has become apparent and has shown the different approaches that can be achieved

 doi:10.20944/preprints202110.0020.v1

using AI. Based on the opinions about the most promising methods and theories, AI has been effectively deployed in some fields, while it remains beyond the reach in others (Aronson 2018; Castro and New 2016; Ortiz Freuler and Iglesias 2018).

Despite AI being a disputed concept, seven subfields emanate from this major field that present significant connections and commonalities among them: (i) machine learning; (ii) computer vision (CV); (iii) speech recognition; (iv) natural language processing (NLP); (v) planning, scheduling, and optimization; (vi) expert systems; and (vii) robotics.

The most popular is machine learning, a "subfield of AI that studies the ability to improve performance based on previous experience" (Russell and Norvig 2021, 19). In other words, machine learning is a branch of AI dedicated to designing algorithms that build models from data without explicitly programming this solution (Castro and New 2016). This has reduced the operational costs and the cost of manpower substantially in the news industry.

The development of machine learning is highly related to deep learning and predictive analytics. In deep learning, statistical techniques are used to solve problems with little human intervention. For this, the models rely on large and complex datasets in an attempt to replicate the human brain's learning capabilities (Chan-Olmsted 2019; Hassaballah and Awad 2020). Some of these models are inspired by the structure and function of neural networks to enable a computer to learn to recognize abstract patterns, simulating large, multilayered webs of virtual neurons (Castro and New 2016). Predictive analytics is a branch of machine learning dedicated to making predictions about future outcomes using historical data (Russell and Norvig 2021).

Another subfield of AI is computer vision. Through the use of mathematical algorithms, CV gives the ability of computers to derive meaningful information from digital images (Szeliski 2011). It does not mean that a CV model can actually see the content of an image like a human, but it is capable of detecting or deducing an object (Marr 2010). CV algorithms can be seen in two strands: image recognition and machine vision. Image recognition represents a set of methods for detecting and analyzing images, which can be used for the automation of specific tasks. Machina vision encompasses the capability of a computer to perceive the environment (Szeliski 2011; Marr 2010).

Speech recognition focuses on automatically and accurately transcribing human speech, converting voice data into text data, commonly found in applications that follow voice commands or answer spoken questions (Deloitte 2014). Natural language processing goes beyond that and refers to the automatic computational processing of human language (Castro and New 2016). In other words, NLP is the ability of computer programs to manipulate text and spoken words in a similar way humans do, such as understanding and responding to text or voice data, extracting meaning from sentences, or generating readable texts (Deloitte 2014). This broader subfield includes other areas such as translation, classification & clustering, and information extraction. Other terms that are relatable to the NLP are Natural Language Generation (NLG) and Natural Language Understanding (NLU), as shown in Figure 1. While the former is responsible for converting structured data into meaningful sentences in the form of natural language, the latter represents the process that turns the unstructured data into understandable structured data (Future Today Institute 2018; Locker et al. 2019).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 October 2021 doi:10.20944/preprints202110.0020.v1

Figure 1. NLP models can be divided into two subsections: NLG and NLU.

Other mature cognitive technologies are the systems that use AI to determine steps to take (planning) and understand when to carry out a certain step (scheduling) in order to achieve a goal. Additionally, this system can find its most optimal way to make all the necessary adjustments in the most efficient manner (optimization). This step is important in some cases due to the trade-offs concerning limited resources and complex decisions required by the model. This subfield is known as planning, scheduling, and optimization (Deloitte 2014; Russell and Norvig 2021).

The rules-based systems simulate the behavior and judgment of humans that have expert knowledge and experience in a particular field to automate the process of making inferences about information. Also known as expert systems, this model uses databases of knowledge and rules to solve complex problems (Russell and Norvig 2021). Robotic is a subfield of AI that integrates different cognitive technologies to enable computers and systems to perform different tasks simultaneously with people in unpredictable environments. Examples include robotic vacuum and unmanned aerial vehicles (Russell and Norvig 2021).

Although these subfields describe different applications of AI, they are interwoven and often mutually reinforcing. For this reason, some authors see AI through its five functions: monitoring, discovering, predicting, interpreting, interacting with the physical environment, humans, or machines (Castro and New 2016).

However, artificial intelligence has not yet been fully displayed in all industries. It is particularly hard to develop these technologies in areas that are suffering from liabilities, such as the news media. In the following section, we describe the nascent literature on artificial intelligence in journalism and show the potential that AI brings to the news industry.

2.2. Artificial Intelligence in the News Industry

Artificial intelligence (AI) has shown promise in numerous experimental studies, particularly in science and technology fields. Although it is hard to estimate the cost of creating and implementing an artificial intelligence application without diving into your project's details, it is known that the development of AI systems has reduced costs in

recent years. However, it still requires specialized expertise that is hard to compete with Silicon Valley companies. These big tech companies leverage their control over AI by actively acquiring startup companies that are deploying AI solutions in an attempt to concentrate power and thwart any competitor (Linden 2017a).

A recent report has shown that the major bottleneck for the development of AI in news media is the talent competition. It is not only about attracting talent but also retaining these professionals in newsrooms, which offer lower salaries in comparison to the tech industry (Cook et al. 2021). This newsroom brain drain works against the adoption of technologies in the news industry (Broussard et al. 2019).

Even with those conditions, news outlets around the world are embracing AI solutions in their newsrooms. In recent years, there was a great trend toward the automation of news stories (Linden 2017b). Although there is a certain level of use of machine learning algorithms in some of these projects, many of them still rely on simple automation that fills in the blanks of template stories, instead of producing stories built on prior data (Biswal and Gouda 2020).

A recent report has shown that the major bottleneck for the development of AI in news media is the talent competition. It is not only about attracting talent but also retaining these professionals in newsrooms, which offer lower salaries in comparison to the tech industry (Cook et al. 2021). This newsroom brain drain works against the adoption of technologies in the news industry (Broussard et al. 2019; Lokot and Diakopoulos 2016). Although many of these bots do not use machine learning or NLP models, they rely on a series of steps to take (planning) and understand when to carry out a certain step (scheduling) in order to send messages. Prior study identified four different categories of news bots: "the inputs and sources of input data; the outputs produced by the news bots; the algorithms that guide how a news bot turns inputs into outputs; and the function or intent of the news bot" (Lokot and Diakopoulos 2016, 696). In Brazil, scholars have shown that newsrooms mainly rely on Twitter to create bots that use AI models, particularly NLP, machine learning as well as planning, scheduling, and optimization, to process large volumes of data and interact on digital media platforms (DalBen and Jurno 2021). On the flip side, bots were used with malicious intent in several recent events, such as the Brexit campaign (Bastos and Mercea 2018).

Even with those conditions, news outlets around the world are embracing AI solutions in their newsrooms. In recent years, there was a great trend toward the automation of news stories (Linden 2017b). Although there is a certain level of use of machine learning algorithms in some of these projects, many of them still rely on simple automation that fills in the blanks of template stories, instead of producing stories built on prior data (Biswal and Gouda 2020).

On the other hand, there is such an expectation placed upon the possibility of AI techniques to reduce the costs of investigative journalism (Broussard 2015). However, these models are built for a particular story, which means that there is a need to create and train again for novel projects. As a result, the high initial investment cannot be amortized over multiple projects (Stray 2019). Similarly, investigative news projects that rely on computer vision require significant investment to build technological infrastructure and hire highly qualified personnel to develop such codes (de-Lima-Santos and Salaverría 2021). Furthermore, AI models are usually trained with old and biased datasets, which can generate many ethical implications (Guzman and Lewis 2020).

Traditional news outlets have been successful in deploying AI projects in their newsrooms, such as The New York Times (NYT), The Washington Post, and Associated Press (Chan-Olmsted 2019). However, even for them, AI is expensive (Broussard et al. 2019). Furthermore, elite news organizations have a long way to go to realize what is possible with AI in their newsrooms. For example, the release of GPT-3 in 2020 evolved the machine learning models for text to the next level. This model has the ability to perform many different tasks, such as article generation, translation, summarization, and prediction,

 doi:10.20944/preprints202110.0020.v1

using less computing power (Gage 2020). On the other hand, this applicability also brings risks, such as distorted content that can be used to deceive the public.

These advances in AI-related technologies have the potential to significantly disrupt the nature of human-machine interactions. AI technologies prove to be in short, medium, and long term part of a broader reconfiguration of the news industry, which started with the digitalization and the deployment of the internet (Broussard et al. 2019; Erdal 2011). However, AI is not a silver bullet for journalism, but it is a new tool that requires more understanding to further support and bolster AI capabilities in newsrooms. AI-enforcement mechanisms are important for ensuring that AI systems adhere to legal and ethical guidelines without explicitly considering the power structures between various stakeholders (Broussard et al. 2019). For this reason, it is important to understand the different subfields of AI.

3. Methods

To understand the evolution of AI in the news industry, we resort to the list of case studies from AIJournalism. This is a global initiative led by the LSE journalism's think-tank, Polis, which aims to expand the knowledge about AI-powered technologies in newsrooms through collaborative projects. The institute offers a network for best practice and innovation sharing, while it also produces research reports and training materials for the news industry. On one of the fronts, AIJournalism built a database of cases that are at the intersection of AI and journalism. In the database, the team collected the best case studies that they came across and there is also a form where organizations can send their applications of AI. According to AIJournalism's portal, the idea is to "ensure that everyone in our global network can learn from and be inspired by these creative applications of AI in journalism" (JournalismAI n.d.).

The list comprises 102 items. We excluded 9 entries, which were not possible to access their website or we could not find information about the project on the internet. Although we recognize that this is not an extensive list, it should give some ideas of how AI is being developed in the news industry. To analyze it, we built on the previous literature and classify these cases according to the type of AI application they use. As shown in Figure 1, the artificial intelligence literature suggests that there are seven major areas of AI: (i) machine learning; (ii) natural language processing (NLP); (iii) speech recognition; (iv) expert systems; (v) planning, scheduling, and optimization; (vi) robotics; and (vii) computer vision (see Figure 2).

Although we recognize that in each of these areas, there are a number of subareas, such as machine learning that can be divided into deep learning and predictive analytics, our objective is to give a macro approach and show how those topics work together in the journalistic field. We also acknowledge that certain projects rely on more than on the subfield of AI. In this case, we classified only the two major subfields of AI found on these projects. Each author qualitatively analyzed each item from the database and coded it separately. During the process, we took annotations about each project to facilitate the discussion of discrepancies. After this process, we reunited and discussed the differences in classification. In total, 102 news products were initially classified differently. We also removed 9 entries because their websites were not available and we could not find information on the web about them.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 October 2021 doi:10.20944/preprints202110.0020.v1

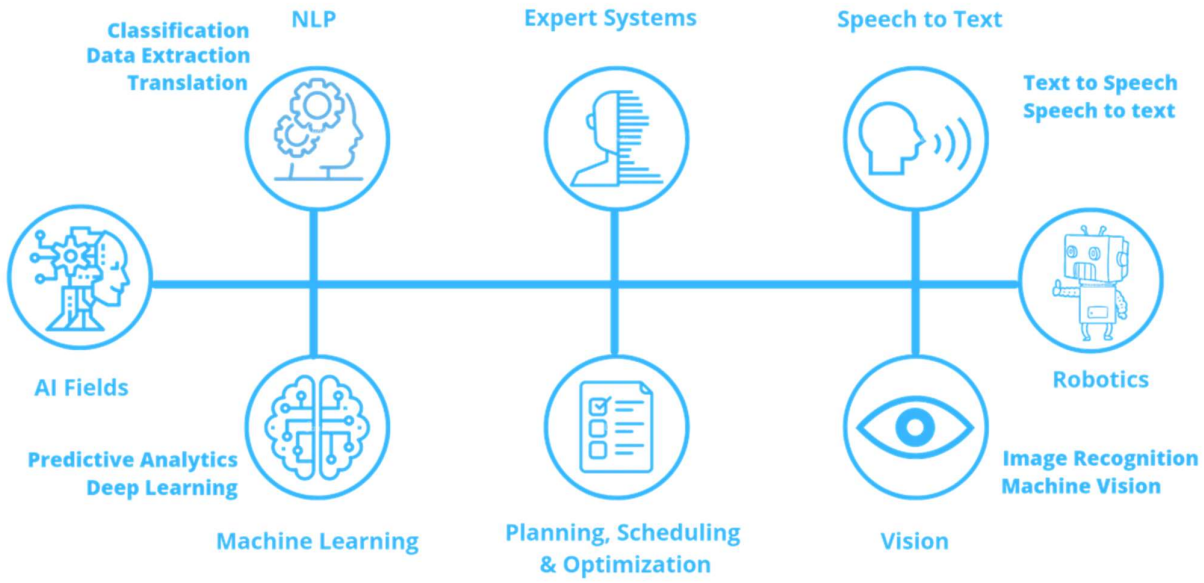

Figure 2. The seven fields of AI.

In addition, we analyzed where these projects were developed to give a better understanding of the distribution of AI-powered technologies in newsrooms worldwide. Our findings bring into perspective the use of AI by newsrooms, especially for bringing out the possibility of its use beyond the traditional forms of machine learning applications and natural language processing algorithms, detailing the most relevant aspects of these cases. In the following section, we present our findings based on the explorative research.

4. Findings

4.1. An Overview of the AI in News Media

Overall, our findings show that most of the AI applications in the news industry are being developed in the Americas (43.01%) and Europe (39.78%). These findings are similar to prior studies on other technological development in the news media ecosystem. For example, data journalism was first developed in the United States and European nations, such as the United Kingdom and the Nordic countries (Appelgren and Nygren 2014; Borges-Rey 2016; Parasie and Dagiral 2013; Young and Hermida 2015) until recently scholars have identified developments in non-Western countries (Mutsvairo 2019). This is important because, as happened with data journalism, there is an assumption of the ubiquity and growing universal acceptance of AI in the news industry, this is not the reality. There is an unequal development of artificial intelligence as it is far too expensive for the majority of media companies. This limits the adoption of cutting-edge technologies by a great part of newsrooms in the Global South (Linden 2017a). In our list, only 5.38% of the cases were from Asia and 2.15% from Oceania. Looking more specifically at the Americas, which also includes Latin America and Canada, only 4.3% of all cases belong to this region. The rest 38.71% belongs to the US. Only 9.68% of these cases encompass global projects, that is, products that reach wider audiences and different continents. Figure 3 depicts these subfields of AI in relation to each continent.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 October 2021 doi:10.20944/preprints202110.0020.v1

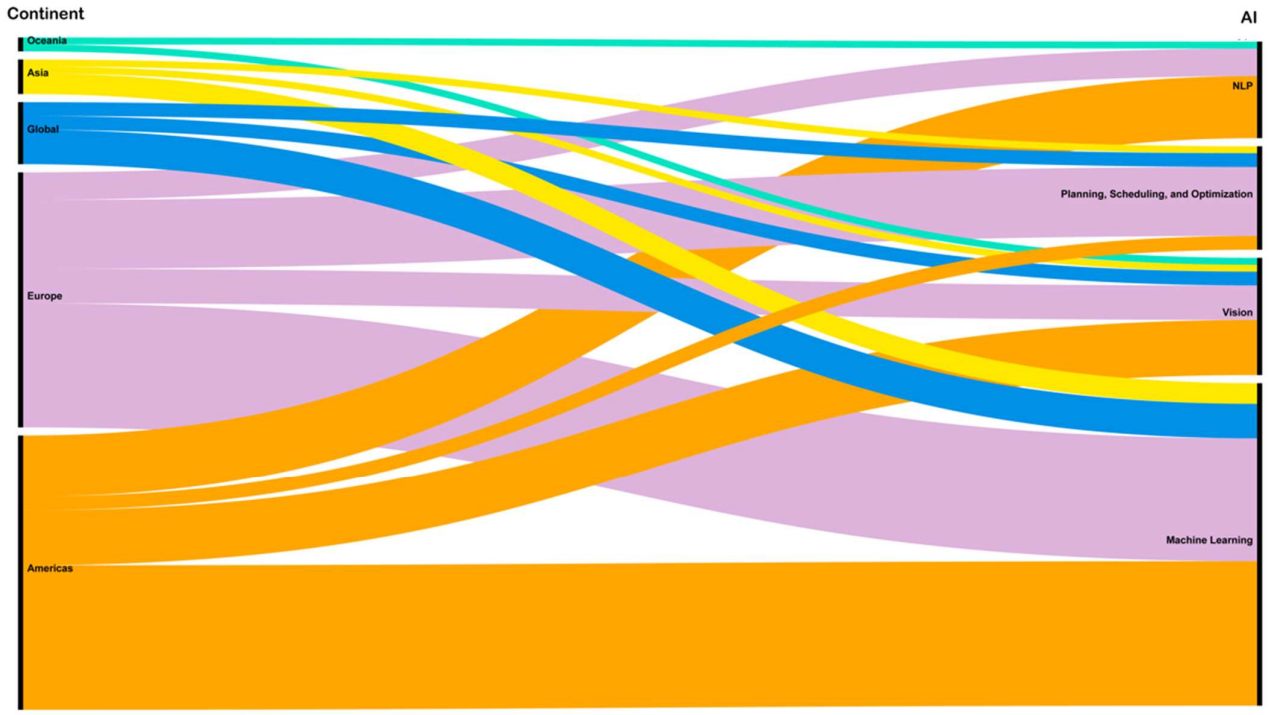

Figure 3. Fields of AI in relation to continents.

In regards to AI subfields, it has been noticed that the most common is machine learning (66.67%). This is followed by applications that rely on computer vision (18.28%) and planning, scheduling, & optimization (16.13%). NLP models learn a non-trivial amount of linguistic knowledge, which makes it less hard to replicate it in different languages, such as Portuguese (Rodrigues, Oliveira, and Gomes 2014). Although it can be found in the computer science literature fine-tuning models for the English language, it requires higher layers of specific knowledge. For this reason, NLP models represent 15.05% of our sample.

On the other hand, the subfields of speech, expert systems, and robotics are not presented in our sample. Within such a fluid environment, there is an increasing complexity to deploy projects on these subfields of artificial intelligence in the news ecosystem, as they require more specialized skills and some of them require an innovative approach to develop novel products that were not previously available in the industry. This is happening at a moment when media companies are trying to survive these uncertain times (Chan-Olmsted 2019). This integration would require significant alignment in the organization and resource investment that news outlets do not foresee in a near future (Locker et al. 2019).

However, we could identify that a great part of these projects relies on external funds to deploy such projects. A great majority is coming from Google's Digital News Innovation (DNI) grants. Currently, news media is working with technology companies in a relationship that is seen as frenemies (Rashidian et al. 2018). On one hand, these tech companies broke news media's business models; on the other hand, these big tech platforms have become the major source of funding and disruptor of innovation in the news industry (Rashidian 2020). Additionally, news organizations are relying on third-party platforms, such as Google's incubator Jigsaw, to help them develop AI solutions for their business, echoing findings from previous studies (Cook et al. 2021). In the following subsections, we will describe how the three main subfields of AI are being developed in the news industry.

4.2. Machine Learning and its Applications in the Journalistic Field

 doi:10.20944/preprints202110.0020.v1

Although many of these organizations relied on external funding to develop some of these projects, the news media is looking for ways to improve their strategies and boost their revenue streams. In our cases, two major trends appear. First, the use of machine learning models was commonly used to learn the readers' interest to boost engagement with a content recommendation engine. For example, The New York Times, New Zealand Media and Entertainment (NZME), and Toutiao (China) were some news media organizations that launched AI-powered news applications with this intent. In the same vein, The Times (UK) halved digital subscriber churn using tailored emails for its subscribers.

Second, there is a growing number of news outlets building paywalls that bend to the individual reader or predict subscription cancellation. The Wall Street Journal is an example of this, the newspaper deployed a machine learning model that allows non-subscribers to sample some stories. Using a machine learning algorithm, The Wall Street Journal could build a more flexible paywall that informs news managers of the kinds of stories and how many stories certain types of users are interested in. Similarly, third-party organizations start to build these solutions and offer them to newsrooms, such as Piano in the US and Deep BI in the UK, echoing findings from previous studies (Cook et al. 2021).

Although machine learning is the most popular subfield of AI, as it gives the ability of computers to improve performance based on previous experience (Russell and Norvig 2021), it is also the most common model because it can be applied with other subfields of AI. We could identify some projects that combined machine learning algorithms with NLP, computer vision as well as planning, scheduling, and optimization. This happens because these different components are squeezed and entangled in the AI space (Castro and New 2016). One of those subfields of AI that is commonly used with machine learning is planning, scheduling, and optimization.

4.3. Computer Vision To Investigative Reporting

Until recently, AI only worked at a limited capacity since technologists had to program a wide array of functions into a system to mimic human intelligence, requiring a massive computing power with low throughput. However, this has changed with better hardware, more data, and better algorithms (Aronson 2018; Castro and New 2016; Hassaballah and Awad 2020; Whittaker 2019). Nonetheless, computer vision is still a subfield of AI that requires a powerful tool to simulate human vision, enabling a machine to learn to recognize abstract patterns in images (Castro and New 2016; Szeliski 2011).

In our dataset, it was noticed in our sample some small news outlets developing news stories using CV. The small news outlet Texty in Ukraine relied on CV models to detect land that turned into lunar landscapes due to illegal amber mining in the country. However, we could identify a major number of traditional news organizations adopting different approaches to the use of computer vision in their newsrooms. The New York Times and Reuters used CV algorithms in their news stories. For example, NYT used computer vision algorithms to estimate 3D poses of sports athletes at live events. Reuters used CV and satellite images to track the urban expansion in the South China Sea. On the other hand, Reuter has used this subfield of AI to help in the search of its video archive. The Chinese Xinhua News Agency is using CV combined with other subfields of AI to rebuild its newsroom to emphasize human-machine collaboration and producing real-time stories.

In another vein, CV is being used by news organizations to detect fake images. In an event promoted by Google, Asian practitioners developed an app called Source, powered by Storyful. This app uses Google's AI technology to give access to an image's public history, allowing one to understand its provenance and any sort of manipulation.

4.4. Planning, Scheduling, and Optimization in News Media

In the subfield of planning, scheduling, and optimization, automated stories are commonly found. The process involves running data through an algorithm that organizes that data into a readable story. Therefore, the use of algorithms to plan, publish and refine stories is usually implemented to produce stories that are data-driven, such as crimes, earthquakes, elections, finance, and sports. After the success of LA Times' QuakeBot that has a write-up within minutes of an earthquake, newsrooms started to embrace the automation of stories (Salaverría and de-Lima-Santos 2020).

In general, the narrative structures are repetitive, allowing automation (Carlson 2015; Dörr 2016; Graefe 2016; van Dalen 2012). For example, AP and Newsday automated the coverage of 124 school districts in the US and The Washington Post has published 850 automated articles in 2016. These examples highlight the potential that AI brings to news production, allowing producing more stories, using less human resources (Broussard et al. 2019). On the flip side, there is a pile of ethical and quality concerns raised by these automated stories (Guzman and Lewis 2020).

During the COVID-19 outbreak, news media saw the opportunity to automate their production processes, as global death toll and infection rates were structured data that can fit into predictable story frames (Danzon-Chambaud 2021). For example, The Times (UK) automated a powerful charting tool to build graphics of the coronavirus coverage. In BBC, a project called Salco (Semi-Automated Local Content) was released to generate over 100 unique stories per month, allowing its coverage to focus on local audiences that could learn about their hospital's performance. These examples bring another level of AI systems which includes other subfields, such as NLG and Computer Vision.

5. Discussion and Conclusions

Overall, this study argues that AI can take different forms in the news industry. Our findings show three major subfields are more present in the news ecosystem: machine learning, planning, scheduling, & optimization, and computer vision. Machine learning is used in different parts of the news workflow. However, two applications were commonly found in our cases. First, there is a great interest in boosting public engagement using machine learning recommendation engines. Second, news outlets use machine learning models to adjust their business strategies to individual readers. For example, it is used to predict subscription cancellation or build paywalls that bend to the individual reader. Thus, machine learning algorithms were used to strengthen news media's business model and boost revenue streams. In line with previous studies, third-party organizations build these solutions and sell them to newsrooms, such as Piano in the US and Deep BI in the UK (Cook et al. 2021). In a similar vein, big tech platforms, mainly Google, are providing some of these solutions, such as Jigsaw, a tool used to help community managers manage toxic comments or posts that might violate community guidelines (Rashidian 2020).

In the subfield of planning, scheduling, and optimization, automated journalism governs. Although journalism is related to textual content, our findings suggest that NLP models are less used in the industry than planning, scheduling, and optimization. We speculate that this might have associated with the fact that it is not easy to replicate NLP models in different languages, such as Portuguese (Rodrigues, Oliveira, and Gomes 2014). Furthermore, automated journalism constitutes a basic application of computational models, which in many cases, are used to fill in the blanks of template stories, instead of adopting machine learning or NLP approaches (Biswal and Gouda 2020).

Computer vision is a subfield of AI that helps practitioners to deal with visual content in different ways. So far, most of the cases deal with CV as a tool for investigative reporting, including fact-checking. In this aspect, CV seems limited to be a one-goal project, which hampers its application on a large scale. Although we could identify small news outlets relying on computer vision algorithms to produce investigative stories, the majority came from large newsrooms. We understand that, as CV requires technological

infrastructure, qualified personnel to develop such codes, and a significant investment, only large newsrooms can afford it (de-Lima-Santos and Salaverría 2021). On the other hand, we found that most of the applications using AI in the news industry rely on grants from big tech companies, such as Google and Facebook (Rashidian et al. 2018; Rashidian 2020). This brings serious challenges to the development of technological innovations in the news media, as these organizations decide to whom, where, and when the money goes.

We were lacking examples that use social bots by these news companies. Most of the bots were news bots that write stories. These social bots are the easiest application of technological assistance in news production and dissemination (DalBen and Jurno 2021; Lokot and Diakopoulos 2016). However, we recognize that there are two reasons why they did not show up on this list. First, social bots, for example on Twitter, are not necessarily using AI. We speculate that this might be the reason social media bots did not show up in the list of cases. Second, this list is not extensive enough, which might leave some examples of AI applications in the news industry out. Because of this potential limitation, we treat the study as an initial approach to the subject. Future research could explore this topic, especially from practitioners' point of view. Furthermore, it would be interesting to have more ethnographic studies about the development of AI in newsrooms, shedding light on limitations and hurdles to deploy AI algorithms in the field.

In conclusion, we were able, despite the limitations discussed, to highlight the different uses of the subfields of AI in the news industry. Our study contributes to scholarly literature by stressing the limits and opportunities of AI in news media and giving inputs for practitioners to expand its applicability.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER, grant number XXX" and "The APC was funded by XXX". Check carefully that the details given are accurate and use the standard spelling of funding agency names at https://search.crossref.org/funding. Any errors may affect your future funding.

Acknowledgments: In this section, you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: The authors declare no conflict of interest.

References

References must be arranged in alphabetical order of the first author in the Reference list and listed individually at the end of the manuscript. We recommend preparing the references with a bibliography software package, such as EndNote, ReferenceManager or Zotero to avoid typing mistakes and duplicated references. Include the digital object identifier (DOI) for all references where available.

Citations and references in the Supplementary Materials are permitted provided that they also appear in the reference list here.

(Aranceta-Bartrina 1999a) Aranceta-Bartrina, Javier. 1999a. Title of the cited article. *Journal Title* 6: 100–10.

(Aranceta-Bartrina 1999b) Aranceta-Bartrina, Javier. 1999b. Title of the chapter. In *Book Title*, 2nd ed. Edited by Editor 1 and Editor 2. Publication place: Publisher, vol. 3, pp. 54–96.

(Baranwal and Munteanu [1921] 1955) Baranwal, Ajay K., and Costea Munteanu. 1955. *Book Title*. Publication place: Publisher, pp. 154–96. First published 1921 (optional).

(Berry and Smith 1999) Berry, Evan, and Amy M. Smith. 1999. Title of Thesis. Level of Thesis, Degree-Granting University, City, Country. Identification information (if available).

(Cojocaru et al. 1999) Cojocaru, Ludmila, Dragos Constatin Sanda, and Eun Kyeong Yun. 1999. Title of Unpublished Work. *Journal Title*, phrase indicating stage of publication.

(Driver et al. 2000) Driver, John P., Steffen Röhrs, and Sean Meighoo. 2000. Title of Presentation. In *Title of the Collected Work* (if available). Paper presented at Name of the Conference, Location of Conference, Date of Conference.

Appelgren, Ester, and Gunnar Nygren. 2014. "Data Journalism in Sweden." *Digital Journalism* 2 (3): 394–405. https://doi.org/10.1080/21670811.2014.884344.

Aronson, Jay D. 2018. "Computer Vision and Machine Learning for Human Rights Video Analysis: Case Studies, Possibilities, Concerns, and Limitations." *Law and Social Inquiry* 43 (4): 1188–1209. https://doi.org/10.1111/lsi.12353.

Bastos, Marco, and Dan Mercea. 2018. "The Public Accountability of Social Platforms: Lessons from a Study on Bots and Trolls in the Brexit Campaign." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376 (2128). https://doi.org/10.1098/rsta.2018.0003.

Biswal, Santosh Kumar, and Nikhil Kumar Gouda. 2020. "Artificial Intelligence in Journalism: A Boon or Bane?," 155–67. https://doi.org/10.1007/978-981-15-0994-0_10.

Boczkowski, P J. 2004. *Digitizing the News: Innovation in Online Newspapers*. Inside Technology. Cambridge and London: MIT Press.

Borges-Rey, Eddy. 2016. "Unravelling Data Journalism: A Study of Data Journalism Practice in British Newsrooms." *Journalism Practice* 10 (7): 833–43. https://doi.org/10.1080/17512786.2016.1159921.

Broussard, Meredith. 2018. *Artificial Unintelligence : How Computers Misunderstand the World*. 1st ed. Cambridge: MIT Press.

Broussard, Meredith, Nicholas Diakopoulos, Andrea L. Guzman, Rediet Abebe, Michel Dupagne, and Ching-Hua Hua Chuan. 2019. "Artificial Intelligence and Journalism." *Journalism & Mass Communication Quarterly* 96 (3): 673–95. https://doi.org/10.1177/1077699019859901.

Carlson, Matt. 2015. "The Robotic Reporter: Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority." *Digital Journalism* 3 (3): 416–31. https://doi.org/10.1080/21670811.2014.976412.

Castro, Daniel, and Joshua New. 2016. "The Promise of Artificial Intelligence." Washington, DC and Brussels. https://www2.datainnovation.org/2016-promise-of-ai.pdf.

Chan-Olmsted, Sylvia M. 2019. "A Review of Artificial Intelligence Adoptions in the Media Industry." *International Journal on Media Management* 21 (3–4): 193–215. https://doi.org/10.1080/14241277.2019.1695619.

Coddington, Mark. 2015. "Clarifying Journalism's Quantitative Turn." *Digital Journalism* 3 (3): 331–48. https://doi.org/10.1080/21670811.2014.976400.

Cook, Clare, Emiliana Garcia, Heghine Gyulnazaryan, Juan Melano, Jakub Parusinski, and Alex Sabadan. 2021. "The next Wave of Disruption: Emerging Market Media Use of Artificial Intelligence and Machine Learning." Edited by Robert Shaw. 1st ed. Copenhagen.

DalBen, Silvia, and Amanda Chevtchouk Jurno. 2021. "More than Code: The Complex Network That Involves Journalism Production in Five Brazilian Robot Initiatives Cite This Paper Related Papers." *ISOJ* 11 (1): 111–37.

Dalen, Arjen van. 2012. "The Algorithms behind the Headlines: How Machine-Written News Redefines the Core Skills of Human Journalists." *Journalism Practice* 6 (5–6): 648–58. https://doi.org/10.1080/17512786.2012.667268.

Danzon-Chambaud, Samuel. 2021. "Covering COVID-19 with Automated News." Columbia Journalism Review. August 6, 2021. https://www.cjr.org/tow_center_reports/covering-covid-automated-news.php.

de-Lima-Santos, Mathias-Felipe, and Lucia Mesquita. 2021. "Data Journalism Beyond Technological Determinism." *Journalism Studies* 22 (11): 1416–35. https://doi.org/10.1080/1461670X.2021.1944279.

de-Lima-Santos, Mathias-Felipe, and Ramon Salaverría, 2021. "From Data Journalism to Artificial Intelligence: Challenges Faced by La Nación in Implementing Computer Vision in News Reporting" News Reporting. *Palabra Clave*, 24(3), e2437.

https://doi.org/10.5294/pacla.2021.24.3.7

Deloitte. 2014. "Demystifying Artificial Intelligence." NYC.

Diakopoulos, Nicholas. 2020. "Computational News Discovery: Towards Design Considerations for Editorial Orientation Algorithms in Journalism." *Digital Journalism* 8 (7): 945–67. https://doi.org/10.1080/21670811.2020.1736946.

Dörr, Konstantin Nicholas. 2016. "Mapping the Field of Algorithmic Journalism." *Digital Journalism* 4 (6): 700–722. https://doi.org/10.1080/21670811.2015.1096748.

Erdal, Ivar John. 2011. "Coming to Terms with Convergence Journalism: Cross-Media as a Theoretical and Analytical Concept." *Convergence: The International Journal of Research into New Media Technologies* 17 (2): 213–23. https://doi.org/10.1177/1354856510397109.

Ferrer-Conill, Raul, and Edson C. Tandoc. 2018. "The Audience-Oriented Editor: Making Sense of the Audience in the Newsroom." *Digital Journalism* 6 (4): 436–53. https://doi.org/10.1080/21670811.2018.1440972.

Future Today Institute. 2018. "2019 Industry Trends: Journalism, Media, Technology." https://doi.org/10.1080/21670811.2018.1498747.

Gage, Justin. 2020. "What's GPT-3? - Technically." Technically. July 20, 2020. https://technically.substack.com/p/whats-gpt-3.

Gelgel, Ni Made Ras Amanda. 2020. "Will Technology Take over Journalism?" *Informasi* 50 (2): v–x. https://doi.org/10.21831/informasi.v50i2.36379.

Graefe, Andreas. 2016. "Guide to Automated Journalism." https://doi.org/https://doi.org/10.7916/D80G3XDJ.

Guzman, Andrea L., and Seth C. Lewis. 2020. "Artificial Intelligence and Communication: A Human–Machine Communication Research Agenda." *New Media & Society* 22 (1): 70–86. https://doi.org/10.1177/1461444819858691.

Harvard, Jonas. 2020. "Post-Hype Uses of Drones in News Reporting: Revealing the Site and Presenting Scope." *Media and Communication* 8 (3): 85–92. https://doi.org/10.17645/mac.v8i3.3199.

Hassaballah, Mahmoud, and Ali Ismail Awad. 2020. *Deep Learning in Computer Vision*. Edited by Mahmoud Hassaballah and Ali Ismail Awad. First edition. | Boca Raton, FL : CRC Press/Taylor and Francis, 2020. |: CRC Press. https://doi.org/10.1201/9781351003827.

Hermida, Alfred, and Mary Lynn Young. 2019. *Data Journalism and the Regeneration of News*. Edited by Alfred Hermida and Mary Lynn Young. 1st ed. London: Routledge. https://doi.org/10.4324/9781315163895.

Hernandez Serrano, Maria Jose, Anita Greenhill, and Gary Graham. 2015. "Transforming the News Value Chain in the Social Era: A Community Perspective." Edited by Dr. Gary Graham. *Supply Chain Management: An International Journal* 20 (3): 313–26. https://doi.org/10.1108/SCM-05-2014-0147.

Jamil, Sadia. 2020. "Artificial Intelligence and Journalistic Practice: The Crossroads of Obstacles and Opportunities for the Pakistani Journalists." *Journalism Practice*, July, 1–23. https://doi.org/10.1080/17512786.2020.1788412.

JournalismAI. n.d. "Case Studies." JournalismAI. Accessed September 30, 2021. https://www.lse.ac.uk/media-and-communications/polis/JournalismAI/Case-studies.

Kang, Seok, Erin O'Brien, Arturo Villarreal, Wansoo Lee, and Chad Mahood. 2019. "Immersive Journalism and Telepresence." *Digital Journalism* 7 (2): 294–313. https://doi.org/10.1080/21670811.2018.1504624.

Krumsvik, Arne H, Stefania Milan, Niamh Ní Bhroin, and Tanja Storsul. 2019. "Making (Sense of) Media Innovations." In *Making Media*, edited by Mark Deuze and Mirjam Prenger, 1st ed., 193–206. Amsterdam: Amsterdam University Press. https://doi.org/10.2307/j.ctvcj305r.16.

Lewis, Seth C., Andrea L. Guzman, and Thomas R. Schmidt. 2019. "Automation, Journalism, and Human–Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News." *Digital Journalism* 7 (4): 409–27. https://doi.org/10.1080/21670811.2019.1577147.

Linden, Carl-Gustav. 2017a. "Algorithms for Journalism: The Future of News Work." *The Journal of Media Innovations* 4 (1): 60–76. https://doi.org/10.5617/jmi.v4i1.2420.

———. 2017b. "Decades of Automation in the Newsroom." *Digital Journalism* 5 (2): 123–40. https://doi.org/10.1080/21670811.2016.1160791.

Locker, Mic, Jeff Loucks, Susanne Hupfer, and David Jarvis. 2019. "Seasoned Explorers: How Experienced TMT Organizations Are Navigating AI." New York.

Lokot, Tetyana, and Nicholas Diakopoulos. 2016. "News Bots: Automating News and Information Dissemination on Twitter." *Digital Journalism* 4 (6): 682–99. https://doi.org/10.1080/21670811.2015.1081822.

Marr, D. 2010. *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. The MIT Press. MIT Press.

Mutsvairo, Bruce. 2019. "Challenges Facing Development of Data Journalism in Non-Western Societies." *Digital Journalism* 7 (9): 1289–94. https://doi.org/10.1080/21670811.2019.1691927.

Nelson, Jacob L., and Edson C. Tandoc. 2018. "Doing 'Well' or Doing 'Good': What Audience Analytics Reveal About Journalism's Competing Goals." *Journalism Studies*. https://doi.org/10.1080/1461670X.2018.1547122.

Örnebring, Henrik. 2010. "Technology and Journalism-as-Labour: Historical Perspectives." *Journalism* 11 (1): 57–74. https://doi.org/10.1177/1464884909350644.

Ortiz Freuler, J., and C. Iglesias. 2018. "Algorithms e Inteligencia Artificial En Latin America: Un Estudio de Implementaciones Por Parte de Gobiernos En Argentina y Uruguay." www.webfoundation.org.

Parasie, Sylvain, and Eric Dagiral. 2013. "Data-Driven Journalism and the Public Good: 'Computer-Assisted-Reporters' and 'Programmer-Journalists' in Chicago." *New Media & Society* 15 (6): 853–71. https://doi.org/10.1177/1461444812463345.

Paulussen, Steve. 2016. "Innovation in the Newsroom." In *The SAGE Handbook of Digital Journalism*, edited by Tamara Witschge, C.W. Anderson, David Domingo, and Alfred Hermida, 1st ed., 192–206. 1 Oliver's Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd. https://doi.org/10.4135/9781473957909.n13.

Rashidian, Nushin. 2020. "Platforms and Publishers: The Great Pandemic Funding Push." Columbia Journalism Review. December 17, 2020. https://www.cjr.org/tow_center_reports/platforms-publishers-pandemic-funding-news.php.

Rashidian, Nushin, Pete Brown, Elizabeth Hansen, Emily Bell, Jonathan Albright, and Abigail Hartstone. 2018. "Friend and Foe: The Platform Press at the Heart of Journalism." Columbia Journalism Review. 2018. https://www.cjr.org/tow_center_reports/the-platform-press-at-the-heart-of-journalism.php.

Rodrigues, Ricardo, Hugo Gonçalo Oliveira, and Paulo Gomes. 2014. "Lemport: A High-Accuracy Cross-Platform Lemmatizer for Portuguese." *OpenAccess Series in Informatics* 38: 267–74. https://doi.org/10.4230/OASIcs.SLATE.2014.267.

Russell, Stuart, and Peter Norvig, eds. 2021. *Artificial Intelligence: A Modern Approach*. 4th ed. Harlow, UK: Pearson. https://www.pearson.com/us/higher-education/program/Russell-Artificial-Intelligence-A-Modern-Approach-4th-Edition/PGM1263338.html.

Salaverría, Ramón, and Mathias-Felipe de-Lima-Santos. 2020. "Towards Ubiquitous Journalism: Impacts of IoT on News." In *Journalistic Metamorphosis: Media Transformation in the Digital Age Volume 70 of Studies in Big Data*, edited by Jorge Vázquez-Herrero;, Sabela Direito-Rebollal;, Alba Silva-Rodríguez;, and Xosé López-García, 1st ed., 1–15. Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-36315-4_1.

Stray, Jonathan. 2019. "Making Artificial Intelligence Work for Investigative Journalism." *Digital Journalism* 7 (8): 1–22. https://doi.org/10.1080/21670811.2019.1630289.

Szeliski, Richard. 2011. *Computer Vision*. Texts in Computer Science. London: Springer London. https://doi.org/10.1007/978-1-84882-935-0.

Whittaker, Jason. 2019. *Tech Giants, Artificial Intelligence, and the Future of Journalism*. *Tech Giants, Artificial Intelligence, and the Future of Journalism*. Routledge. https://doi.org/10.4324/9781351013758.

Young, Mary Lynn, and Alfred Hermida. 2015. "From Mr. and Mrs. Outlier To Central Tendencies." *Digital Journalism* 3 (3): 381–97.

https://doi.org/10.1080/21670811.2014.976409.