Check for updates

From Data Journalism to Artificial Intelligence: Challenges Faced by *La Nación* in Implementing Computer Vision in News Reporting*

Mathias Felipe de Lima Santos[1]
Ramón Salaverría[2]

Recibido: 09/10/2020
Aprobado por pares: 11/08/2021

Enviado a pares: 13/01/2021
Aceptado: 09/09/2021

DOI: 10.5294/pacla.2021.24.3.7

To reference this article / para citar este artículo / para citar este artigo
de-Lima-Santos, M-F., & Salaverría, R. (2021). From Data Journalism to Artificial Intelligence: Challenges Faced by *La Nación* in Implementing Computer Vision in News Reporting. *Palabra Clave*, *24*(3), e2437. DOI: https://doi.org/10.5294/pacla.2021.24.3.7

Abstract

Journalism is at a radical point of change that requires organizations to come up with new ideas and formats for news reporting. Additionally, the notable surge of data, sensors and technological advances in the mobile segment has brought immeasurable benefits to many fields of journalistic practice (data journalism in particular). Given the relative novelty and complexity of implementing artificial intelligence (AI) in journalism, few areas have managed to deploy tailored AI solutions in the media industry. In this study, through a mixed-method approach that combines both participant observations and interviews, we explain the hurdles and obstacles to deploying

* This project was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska Curie grant agreement No 765140.

1 ✉https://orcid.org/0000-0001-8879-7381. Universidad de Navarra, Spain. mdelimas@unav.es
2 https://orcid.org/0000-0002-4188-7811. Universidad de Navarra, Spain. rsalaver@unav.es

computer vision news projects, a subset of AI, in a leading Latin American news organization, the Argentine newspaper *La Nación*. Our results highlight four broad difficulties in implementing computer vision projects that involve satellite imagery: a lack of high-resolution imagery, the unavailability of technological infrastructure, the absence of qualified personnel to develop such codes, and a lengthy and costly implementation process that requires significant investment. This article concludes with a discussion of the centrality of AI solutions in the hands of big tech corporations.

Keywords (Source: Unesco Thesaurus)

Computer vision; artificial intelligence; machine learning; data journalism; *La Nación*; technology; journalism; computer science; advanced technology.

Del periodismo de datos a la inteligencia artificial: desafíos que enfrenta *La Nación* en la implementación de la visión artificial para la producción de noticias*

Resumen

El periodismo se encuentra en un punto de cambio radical que exige a las organizaciones desarrollar nuevas ideas y nuevos formatos para la presentación de noticias. Además, el aumento de los datos, los sensores y los avances tecnológicos, especialmente en el segmento móvil, han traído beneficios inconmensurables a muchos campos de la práctica periodística, en particular al periodismo de datos. Dada la relativa novedad y complejidad de la implementación de la inteligencia artificial (IA) en el periodismo, pocas áreas han logrado implementar hasta ahora soluciones de IA personalizadas en la industria de los medios. En este estudio, a través de un enfoque de método mixto que combina la observación participante y la entrevista, se explican los obstáculos y dificultades de implementar proyectos de noticias mediante la visión artificial, un subconjunto de la IA, en una organización de noticias líder en América Latina como el diario argentino *La Nación*. Los resultados destacan cuatro grandes dificultades para implementar la IA, más específicamente los proyectos de visión artificial que involucran el uso de imágenes satelitales: la falta de imágenes de alta resolución, la falta de disponibilidad de infraestructura tecnológica, la ausencia de personal cualificado para desarrollar dichos códigos y un proceso de implementación prolongado y costoso que requiere de una inversión significativa. El artículo concluye con una discusión sobre la centralidad de las soluciones de IA en manos de las grandes corporaciones tecnológicas.

Palabras clave (Fuente: Unesco Thesaurus)

Visión artificial; inteligencia artificial; aprendizaje automático; periodismo de datos; *La Nación*; tecnología; periodismo; ciencias de la computación; tecnología avanzada.

* Este proyecto cuenta con la financiación del programa de investigación e innovación Horizon 2020 de la Unión Europea bajo el acuerdo de subvención Marie Sklodowska Curie n.º 765140.

Do jornalismo de dados à inteligência artificial: desafios enfrentados pelo *La Nación* na implementação da visão computacional para a produção de notícias*

Resumo

O jornalismo está em um ponto radical de mudança que exige que as organizações abracem novas ideias e novos formatos para a apresentação de notícias. Além disso, o aumento de dados, sensores e avanços em tecnologia, em particular no segmento móvel, trouxe benefícios imensuráveis para a prática jornalística e, principalmente, para o jornalismo de dados. Dado o seu carácter inovador e a complexidade de implementar a inteligência artificial (IA) em jornalismo, poucas áreas conseguiram implantar soluções de IA personalizadas na indústria de mídia. Neste estudo, a partir de uma abordagem mista, combinando observação participante e entrevistas, explicamos os obstáculos e os desafios para implementar projetos noticiosos por meio da visão computacional, um subconjunto da IA, em uma organização de notícias líder na América Latina, o jornal *La Nación* (Argentina). Os resultados mostram quatro amplas dificuldades para implementar a IA, especialmente projetos de visão computacional que envolvem imagens de satélite: a falta de imagens de alta resolução, a indisponibilidade de infraestrutura tecnológica, a ausência de pessoal qualificado para desenvolver tais códigos e um longo e caro processo de implementação que requer um investimento significativo. O artigo conclui com uma discussão sobre a centralidade das soluções de IA nas mãos de grandes corporações de tecnologia.

Palavras-Chave (Fonte: Unesco Thesaurus)

Visão computacional; inteligência artificial; aprendizado de máquina; jornalismo de dados; *La Nación*; tecnologia; jornalismo; ciência da computação; tecnologia avançada.

* Este projeto foi financiado pelo programa de pesquisa e inovação Horizon 2020 da União Europeia, com o patronício Marie Sklodowska Curie n.o 765140.

Introduction

The pervasive nature of information and communications technology (ICT) in the news and the datafication of society have expanded opportunities to use new formats to produce content (Knight, 2015; Lewis & Usher, 2013). Through improved access to data and metadata, journalists have found a logical step: adapting the field to the emergence of a new journalistic sub-field, commonly referred to as data journalism. In this way, the practice can assist political watchdog reporting (Carson & Farhall, 2018). Data can be processed to derive valuable and meaningful information that assists in analyzing and reporting issues that impact the public. In the beginning, data journalism was very Western-centric since it engaged mostly elite news outlets from countries in the Northern Hemisphere, such as the United States, the United Kingdom, the Nordic countries, and Germany (Anderson, 2019; Appelgren & Nygren, 2014; Borges-Rey, 2016; Coddington, 2015; De Maeyer et al., 2015; Felle, 2016; Fink & Anderson, 2015; Karlsen & Stavelin, 2014; Lewis & Westlund, 2015; Stalph, 2018). In recent years, scholars have become more interested in data journalism, which has stimulated a large body of literature on the evolution of data reporting in the Global South (Mutsvairo et al., 2019).

In the mythos of data journalism, practitioners were empowered by the superior computational prowess of the 21st century since it allowed them to move beyond the age of big data to the era of artificial intelligence (AI), when principles of computing and data science became ever more pervasive in journalism (Borges-Rey, 2021).

In reality, few cases outside the Western world have achieved success and recognition in both the journalistic and academic communities; one such case is the Argentine newspaper *La Nación* (Mazotte, 2017; Palomo et al., 2019). Despite a lack of publicly available statistical data affecting most countries in the Global South, specifically Latin America (Salaverría, 2016, p. XVI), *La Nación* has been able to innovate and engage its audience through news production. The effective participation of the audience in the news process has ensured a gold standard for the outlet's data journalism department, which has led the team to gain internal and external recognition (Palomo et al., 2019).

Founded in 2011, the team responsible for data journalism projects, *La Nación Data (LN Data),* has also experimented with open data, civic and public journalism (Lambeth, 1998), and automated journalism (Carlson, 2016) to enhance readers' experiences and create more value for the audience. Journalism is at a radical point of change (Franklin, 2014; Picard, 2014): it requires organizations to develop new ideas and formats for news reporting. To envision the future of the profession, *LN Data* has continuously searched for novel ideas to engage with and present information to its audience. Despite being recognized for its forward-thinking data skills, *La Nación* is still restricted by the fact of being a legacy media organization located in a country continuously shaken by economic and political crises. As with many other Latin American media outlets, resources are scarce, and innovation is gradual (Salaverría & de-Lima-Santos, 2021). It is also worth noting that incremental innovation tends to prevail in more mature technological markets, resulting in improved processes and products instead of radical change (Eisenman, 2013; Verganti, 2006).

To attempt to burst this bubble of incremental innovation in newsrooms (Paulussen, 2016), *LN Data* has analyzed ways to use AI techniques in journalism production. AI uses algorithms to enable machines to learn from experiences, adjust to new inputs, and perform human-like tasks. Amongst the several fields in which AI is utilized, computer vision (CV) focuses on image processing. Contrary to what the name suggests, CV models cannot see the content of an image like a human eye can (Marr, 2010). Instead, they use mathematical algorithms to deduce what content is being shown (Szeliski, 2011). By using a computer to classify and organize a plethora of images and videos quickly, CV can accelerate the editing process and "enable journalists to source evidence for investigative pieces" (Marconi et al., 2017, p. 14). With this idea in mind, *LN Data* designed a CV-based project to map solar farms in full expansion in Argentina that are in line with the Paris agreement to reduce their carbon footprints and adopt 100-percent renewable energy to limit the increase of the global average temperature to below 2 °C (Masson-Delmotte et al., 2018).

Drawing upon literature on management, journalism, and computer science theories, we explore the hurdles inherent in executing a CV project

that is rapidly evolving in the machine-learning space. In comparison to other applications of AI in journalism, this area is still largely unexplored since much of the literature concerns the impact of automated journalism on news reporting (Salaverría & de-Lima-Santos, 2020), which often involves a model that fills in the blanks of pre-written template stories with no machine-learning model behind it (Biswal & Gouda, 2020).

To understand this phenomenon, we relied on an experiment conducted by *La Nación* between January and March 2020. The potential of qualitative research methods to effectively draw out insights depends on the researcher's privileges inside the studied organization. During an ethnographic study, participant observation allows the researcher to meet with and speak to all levels of employees. Thus, we conducted a participant observation for this study; the first author was able to work alongside *La Nación*'s data team at its office in Buenos Aires, Argentina.

The observation data was then triangulated with the ancillary data from interviews, which validated the collected data and allowed us to check for discrepancies in reporting (O'Donoghue, 2003). Therefore, this study followed a multi-method approach since it combined both participant observation and interviews. We conducted this ethnographic study that explores the challenges *La Nación*'s data department faces to introduce new technological skills to their production of data stories. In the end, our ultimate goal was to answer the following research question: What challenges does *LN Data* face in adopting CV, a form of AI, for journalistic purposes?

The contributions of this article are twofold. First, this study proposes a practical reflection based on the application of CV to journalism, which contributes not only to the academic understanding of the field but also offers insights into practice that can help journalists advance their abilities. Second, this article adds to the scholarly work of journalism by shedding light on the emerging challenges and threats to deploying cutting-edge technological projects and focusing on groups in the Global South, such as *La Nación*.

Finally, this work underlines the possible challenges news organizations face when deploying AI projects in their newsrooms while present-

ing the practical phases of implementing satellite imagery, CV, or imagery analysis projects in journalism.

Literature review

Transformation: Bringing dynamic capabilities and new corporate strategies

Digital technologies have catalyzed significant shifts and market disruption in the media industry (Oliver, 2018; Teece, 2010). In this time of complex change, most media companies fail to profit from digital products, resulting in a high level of risk and uncertainty involving the reconfiguration of resources and capabilities (Paulussen, 2016).

To achieve their goals, media organizations must adapt their strategies to this massive technological shift and rely on a series of processes involving technology in news production. The diffusion of innovations plays a vital role in achieving sustained development, which is only possible through a successful process of transformation that requires new capabilities due to the largely unknown technologies that have emerged in the scene. Thus, organizations must possess an integrated perspective for managing innovation and change, which are two emergent components of strategic management (Gade, 2004; Oliver, 2018; Porcu, 2020). The management of innovation processes in the industry serves to trigger and deploy a new set of tools to facilitate the different stages of this evolutionary process (Tidd & Bessant, 2009).

In journalism, this process is usually conservative and slow (Kueng, 2017; Lischka, 2019) due to the "yes, but" syndrome (YBS), which limits new digital strategies because of economic challenges and other excuses such as a lack of resources and staff, economic viability, or the risk of failure (Paulussen, 2016). However, innovation is not a temporary action but a long-term strategy. It also involves offering new ways of engaging established and mature audiences beyond the technological determinism that characterizes the discourse about information and communication technology (ICT) (Verganti, 2006). In this matter, journalism has been extended

to include new ways of reporting and narrating stories, such as blogging, citizen journalism, virtual and augmented realities, data, and collaboration, and bring profitability to digital business models. Furthermore, a new revolution is upon us and promises to bring other technologies for news reporting, such as open data, automation, and AI. Yet, the adoption of these strategies currently remains restricted mainly to media conglomerates that have more financial resources.

On the other hand, data journalism is perceived as a commercial opportunity to innovate and "create a distinctive online offering" (Usher, 2016, p. 142) that, in the past, was considered a complicated endeavor that required extensive time, knowledge, and skills. Data journalism has evolved and expanded due to today's wealth of information (Loosen, 2021; Loosen et al., 2020) and the evolving infrastructure of technological tools (de-Lima-Santos et al., 2020). These developments have attracted significant attention not just across the journalism industry but also within scholarship. There is a strong interest, particularly in the Global South, to better understand the challenges and possibilities of data journalism's development in different social and developmental contexts (Mutsvairo, 2019; Mutsvairo et al., 2019).

Data journalism: Challenges and opportunities in the Global South

Until recently, little was known about the development of data journalism beyond Western societies; much of the academic literature and practical cases were concentrated in early-adopter nations such as the US, the UK, and the Nordic countries (Anderson, 2019; Appelgren & Nygren, 2014; Borges-Rey, 2016; Felle, 2016; Fink & Anderson, 2015; Karlsen & Stavelin, 2014; Lewis & Westlund, 2015; Stalph, 2018). Over the years, the practice has noticeably expanded and evolved from a niche to a mainstream audience because of the rapid advances in technological development and the increasing availability of data (Loosen, 2021; Loosen et al., 2020). However, major concerns relate to challenges in data journalism since this practice requires reporters to possess novel skill sets to adopt it in newsrooms in the Global South.

The origin of data journalism is rooted in computer-assisted reporting (CAR), which dates back to 1973 and is based on the work of Philip Meyer (2002). By combining surveys, content analyses, and statistics, Meyer validated journalists' questions and embraced a new form of journalistic reporting (McGregor, 2013). This required a new level of literacy, writing skills and numeracy, knowledge of technology, and an analytical mindset, which might explain the dominance of the US, the UK, and the Nordic countries in data journalism. These nations have historically emphasized transparency, access to information, and digitalization to ensure open access to their governments and freedom of the press (Appelgren & Salaverría, 2018; Berliner, 2014).

In contrast, the Global South has focused on immediate and local changes such as reducing illiteracy and alleviating poverty. This developing section faces different challenges in increasing their efficiency and using working models such as those already commonly used in Western democracies (Mutsvairo et al., 2019). Nevertheless, the region is plural and diverse and features different levels of development, behavioral patterns, and attitudes (Milan & Treré, 2019). Recently, scholars have paid more attention to better understand this phenomenon with the particular nuances of each scenario.

In Asia, concerns revolve around the external and internal struggles that have challenged the evolution of the practice, such as "non-existent, inaccessible or unreliable data" (Kashyap et al., 2020, p. 128), as well as a lack of standardized governmental procedures and data literacy. Additionally, journalists face a growing risk when reporting on governments' social and economic policies or questioning the reliability of sources and data obtained through the government due to potential censorship in the region (Jamil, 2019). China, for example, also suffered from an initial lack of motivation that delayed the expansion of this practice (Zhang & Feng, 2019).

To overcome these problems and tackle structural impediments, African data journalists have joined forces with civic technologists to form a community to address the lack of skills and complementary ambitions to

create a participatory culture and "therefore more directly engage in those practices where data journalism is lacking" (Cheruiyot et al., 2019, p. 12). The realities of Latin American cities are not very different since a lack of open data culture (which lasted until the 2010s) and the scarcity of an "investigative reporting culture" overlapped with "high levels of crime and corruption" (Palomo et al., 2019, p. 1273). In Brazil, the digital news outlet from *favelas* found in citizen participation a way to help these marginalized communities produce their data to fill in the gaps of official data (de-Lima-Santos & Mesquita, 2021b).

Despite these limitations, the Argentine *La Nación* has been able to uncover meaningful data stories. The legacy news outlet was founded in 1870 and became an example of success in data storytelling. Their dedicated team, *LN Data,* has received significant recognition based on their efforts to overcome a lack of data and resources to tell stories. One example of their efforts is the civic hackathon during which *LN Data* asked for volunteers from the public to help analyze 4,800 verified public documents related to senate expenses between 2010 and 2012 (Palomo et al., 2019). Thus, *LN Data* bridged the gap between the public and journalism; the outlet automatically became more invested in what was covered thanks to the community's involvement in the storytelling process, which somehow attenuated the lack of staff and resources.

LN Data constantly innovates to ensure that its offerings are inclusive and sufficiently broad to support journalistic production. In this sense, the team develops multiple alternative solutions based on an understanding of the critical requirements of the team and the newsroom, including designing scripts and algorithms to automate tasks and solve computational problems and deploying in-house solutions to support the work of data and infographic teams. This aligns with findings from studies suggesting many experience a degree of frustration with generic third-party solutions (de-Lima-Santos et al., 2020) since they "were not compatible with the software infrastructure of their news outlets or were not for certain projects they pursued" (Borges-Rey, 2016, p. 840). In the next few years, investments will become essential to ensure that media companies can ful-

ly deploy technologies and innovations in newsrooms. Thus, data journalism has been pivotal in this transformation towards utilizing AI strategies in newsrooms, as discussed in the following section.

Artificial intelligence: A booming industry to be fully explored by journalism

The first use of the term "artificial intelligence" dates back to 1955, when John McCarthy, a professor at Stanford University, described the science and engineering of making intelligent machines (McCarthy, 1998). Since then, the term has spread widely, and no universal definition of AI currently exists. Instead, different definitions have been provided by experts and scholars who have tried to refine them over the years. AI can be simply described as the process of automation of cognitive tasks (Chan-Olmsted, 2019) or, in other words, as an area of knowledge that is "devoted to creating computing machines and systems that perform operations analogous to human learning and decision-making" (Castro & New, 2016, p. 2).

Numerous studies on AI have been part of computer science for decades, but AI has not been implemented effectively in other areas since it remains beyond the reach of technology (Aronson, 2018; Castro & New, 2016; Ortiz Freuler & Iglesias, 2018). This has changed recently due to countless developments relating to data, sensors, and technology, including the surge of the smartphone segment, which allowed data to be collated and stored in massive databases and moved across multiple devices using the Internet (Lewis, 2015). The ubiquity of computing has become apparent and brought immeasurable benefits to every field of study, especially science. However, according to some authors, AI is a disputed concept that has not yet been fully displayed in journalism; at best, what currently exists is simple automation that fills in the blanks of template stories (Biswal & Gouda, 2020). In Latin America, a recent study has shown an increase in the number of organizations embracing AI in their newsrooms, albeit to a limited extent, but many of these organizations come from countries with public AI strategies (Cook et al., 2021).

Given the relative novelty and complexity of the subject matter, few organizations have managed to deploy tailored solutions for their business

needs. Big tech companies such as Facebook, Amazon, Apple, Netflix, and Google leverage their control over AI by actively acquiring unknown companies developing AI solutions to concentrate more power (Linden, 2017a, 2017b). AI is beginning to pave its way into a diverse range of projects in the media industry, from services to products, including, for example, advanced AI software that automates fact-checking by analyzing vast amounts of data. For this, "new alliances inside newsrooms and with academia and technology are being forged" in Latin America (Cook et al., 2021, p. 30).

Automated processes have influenced the production, distribution, and curation of content. Thus, automation is one of the areas in which AI has become increasingly popular. In journalism, its proponents argue that automated journalism could free journalists of mundane, repetitive, and boring chores and allow them to focus on investigative journalism or other valuable tasks (Wu et al., 2019). Latin American news organizations lack hiring resources to invest in professionals specialized in the field, requiring greater reliance on third-party solutions, usually cheaper than those developed inside the newsrooms. However, these tools operate better in English than Portuguese or Spanish (Cook et al., 2021).

Nonetheless, AI is not limited to textual content. Recently, researchers have found promising areas that can be empowered by AI. For example, Diakopoulos (2015) presented challenges and opportunities related to utilizing automated journalism in newsrooms, while Graefe (2016) introduced the audience's perspective on automated texts. Linden (2017b) studied the media's response to automated content in the US and five other European countries. Chan-Olmsted (2019) took this further by demonstrating how AI can potentially disrupt the existing traditional workflow of a media company (on the supply side) and the audience (the demand side). In her review, she mentioned eight main functional areas that can adopt AI solutions in the media industry: content recommendations and discovery, audience engagement, augmented experiences, message optimization, content management, content creation, audience insights, and operational automation (Chan-Olmsted, 2019). In Latin America, AI can mainly be found in four fields: subscriptions, content recommendation, automated

content production, and automated data processing. These areas rely on textual data to feed AI algorithms (Cook et al., 2021).

Chan-Olmsted's (2019) findings demonstrate that AI will undoubtedly have a fundamental impact on global labor. For economic reasons, numerous activities will be carried out by intelligent software, which will increase the number of stories that the media can produce at a minimal cost with speed, accuracy, and scale (van Dalen, 2012). In 2020, the beta release of GPT-3, an evolved machine-learning model that generates text, took place. This new language-generation model can perform a great range of tasks such as summarization, article generation, and translation, making the future of AI more promising. Recent tests conducted with GPT-3 revealed that it could also complete many other tasks such as unscrambling words and writing poetry and other creative pieces. One can understand what GPT-3 does through text generation. When someone spends time with a close group of friends who have their own mannerisms and preferred topics of conversation, for instance, text generation might help that person predict what their friends will say next. GPT-3 focuses on making predictions (Gage, 2020).

These advances in AI-related technologies have the potential to disrupt the nature of human-machine interactions significantly. However, a clear understanding of the near- and long-term opportunities and risks of AI is still missing. AI-enforcement mechanisms are essential for ensuring that AI systems adhere to legal and ethical guidelines without explicitly considering the power structures between various stakeholders. For this reason, it is crucial to understand the different domains and sub-sets of AI.

Computer vision, a branch of artificial intelligence

AI uses algorithms —a step-by-step procedure for solving problems— to enable machines to learn from experiences, adjust to new inputs, and perform human-like tasks. Until recently, AI only worked at a limited capacity since computer scientists had to program a wide array of functions into a system to mimic human intelligence. In recent years, this AI capability has appeared due to better hardware, more data, better algorithms, and more

abundant storage as a result of the development of machine learning (ML) (Aronson, 2018; Castro & New, 2016; Hassaballah & Awad, 2020; Whittaker, 2019). "Machine learning is a powerful tool that allows humans to see things they otherwise could not when datasets get above a certain size" (Marconi, 2020); consequently, ML is a subset of AI that performs a task through a system without explicit instructions or with minimal assistance from programmers.

The development of ML is highly related to deep learning. In deep learning, algorithms use statistical techniques to solve problems through models based on large and complex datasets with little human intervention. In other words, deep learning replicates the human brain's learning capabilities and is inspired by the structure and function of neural networks (Chan-Olmsted, 2019; Hassaballah & Awad, 2020). These networks simulate "large, multilayered webs of virtual neurons, which enable a computer to learn to recognize abstract patterns" (Castro & New, 2016, p. 3).

Depending on the human-activity level, deep learning can be classified as supervised, semi-supervised, or unsupervised (Castro & New, 2016; Marconi, 2020). Supervised learning involves a machine-learning task trained to associate an object with the desired output. Therefore, supervised learning-based techniques require a large amount of human-annotated training data to learn an adequate model, and it is common in problems that involve image classification, image captioning, instance segmentation, visual question answering, and other tasks.

In contrast, unsupervised ML is free to deliver relationships between input and output without a target result. Usually, this task is performed by comparing similarities and deviations between the information found in a dataset (Marconi, 2020). Semi-supervised learning algorithms represent a middle ground between supervised and unsupervised methods. Semi-supervised learning labels a small amount of data while leaving another large part unlabeled during training. By falling between these two domains, semi-supervised learning requires less human interaction than unsupervised ML (Aronson, 2018; Marconi, 2020).

CV is a sub-set of ML, but it can also be considered a direct sub-set of AI (Aronson, 2018; Marr, 2010; Szeliski, 2011). CV analyzes digital image-processing through mathematical algorithms to deduce what content is being displayed (Szeliski, 2011). Additionally, CV is a multidisciplinary field that could be broadly called a sub-field of ML since it may involve specialized methods and learning algorithms. In journalism, CV can accelerate the editing process and enable journalists to source evidence for investigative pieces. So far, scholars have paid little attention to CV and its implications for journalism (Chan-Olmsted, 2019).

Method

To answer the research question (RQ), we relied on a qualitative research strategy using a multi-method design that utilized participant observation and in-depth interviews. Thus, this research aims to provide a more holistic and nuanced view of the challenges that newsrooms face in adopting AI, especially CV.

The newsroom observation consisted of about three months of fieldwork in Q1 2020 before the pandemic hit Latin America when the first author was on a secondment in the organization. On average, the first author spent 40 hours per week with the team. The observation included qualitative notes that the first author recorded during or after each day in the newsroom. Additionally, the first author took part in meetings that usually lasted 60 minutes. During these meetings and the period spent in the host institution, several internal and external individuals and organizations were mobilized around these themes, which allowed the first author to engage in further interactions with them. These stakeholders in the meetings offered other perspectives on environmental and technological issues that were useful for deploying this new project.

There are many ethnographic contexts in which the ethnographer's active participation is advantageous, if not essential, to the collection of quality data (Johnson et al., 2006, p. 111); for this research, this was the case. Since this research involved an innovative project for the data team, having an active role in a given ethnographic setting helped the researcher to better

understand the challenges and threats related to implementing cutting-edge technological projects in a legacy newsroom.

The obtained data were analyzed using the method proposed by Emerson et al. (2011), which consists of a three-step process. First, the first author closely and systematically examined the field notes to find patterns. Second, the authors conducted qualitative analytic coding in two phases: *open coding* that allowed the ethnographer to read the fieldnotes line-by-line and formulate all the themes and ideas in them and, afterward, *focused coding* that subjected the notes to a fine-grained analysis that used a smaller set of ideas and categories based on those themes. In the last phase, we wrote our findings as "tales" (Richardson, 1990). Unlike quantitative researchers who express their findings in tables and charts, qualitative researchers must depend upon words. The decision to report the findings as tales invigorated the world studied in all its complexities for the reader and incorporated the themes in a series of steps that "move progressively toward creating a thematic narrative that is fieldnote-centered" (Emerson et al., 2011, p. 202). This format has proved helpful in conveying information about observation data clearly and effectively (Lewis-Beck et al., 2004).

To conclude, observation data was then triangulated with ancillary data from the in-depth interviews. By listening to interviewees' recounts and making notes of the important events in the narratives, we generated reliable insights into the phenomenon under investigation, which enabled us to go beyond our initial observations to supplement missing information (Becker & Geer, 1957; Bryman, 2012; Kawulich, 2005). The interviews varied in length, but the typical interview lasted about 40–75 minutes. They were conducted to understand the scope of the project better. This method was valuable for understanding motivations and processes. The identities of the interviewees have been disclosed with their permission.

This work can be described as a hybrid ethnography (Usher, 2016). This method produces "a single, unified and complete description of the world" (Longino, 1999, p. 339 as cited in Charles, 2020) that stands alone and apart from others. It allows us to undergo a unique experience through the eyes of the key actors involved in the news production (Spradley, 2016).

Findings

Inspiration

It is essential to describe the preliminary work carried out before this idea was brought to the newsroom to explain why *La Nación* decided to experiment with CV in news production. The *LN Data* team comprises professionals who come from a wide range of backgrounds: the team includes a journalist, a technologist, a lawyer, and a librarian. Angélica "Momi" Peralta, a technologist, has been leading the team since its foundation. Her values reflect in the team's multicultural mindset. Their diverse backgrounds are one of the most substantial reasons for delivering tremendous results. *LN Data* is always open to alliances and collaborations with other professionals. Once a year, they invite a student from Northwestern University to complete a three-month internship with them. Similarly, the team also hosts students and researchers from other universities. Recently, some researchers found that this collaboration is also part of the company's culture and takes advantage of co-creation or crowdsourcing to support its journalistic production (Palomo et al., 2019).

A critical role in the team belongs to Florencia Coelho. She serves as the New Media Research and Training Manager and is responsible for finding new ideas and opportunities that suit the team's goals. Her job also consists of following international trends and understanding what could be useful for their newsroom. With this in mind, she understood that the team needed to begin experimenting with more advanced levels of technology application for news reporting, mainly from an AI perspective. However, as she related in an interview, "AI seemed to be a very interesting topic, but I did not understand anything about it before I saw it for the first time" (Florencia Coelho, *LN Data*, January 2020). Therefore, in 2019, she became a JSK Fellow at Stanford to fully devote her time learning about opportunities for applying AI to the news ecosystem.

Meanwhile, *LN Data* was inspired to report about the climate emergency after collaborating with Covering Climate Now from *Columbia Jour-*

nalism Review and the #6Dnow, a project promoted by Avina Foundation. For these projects, the team has covered environmental world news such as the fires in Australia and the Amazon Forest and Greta Thunberg's talks in international meetings (UN Summit, COP 25) while also trying to leverage local concerns about environmental issues such as glaciers melting in Patagonia and Antarctica (LN Data, 2020).

Driven by these climate change concerns, the renewable energy industry worldwide has experienced significant growth. In early 2016, during the mandate of former president Mauricio Macri, Argentina launched an ambitious program called *RenovAr*, which promised to take advantage of the country's abundant clean energy resources by promoting private renewable energy generation via an auction. *RenovAr*, an acronym for Argentina Renewable Energy Auctions, endeavored to increase renewable energy sources to 20 percent of the national energy matrix by 2025 (IFC, 2018).

The project ensures the country some points in the ratified Paris Agreement, the landmark accord signed in 2016 that brought together 189 parties across the world to combat climate change. Based on the International Renewable Energy Agency's (IRENA) findings from multiple studies, renewable energy has the potential to meet 90 percent of the Paris Agreement's energy-related emission reduction goals (UNFCCC, 2017). Argentina possesses vast natural resources for creating clean energy, especially wind and solar power, allowing the country to shift towards renewable energy.

The sector was supervised by the *Ministerio de Energía y Minería* (MINEM [Ministry of Energy and Mining]), which was in charge of defining policies for the energy sector and overseeing their application. *RenovAr* was divided into several auction rounds, which set up the process to attract international bidders and create a new market for private investment in renewable energy. By the end of 2016, Argentina had successfully conducted two auction rounds. Other rounds made this a successful project. However, accepting a bid in an auction did not guarantee its delivery by a given date nor the possibility of project changes.

Coelho mentioned that she was inspired by other projects that have used satellite imagery for public service. One example is a machine-learning project that reads satellite images to help environmental regulators identify potentially hazardous agricultural facilities and animal farms (de Witte, 2019). Another project that served as a model for *LN's* AI project was "Leprosy of the Land," produced by the award-winning Ukrainian organization *TEXTY*. For this project, the *TEXTY* team detected illegal amber mining in northern Ukraine, an area of 70,000 km², through satellite images from Microsoft's Bing (Merrill, 2019; Tymoshchuk et al., 2019).

Similarly, other newsrooms were also using satellite imagery to improve their news reporting. At *Reuters*, the team used these images to track the expansion of man-made islands in the South China Sea (Wu et al., 2018). *Associated Press (AP)* used satellite imagery to investigate sea vessels in Southeast Asia. This story revealed abuses in the seafood industry and was recognized for its public service with a Pulitzer Prize in 2006 (Bonfanti & Bordignon, 2017).

One of the greatest inspirations came from Stanford's Deep Solar project. The team that worked on this project used a deep learning model to detect solar panels on satellite imagery (see Yu et al., 2018). Thus, *LN Data* decided to use a machine-learning-based predictive model to estimate solar farm deployment in the country while simultaneously testing and applying new technologies to news production. Our findings will be described in the following sections.

New technologies in journalism: A constant learning process

Using new technologies in journalism raises novel questions for the community regarding how it can best report issues, turn data into insights, and tell compelling stories that impact the audience. In the *RenovAr* project, a sub-set of AI and ML was necessary for developing the project. CV helped detect the location of solar farms and identify the status of projects to check whether farms were implemented within the given timeframe proposed by the government.

However, during the process, *LN Data* faced some issues while building the experiment to prototype AI-powered solutions using satellite imagery. Among the key challenges, "the team did not have the expertise to implement computer vision algorithms" (Florencia Coelho, *LN Data*, January 2020). Their first step was to look for other projects that adopted similar approaches. A couple of studies addressed the use of satellite imagery to recognize solar panels, such as Google's Sunroof project, OpenClimateFix's solar PV mapping, Stanford's Deep Solar in the US, a Polish study from Poznań University of Economics and Business, and Sunroof India, created by a start-up. By analyzing these studies, the team understood that the task complexity increased due to the quality of satellite images in the developing world, including Argentina. A lack of a research unit in space such as NASA in the US, the European Space Agency (ESA), or INPE (National Institute for Space Research) in Brazil made it even more complicated to access high-quality imagery. This challenge imposed specific technical requirements that need to be dealt with in developing such approaches.

Satellite imagery: A precious treasure, yet restricted to a minimal number of actors

The first challenge, which was identifiable at the local and national levels, involved understanding the required images to feed the algorithm and the desired level of precision. A CV algorithm seeks to understand and automate tasks that the human visual system can do, which is possible only by providing images visible to the human eye. This can be measured using spatial resolution, which refers to the pixel size of a satellite image, the basic unit of programmable color on a computer image.

Most satellite imagery options are costly; acquiring these was unrealistic for *La Nación*. Sentinel-2, which comprises twin satellites flying in the same orbit but phased at 180°, is part of the European Union's Earth observation program coordinated and managed by the European Commission. Due to its free data available at a relatively high spatial resolution and high revisit time (5-day revisit time), Sentinel-2 had the highest potential for providing the imagery required to identify solar farms. These twin satellites are one of the few free-of-charge options with an open policy that allows their use for this purpose.

Although the pixel size was of moderate scope compared to other options available on the market (which were not free), solar farms make a large footprint and recognizable pattern since they require a significant number of solar photovoltaics. Geostationary satellite imagery also comes in handy for large areas, mainly countries with large territorial extensions, such as Argentina, Brazil, and the US. Other solutions, like Planet, offer much better multispectral images with higher resolutions, but access to them is costly for a newsroom in the Global South.

As mentioned previously, another important issue was update frequency. A project under construction was mapped while the data was collected; thus, imagery from a satellite designed to provide a high revisit frequency was necessary. Although Google offers high-quality imagery through its API, as illustrated in Figure 1, it does not have a fixed schedule of updates, and image quality varies by area. The interviewees argued that most populated areas and highly developed countries tend to have a high update frequency, while updates take place annually in other parts of the world.

Figure 1. Solar farms in Nonogasta, La Rioja, Argentina (-29.328923, -67.423753). Source: Images captured from Google Maps, Sentinel-2, and Planet (PLANET, 2020; SENTINEL-2, 2020).

Google Maps Sentinel-2 Planet

In this context, the team decided that Sentinel-2 was their best option because it is designed to provide a revisit frequency of five days at the Equator, similar to commercial services such as Planet but with better resolution. One must be careful with what they use to map satellite imagery.

The resolution, similarities, and uncertainties pose a risk to news reporting, as echoed in other studies (Marconi et al., 2017).

The prerogative is limited to tech companies: Massive computing power and infrastructure

The goal was to explore the development of RenovAr in the whole country, which corresponds to an area of 2,780,000 km^2. It represents a tremendous territorial area since Argentina is the eighth-largest country in the world. As such, the computing power required to analyze the images downloaded from Sentinel-2 required advanced hardware and cloud computing resources, which *LN Data* did not have. The main benefit of using cloud computing is receiving access to computing power and infrastructure that would be difficult and expensive to establish individually; however, cloud computing is still expensive and limited to big tech companies, much like AI (Chan-Olmsted, 2019).

To overcome these limitations, we suggested a partnership with a company that could offer these resources. After consecutive meetings with different organizations, the *LN Data* team partnered with a start-up specialized in satellite data applications that make various objects detectable through CV. The Argentine start-up, Dymaxion Labs, offered a cloud-based GeoML API for extracting Earth Observation insights and provided the infrastructure not found in the newsroom. It was a win-win situation for both: *La Nación* could use its resources to produce more knowledge of an area that the team did not have expertise in yet, and Dymaxion Labs could apply its software to a public service project.

Using computer vision to map Argentine solar farms

First, we must provide some context. ML and CV are two fields that have become closely related to one another. However, they are not the same, as discussed before. For the story about the solar farms, a supervised learning algorithm was used. A supervised learning algorithm takes a known set of input data, known as the "learning set," and forms a model to generate reasonable predictions and respond to the new input data and uses it to analyze the whole dataset's known responses to the data, which comprise the output (Aronson, 2018; Marconi, 2020; Marconi et al., 2017).

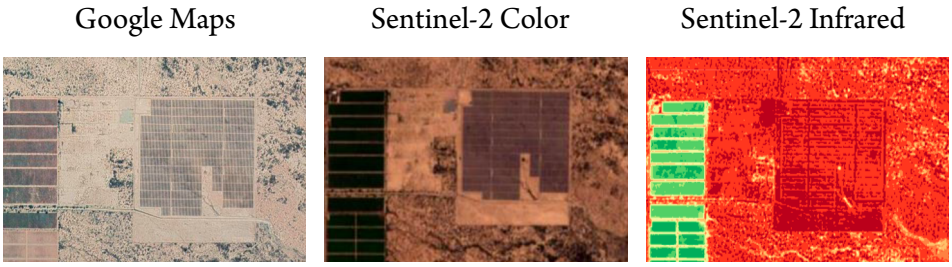
In this case, *MINEM* provided a list of places where solar farms would be built. Based on this data, a list of places that already had solar farms was created, but the number of locations with solar farms constructed by December 2019 was not high enough to generate a training dataset. To solve this, the teams discovered geolocations in OpenStreetMap of Chile's solar farms. By joining these two datasets, a final dataset was created. Both teams analyzed a sample of 100 geocoordinates used to train the dataset. These solar farms were checked manually and classified. Summarily, the steps of the algorithm were as follows:

1. Polygons were drawn around known solar farms based on the built dataset;
2. Geo-located outlines of known solar farms were cropped from the satellite imagery from June 2019 to mid-January 2020 and combined to obtain the best captures. The teams used image segmentation techniques to help in this process;
3. The images were cropped into chunks of 100x100 pixels with positive and negative examples of solar farms to train the algorithm, which used a convolutional neural network (CNN) to distinguish solar farms from any other negative patterns, such as agricultural and mountain areas, cities, and bare soil. In this way, the algorithm could distinguish patterns such as borders and color combinations with greater precision;
4. Finally, the teams scored their model on imagery across the entire country, in partitions of 100x100 pixels. A total of seven million images were processed.

One of the "biggest challenges in mapping solar farms is the similarity to farms" (Federico Bayle, CEO at Dymaxion Labs, February 2020). To solve this, the applied CNN, which comprised computing systems vaguely inspired by the biological neural networks in the human brain, considered the colors in the RGB (red, green, and blue) channels of the images; in this case, the infrared channel was considered since the distinction from crop fields was better. This channel has low reflectance values for solar panels and high reflectance values for agricultural plots, which helps make the distinction, as illustrated in Figure 2. Therefore, an image-recognition sys-

tem is only as accurate as the data used to train it, which is why high-resolution satellite images were needed.

Figure 2. Solar farms in Nonogasta, La Rioja, Argentina (-29.328923, -67.423753). Source: Images captured from Google Maps and both versions of Sentinel-2: RGB and Infrared (SENTINEL-2, 2020).



In total, 10,999 images were used to train the model; 70 % had a negative class, which means that most were not solar farms. Only 1,222 images were used to evaluate the results (images that were not employed by the algorithm to learn the patterns). With the solution deployed by Dymaxion Labs on Google Cloud Platform (a suite of cloud computing services), the training process took 30 hours and involved activating 30 graphic processing nodes (GPUs). In numbers, the process unfolded in the following manner:

- The algorithm precision was close to 94 %;
- Seven million images were processed;
- 2,780,400 square kilometers were analyzed;
- Thirty hours of model training took place.

Therefore, these results demonstrate technical knowledge that is particularly important in the AI industry. A lack of qualified personnel needed to program the complex systems and technical resources was a significant hurdle. This obstacle stemmed from the socio-educational problems observed in deploying AI solutions in any industry that requires significant capital investments for staff training and development. In the end, the project

identified 20 of 24 solar farms included in the list provided by the government. The algorithm also detected two solar farms that were private facilities.

Four facilities were too small to be identified by the AI solution, which found one false positive. Since the quality of the images from Sentinel-2 were not high resolution for the human eye, the team combined them with the images provided by a Planet Labs limited educational account for research, which offered higher quality satellite images. Besides these limitations, the algorithm was highly precise, reflecting the major challenges we faced while conducting this analysis. Due to the country's size, running such an algorithm has proven relatively expensive for any organization. As explained by the start-up's CEO,

> The major challenges for replicating this analysis revolve around two key points: the scale and the representativeness of the sample. Choosing an appropriate sample of data from the analyzed area guarantees representativeness. Also, higher resolution images would improve the results, allowing for the detection of smaller areas. (Federico Bayle, Dymaxion Labs, February 2020)

"The idea is to replicate [the] algorithm yearly to identify further solar farms and check if the projects that the government proposed meet the deadlines established" (Florencia Coelho, *LN Data*, March 2020). However, this agreement was a "single undertaking," meaning a single term. Next time, *LN Data* will need to establish a new partnership with this or another company or develop its team's skills since the team does not have "the skills or competencies to solve the problem in the newsrooms yet" (Momi Peralta, *LN Data*, March 2020).

Conclusion

In this study, we described the process of using CV, a sub-set of ML and AI, in news reporting. It is a largely unknown technology that has emerged in the scene, but journalists do not possess the required capabilities to use it in the newsroom (Biswal & Gouda, 2020).

We could notice four great challenges when implementing AI, especially concerning projects involving satellite imagery and CV in newsrooms. First, high-resolution imagery is highly expensive for any media

organization. Second, the unavailability of technological infrastructure in the newsroom to process such amounts of data created difficulties. Third, the absence of qualified personnel to develop such codes presented another obstacle. Finally, the lengthy and costly implementation process required significant investment. For this solar farm project, we found great potential for applying CV to journalistic work, including gathering information, providing signals to assess the veracity of content, and evolving investigative journalism practices (Diakopoulos, 2020).

That being said, news organizations have yet to unlock AI (Cook et al., 2021). To integrate AI into the daily news routine, organizations need to adapt to this transformation (Gade, 2004; Oliver, 2018; Porcu, 2020), much like how *La Nación* did by collaborating with other actors who are more familiar with the matter. To innovate in the news industry, newsrooms leaders must manage innovation processes to trigger and advance their teams' creative capabilities (Tidd & Bessant, 2009), as LN Data did by experimenting with CV to improve their ability to develop data-driven skills and take their reporting to the next level. This expands the theory of AI in news production by providing a new perspective on studying CV applicability and limitations, mainly in Latin America.

Within such a fluid environment, news media outlets are always seeking sustainable competitive advantages that might help them prosper in a time of uncertainty. News media organizations can use AI capacities to investigate stories that could not have been reported before due to a lack of information; for instance, news outlets could use satellite imagery to report on deforestation in areas considered of critical importance for conservation that are inaccessible to humans or affected by forest fires. Additionally, technology can assist reporters in gathering data as part of the future for watchdog reporting, which has evolved since the emergence of data journalism.

Of course, this kind of work has inherent limitations. While a key contribution of this study is the presentation of the process and hurdles that are part of systematically developing a news story using an AI system to detect objects in images, the focus of this project was also limited. This article has revealed some of the inputs and instructions *LN Data*'s solar

farm project relied on. The team's partnership with other organizations interested in *La Nación*'s work is critical to the development of this project. Without it, it would not have been realized. Innovation is not a temporary action but a long-term strategy that should be highly considered when deploying AI projects (Paulussen, 2016).

Newsrooms are struggling to cope with the switch to digital media and stay competitive in a technological environment that is constantly changing. Historically, innovation in news media is incremental, not disruptive, and is composed of a series of products that provide new features, benefits, or improvements to existing technology (Oliver, 2018; Paulussen, 2016). One example is data journalism, which took many years to establish a status quo in newsrooms around the globe (Lewis & Nashmi, 2019; Stalph & Borges-Rey, 2018). Recently, this practice has become popular in the Global South, despite the significant levels of disparity between countries (Jamil, 2019; Kashyap et al., 2020; Mutsvairo et al., 2019; Salaverría & de-Lima-Santos, 2021; Zhang & Feng, 2019). In a similar vein, Latin America faces many challenges related to deploying data-driven storytelling in newsrooms due to a lack of open data culture (Mazotte, 2017; Palomo et al., 2019; Salaverría, 2016). However, only a few Latin American news media organizations have managed to overcome this challenge with innovative solutions, such as *La Nación* (Argentina), *Ojo Público* (Peru), *Postdata.club* (Cuba), and *Rutas del Conflicto* (Colombia) (de-Lima-Santos & Mesquita, 2021a).

Our study has also highlighted how satellite imagery spatial resolutions may potentially impact the accuracy of image-use classification. High-resolution satellite imagery is highly expensive and restricted to well-resourced news outlets. Additionally, it requires computing power and infrastructures, which are in the hands of big tech corporations; as such, the adoption of these technologies by Global South newsrooms represents a challenging task. For this reason, top high-resolution satellite imagery providers, including Airbus, Maxar, and Planet, should offer advanced satellite imagery data for newsrooms in the Global South and create collaborative projects with them.

Collaboration should be utilized whenever possible for purposeful activities. Cook et al. (2021) pointed out that new alliances are being forged

to overcome these limitations in Latin America. Alliances with universities, civic organizations, and start-ups that have access to these technologies may be a possible solution to this issue. Currently, the symbiotic relationship between technology companies and media outlets is still limited to social media platforms (Rashidian et al. 2018); soon, this association could extend to accessing and delivering cutting-edge technology that is far too expensive for the majority of media companies (Linden, 2017a, 2017b). Future researchers should address the role that big technology companies play in limiting or enabling innovation in newsrooms.

As technology evolves, data journalism and AI will become intertwined and require skilled reporters in newsrooms. Therefore, newsrooms will need more information on the effective application of AI solutions to different areas of news media organizations, ranging from audience engagement to business decisions and innovation perspectives. However, a "lack of journalists' AI-related and digital journalism education and training" (Jamil, 2020, p. 16) is a problem present not only in Argentina but throughout the Global South and in many Western countries. The available scholarly literature offers very little insight into how training alone can solve the problem, even though there are countries where ICT development and the digital divide still represent significant issues. Future researchers could explore this phenomenon.

To conclude, it is crucial to understand how CV can be applied to other areas of the newsroom and whether it will become necessary to establish teams dedicated to developing AI and ML in news media organizations. To harness the utilities of AI, media companies must tackle various challenges. Thus, this study aimed to provide insight into the development of the competencies necessary for integrating CV into projects implemented in newsrooms. To our knowledge, no prior study has investigated the use of CV in news production.

References

Anderson, C. W. (2019). Between Certainty and Uncertainty: The Historical, Political, and Normative Contexts of Inferential Journalism

Claims. *Journalism and Communication Monographs, 21*(4), 358–361. https://doi.org/10.1177/1522637919878730

Appelgren, E., & Nygren, G. (2014). Data Journalism in Sweden. *Digital Journalism, 2*(3), 394–405. https://doi.org/10.1080/21670811.2014.884344

Appelgren, E., & Salaverría, R. (2018). The Promise of the Transparency Culture: A comparative study of access to public data in Spanish and Swedish newsrooms. *Journalism Practice, 12*(8), 986–996. https://doi.org/10.1080/17512786.2018.1511823

Aronson, J. 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

Becker, H. S., & Geer, B. (1957). *Field methods and techniques participant observation and interviewing: a comparison. Human organization* (Vol. 16).

Berliner, D. (2014). The political origins of transparency. *Journal of Politics, 76*(2), 479–491. https://doi.org/10.1017/S0022381613001412

Biswal, S. K., & Gouda, N. K. (2020). Artificial intelligence in journalism: A boon or bane? *Optimization in Machine Learning and Applications* (pp. 155-167). Springer. https://doi.org/10.1007/978-981-15-0994-0_10

Bonfanti, A., & Bordignon, M. (2017). 'Seafood from Slaves': The Pulitzer Prize in the Light of the UN Guiding Principles on Business and Human Rights. *Global Policy, 8*(4), 498–504. https://doi.org/10.1111/1758-5899.12495

Borges-Rey, E. (2016). Unravelling Data Journalism. *Journalism Practice, 10*(7), 833–843. https://doi.org/10.1080/17512786.2016.1159921

Borges-Rey, E. (2021). Journalism With Machines? From Computational Thinking to Distributed Cognition. In L. Bounegru & J. Gray (Eds.), *The Data Journalism Handbook* (2nd ed., pp. 147–156). Amsterdam University Press. https://doi.org/10.5117/9789462989511_ch22

Bryman, A. (2012). *Social Research Methods* (4th ed.). Oxford University Press.

Carlson, M. (2016). Automated journalism: A posthuman future for digital news? *The Routledge Companion to Digital Journalism Studies* (pp. 226-234). Routledge. https://doi.org/10.4324/9781315713793-23

Carson, A., & Farhall, K. (2018). Understanding Collaborative Investigative Journalism in a "Post-Truth" Age. *Journalism Studies, 19*(13), 1899–1911. https://doi.org/10.1080/1461670X.2018.1494515

Castro, D., & New, J. (2016). *The Promise of Artificial Intelligence*. Center for Data Innovation.

Chan-Olmsted, S. 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

Charles, M. (2020). Understanding Trauma for Reconciliation and Peace-building Journalism in Colombia. *Journalism Practice, 15*(2), 259–270. https://doi.org/10.1080/17512786.2020.1713857

Cheruiyot, D., Baack, S., & Ferrer-Conill, R. (2019). Data Journalism Beyond Legacy Media: The case of African and European Civic Technology Organizations. *Digital Journalism, 7*(9), 1–15. https://doi.org/10.1080/21670811.2019.1591166

Coddington, M. (2015). Clarifying Journalism's Quantitative Turn. *Digital Journalism, 3*(3), 331–348. https://doi.org/10.1080/21670811.2014.976400

Cook, C., Garcia, E., Gyulnazaryan, H., Melano, J., Parusinski, J., & Sabadan, A. (2021). *The next wave of disruption: Emerging market media use of artificial intelligence and machine learning* (R. Shaw (ed.); 1st ed.). International Media Support (IMS).

de-Lima-Santos, M. F., Schapals, A. K., & Bruns, A. (2020). Out-of-the-box versus in-house tools: how are they affecting data journalism in Australia? *Media International Australia*, 1-15. https://doi.org/10.1177/1329878X20961569

de-Lima-Santos, M.-F., & Mesquita, L. (2021a). The Strategic Value of Data Journalism. In R. Salaverría & M.-F. De-Lima-Santos (Eds.), *Journalism, Data and Technology in Latin America* (1st ed., pp. 97–136). Palgrave Macmillan Ltd. https://doi.org/10.1007/978-3-030-65860-1_4

de-Lima-Santos, M. -F., & Mesquita, L. (2021b). Data Journalism in favela: Made by, for, and about Forgotten and Marginalized Communities. *Journalism Practice*, 1–19. https://doi.org/10.1080/17512786.2021.1922301

De Maeyer, J., Libert, M., Domingo, D., Heinderyckx, F., & Le Cam, F. (2015). Waiting for Data Journalism. *Digital Journalism*, *3*(3), 432–446. https://doi.org/10.1080/21670811.2014.976415

de Witte, M. (2019). *Stanford scholars show how machine learning can help environmental monitoring and enforcement*. Stanford. https://news.stanford.edu/press-releases/2019/04/08/how-machine-learning-can-help-the-environment/

Diakopoulos, N. (2015). Algorithmic Accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398–415. https://doi.org/10.1080/21670811.2014.976411

Diakopoulos, N. (2020). Computational News Discovery: Towards Design Considerations for Editorial Orientation Algorithms in Journal-

ism. *Digital Journalism, 8*(7), 945–967. https://doi.org/10.1080/21670811.2020.1736946

Eisenman, M. (2013). Understanding Aesthetic Innovation in the Context of Technological Evolution. *Academy of Management Review, 38*(3), 332–351. https://doi.org/10.5465/amr.2011.0262

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). *Writing Ethnographic Fieldnotes* (2nd ed.). The University of Chicago Press.

Felle, T. (2016). Digital watchdogs? Data reporting and the news media's traditional 'fourth estate' function. *Journalism: Theory, Practice & Criticism, 17*(1), 85–96. https://doi.org/10.1177/1464884915593246

Fink, K., & Anderson, C. W. (2015). Data Journalism in the United States: Beyond the "usual suspects." *Journalism Studies, 16*(4), 467–481. https://doi.org/10.1080/1461670X.2014.939852

Franklin, B. (2014). The Future of Journalism: In an age of digital media and economic uncertainty. *Journalism Studies, 15*(5), 481–499. https://doi.org/10.1080/1461670X.2014.930254

Gade, P. J. (2004). Newspapers and Organizational Development: Management and Journalist Perceptions of Newsroom Cultural Change. *Journalism and Communication Monographs* (Vol. 6). https://doi.org/10.1177/152263790400600101

Gage, J. (2020, July 20). *What's GPT-3? - Technically*. https://technically.substack.com/p/whats-gpt-3

Graefe, A. (2016). *Guide to Automated Journalism*. Tow Center for Digital Journalism, Columbia University. https://doi.org/10.7916/D80G3XDJ

Hassaballah, M., & Awad, A. I. (Eds.) (2020). *Deep Learning in Computer Vision*. CRC Press/Taylor and Francis. https://doi.org/10.1201/9781351003827

IFC. (2018). *RenovAr (Argentina): Scaling 'Express Edition.'* https://www.ifc.org/wps/wcm/connect/industry_ext_content/ifc_external_corporate_site/infrastructure/resources/scaling+infra+-+renovar+-+argentina

Jamil, S. (2019). Increasing Accountability Using Data Journalism: Challenges for the Pakistani Journalists. *Journalism Practice,* 15(1), 1–22. https://doi.org/10.1080/17512786.2019.1697956

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

Johnson, J. C., Avenarius, C., & Weatherford, J. (2006). The Active Participant-Observer: Applying Social Role Analysis to Participant Observation. *Field Methods, 18*(2), 111–134. https://doi.org/10.1177/1525822X05285928

Karlsen, J., & Stavelin, E. (2014). Computational Journalism in Norwegian Newsrooms. *Journalism Practice, 8*(1), 34–48. https://doi.org/10.1080/17512786.2013.813190

Kashyap, G., Bhaskaran, H., & Mishra, H. (2020). "We need to find a revenue model": Data journalists' perceptions on the challenges of practicing data journalism in India. *Observatorio (OBS*) Journal, 14*(2), 121–136. http://obs.obercom.pt/index.php/obs/article/view/1552

Kawulich, B. B. (2005). Participant Observation as a Data Collection Method. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 6*(2). http://www.qualitative-research.net/index.php/fqs/article/view/466/996

Knight, M. (2015). Data journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice, 16*(1), 55–72. https://doi.org/10.1080/14682753.2015.1015801

Kueng, L. (2017). *Going Digital. A Roadmap for Organisational Transformation*. Reuters Institute.

Lambeth, E. B. (1998). Public Journalism as a Democratic Practice. In E. B. Lambeth, P. E. Meyer, & E. Thorson (Eds.), *Assessing Public Journalism* (pp. 15-35). University of Missouri Press.

Lewis, N. P., & Nashmi, E. Al (2019). Data Journalism in the Arab Region: Role Conflict Exposed. *Digital Journalism*, *7*(9), 1200–1214. https://doi.org/10.1080/21670811.2019.1617041

Lewis, S. C. (2015). Journalism in an Era of Big Data. *Digital Journalism*, *3*(3), 321–330. https://doi.org/10.1080/21670811.2014.976399

Lewis, S. C., & Usher, N. (2013). Open source and journalism: Toward new frameworks for imagining news innovation. *Media, Culture and Society*, *35*(5), 602–619. https://doi.org/10.1177/0163443713485494

Lewis, S. C., & Westlund, O. (2015). Big Data and Journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*, *3*(3), 447–466. https://doi.org/10.1080/21670811.2014.976418

Lewis-Beck, M., Bryman, A., & Futing Liao, T. (Eds.). (2004). *The SAGE Encyclopedia of Social Science Research Methods* (1st ed.). Sage Publications, Inc. https://doi.org/10.4135/9781412950589

Linden, C.-G. (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

Linden, C.-G. (2017b). Decades of Automation in the Newsroom. *Digital Journalism*, *5*(2), 123–140. https://doi.org/10.1080/21670811.2016.1160791

Lischka, J. A. (2019). Strategic renewal during technology change: Tracking the digital journey of legacy news companies. *Journal of Me-*

dia Business Studies, 16(3), 182–201. https://doi.org/10.1080/16522354.2019.1635349

LN Data. (2020, June 9). *The Nature Project from LA NACION*. https://blogs.lanacion.com.ar/projects/english/the-nature-project-at-la-nacion/

Loosen, W. (2021). Data-Driven Gold-Standards: What the Field Values as Award-Worthy Data Journalism and How Journalism Co-Evolves with the Datafication of Society. In L. Bounegru & J. Gray (Eds.), *The Data Journalism Handbook: Towards A Critical Data Practice* (pp. 360-369). Amsterdam University Press. https://doi.org/10.1515/9789048542079-050

Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2020). Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism*, 21(9), 1246–1263. https://doi.org/10.1177/1464884917735691

Marconi, F. (2020). *Newsmakers*. Columbia University Press.

Marconi, F., Siegman, A., & Machine Journalist. (2017). The Future of Augmented Journalism: A guide for newsrooms in the age of smart machines. *AP Insights*.https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf

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

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J. B. R., Chen, Y., Zhou, X., Gomis, M. I., Lonnoy, E., Maycock, T., Tignor, M., & Waterfield, T. (2018). *Global warming of 1.5°C*. (Vol. 1). IPPC. https://www.ipcc.ch/sr15/

Mazotte, N. (2017, April 20). How the Argentinian daily La Nación became a data journalism powerhouse in Latin America. *Nieman Journalism Lab*. https://www.niemanlab.org/2017/04/how-the-argentinian-daily-la-nacion-became-a-data-journalism-powerhouse-in-latin-america/

McCarthy, J. (1998). What is artificial intelligence? *CogPrints*. http://cogprints.org/412/2/whatisai.ps

McGregor, S. (2013). CAR hits the mainstream. https://archives.cjr.org/data_points/computer_assisted_reporting.php?page=all

Merrill, J. B. (2019, March 7). My eyes sting from searching these images for the same thing. *Quartz AI Studio*. https://qz.ai/my-eyes-sting-from-searching-these-images-for-the-same-thing/

Meyer, P. (2002). *Precision Journalism: A Reporter's Introduction to Social Science Methods.* Rowman & Littlefield Publishers.

Milan, S., & Treré, E. (2019). Big Data from the South(s): Beyond Data Universalism. *Television & New Media, 20*(4), 319–335. https://doi.org/10.1177/1527476419837739

Mutsvairo, B. (2019). Challenges Facing Development of Data Journalism in Non-Western Societies. *Digital Journalism, 7*(9), 1289–1294. https://doi.org/10.1080/21670811.2019.1691927

Mutsvairo, B., Bebawi, S., & Borges-Rey, E. (2019). *Data Journalism in the Global South*. Springer International Publishing. https://doi.org/10.1007/978-3-030-25177-2

O'Donoghue, T. (2003). *Qualitative Educational Research in Action*. Routledge. https://doi.org/10.4324/9780203506301

Oliver, J. J. (2018). Strategic transformations in the media. *Journal of Media Business Studies, 15*(4), 278–299. https://doi.org/10.1080/16522354.2018.1546088

Ortiz Freuler, J., & Iglesias, C. (2018). *Algoritmos e Inteligencia Artificial en Latinoamérica: un Estudio de implementaciones por parte de Gobiernos en Argentina y Uruguay*. https://webfoundation.org/

Palomo, B., Teruel, L., & Blanco-Castilla, E. (2019). Data Journalism Projects Based on User-Generated Content. How La Nacion Data Transforms Active Audience into Staff. *Digital Journalism, 7*(9), 1270–1288. https://doi.org/10.1080/21670811.2019.1626257

Paulussen, S. (2016). Innovation in the Newsroom. In T. Witschge, C. W. Anderson, D. Domingo, & A. Hermida (Eds.), *The SAGE Handbook of Digital Journalism* (1st ed., pp. 192–206). SAGE Publications Ltd. http://dx.doi.org/10.4135/9781473957909.n13

PLANET (2020). Planet Application Program Interface: In Space for Life on Earth. https://api.planet.com/

Picard, R. G. (2014). Twilight or new dawn of journalism? Evidence from the changing news ecosystem. *Journalism Studies, 15*(5), 500–510. https://doi.org/10.1080/1461670X.2014.895530

Porcu, O. (2020). Exploring innovative learning culture in the newsroom. *Journalism, 21*(10), 1556–1572. https://doi.org/10.1177/1464884917724596

Rashidian, N., Brown, P., Hansen, E., Bell, E., Albright, J., Hartstone, A., Hansen-With, E., Bell, E., Albright, J., & Hartstone, A. (2018). Friend and foe: The platform press at the heart of journalism. In *Columbia Journalism Review*. https://doi.org/10.7916/D8-15PQ-X415

Richardson, L. (1990). *Writing Strategies: Reaching Diverse Audiences*. SAGE Publications.

Salaverría, R. (2016). *Ciberperiodismo en Iberoamérica*. Editorial Ariel and Fundación Telefónica.

Salaverría, R., & de-Lima-Santos, M.-F. (2020). Towards Ubiquitous Journalism: Impacts of IoT on News. In J. Vázquez-Herrero, S. Direito-Rebollal, A. Silva-Rodríguez, & X. López-García (Eds.), *Journalistic Metamorphosis: Media Transformation in the Digital Age Volume 70 of Studies in Big Data* (1st ed., pp. 1–15). Springer International Publishing. https://doi.org/10.1007/978-3-030-36315-4_1

Salaverría, R., & de-Lima-Santos, M.-F. (Eds.). (2021). *Journalism, Data and Technology in Latin America*. Palgrave Macmillan. https://www.doi.org/10.1007/978-3-030-65860-1

SENTINEL-2 (2020), *ESA's Optical High-Resolution Mission for GMES Operational Services (ESA), SP-1322/2*. http://esamultimedia.esa.int/multimedia/publications/SP-1322_2

Spradley, J. P. (2016). *The Ethnographic Interview*. Waveland Press.

Stalph, F. (2018). Classifying Data Journalism: A content analysis of daily data-driven stories. *Journalism Practice*, *12*(10), 1332–1350. https://doi.org/10.1080/17512786.2017.1386583

Stalph, F., & Borges-Rey, E. (2018). Data Journalism Sustainability. *Digital Journalism*, *6*(8), 1078–1089. https://doi.org/10.1080/21670811.2018.1503060

Szeliski, R. (2011). *Computer Vision*. Springer London. https://doi.org/10.1007/978-1-84882-935-0

Teece, D. J. (2010). Business models, business strategy and innovation. *Long Range Planning*, *43*(2–3), 172–194. https://doi.org/10.1016/j.lrp.2009.07.003

Tidd, J., & Bessant, J. (2009). Innovation - What it is and why it matters. In *Managing Innovation: Integrating Technological, Market and Organizational Change*. Wiley.

Tymoshchuk, Y., Herasymenko, V., Kelm, N., Bondarenko, A., & Drozdova, Y. (2019). Leprosy of the land. https://texty.org.ua/d/2018/amber_eng/

UNFCCC. (2017, July 5). *Clean Energy Can Meet 90% of Paris Energy-Related Goals*. https://unfccc.int/news/clean-energy-can-meet-90-of-paris-energy-related-goals

Usher, N. (2016). *Interactive Journalism: Hackers, Data, and Code*. University of Illinois Press.

van Dalen, A. (2012). The algorithms behind the headlines. *Journalism Practice, 6*(5–6), 648–658. https://doi.org/10.1080/17512786.2012.667268

Verganti, R. (2006). Innovating Through Design. *Harvard Business Review*, 1–17.

Whittaker, J. (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

Wu, S., Tandoc, E. C., & Salmon, C. T. (2019). A Field Analysis of Journalism in the Automation Age: Understanding Journalistic Transformations and Struggles Through Structure and Agency. *Digital Journalism, 7*(4), 428–446. https://doi.org/10.1080/21670811.2019.1620112

Yu, J., Wang, Z., Majumdar, A., & Rajagopal, R. (2018). DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States. *Joule, 2*(12), 2605–2617. https://doi.org/10.1016/j.joule.2018.11.021

Zhang, S., & Feng, J. (2019). A Step Forward? *Journalism Studies, 20*(9), 1281–1300. https://doi.org/10.1080/1461670X.2018.1513814