Ca' Foscari
University
of Venice

Master Degree programme

in Digital and Public
Humanities

Final Thesis

AI and Sport Journalism : a perfect duo?

Supervisor
Ch. Prof. Gianluca Lebani

Assistant supervisor
Ch. Prof. Stefano Dall'Aglio

Graduand
Alessio Tesse
Matriculation Number 892774

Academic Year
2023 / 2024

Index

Abstract

The intersection of artificial intelligence (AI) and journalism has revolutionized media, particularly in sports reporting. This thesis explores the growing integration of AI, focusing on Natural Language Generation (NLG), and its impact on sports journalism. AI technologies, such as automated systems for real-time data analysis and content generation, offer sports journalists tools to enhance reporting efficiency, allowing faster coverage and personalized content. However, these advancements raise significant questions about human creativity, journalistic integrity, and the evolving role of journalists. This research critically examines current AI applications in sports journalism, from automated match reports to sentiment-driven sports narratives, and assesses the broader implications of AI's influence on the profession. Ethical considerations, particularly regarding accuracy, bias in algorithms, and the potential decline of in-depth investigative reporting, are also explored. By analyzing both the opportunities and challenges posed by AI, this thesis aims to provide a balanced understanding of how technology is reshaping sports journalism while proposing strategies to maintain journalistic values in an AI-driven future.

Introduction

The intersection of technology and journalism has drastically reshaped the media landscape, and few innovations have been as transformative as artificial intelligence (AI). In the realm of sports journalism, AI is increasingly being utilized to automate processes, analyze vast datasets, and even generate content, offering both promising advancements and significant challenges. From real-time data analysis and automated match reports to personalized content delivery, AI is rapidly changing how sports stories are produced, disseminated, and consumed by the public.

As AI technologies evolve, they bring numerous opportunities for sports journalism. Automated systems can instantly compile statistics, generate game summaries, and even predict future match outcomes, offering sports journalists new tools to enhance their reporting. These advancements allow for faster coverage, deeper analysis, and more engaging storytelling for audiences, particularly in the fast-paced, content-driven world of modern sports media.

However, the increasing reliance on AI also raises important questions about the role of human creativity, journalistic integrity, and the future of the profession itself. Will the automation of sports reporting diminish the value of human journalists, or will AI serve as a complementary tool that enhances their work? How can AI-generated content be trusted in terms of accuracy, nuance, and objectivity? Furthermore, the rise of AI in journalism raises ethical concerns about bias in algorithms, loss of jobs, and the potential decline of in-depth investigative reporting, which is central to the role of traditional journalism.

This thesis will argue that while AI offers remarkable possibilities for the enhancement of sports journalism, it also presents complex challenges that must be addressed to ensure a balance between technological innovation and the core values of journalism. By analyzing current AI applications in sports reporting, the evolving role of sports journalists, and the broader implications for the industry, this research will critically examine the ways AI is reshaping sports journalism and explore strategies for navigating these changes in a manner that benefits both professionals and audiences alike.

Chapter 1 begins by defining AI as the ability of machines to imitate human intelligence and outlines key milestones in AI’s development, from early computing models like Charles Babbage's Analytical Engine and Alan Turing's work, to the rise of neural networks and logic systems.

Early AI breakthroughs included expert systems like DENDRAL and MYCIN, which incorporated specialized knowledge, but the field also faced setbacks during "AI winters" due to unmet predictions. AI research took a scientific turn with probabilistic

reasoning, and advancements in big data and deep learning accelerated progress, particularly in complex tasks like image and speech recognition.

Deep learning, with its multi-layered neural networks, transformed natural language processing (NLP) and led to advancements in sentiment analysis, word embeddings, and chatbots. However, challenges in reasoning, explainability, and handling complex data structures remain.

NLG, which focuses on generating human-like text from data, has also evolved, moving from rule-based to deep learning approaches. This has improved applications like summarization, information retrieval, and report generation.

Overall, the document emphasizes AI's growth from theoretical foundations to practical applications, highlighting deep learning and NLP as key factors in modern AI systems.

Chapter 2 explores the increasing role of Natural Language Generation (NLG**)** in journalism, with a particular emphasis on sports reporting. NLG technologies, including systems like **Wordsmith**, **ArriaNLG**, and **OpenAI's GPT series**, are widely utilized in areas such as financial news and sports journalism due to their ability to transform structured data into coherent, audience-specific content. These systems allow for faster production of routine reports, enabling journalists to focus on more complex tasks, such as in-depth analysis. However, challenges associated with accuracy, fluency, and maintaining reader engagement remain critical considerations in their implementation.

Automated journalism, defined as the use of algorithms to generate news articles without human intervention, plays an increasingly prominent role in newsrooms. While automation allows for high-volume and timely coverage—especially in cases where human reporters may not be present—it must still adhere to essential journalistic standards, such as accuracy, transparency, and modifiability. Ethical concerns, particularly regarding misinformation, remain a major challenge for automated systems.

In sports journalism, NLG tools such as **PASS** and **GoalGetter** demonstrate how automation can generate real-time sports reports. These systems often tailor their output to the emotional tone of audiences, enhancing user engagement. Despite these advancements, automated content is often criticized for its formulaic style and inability to capture the emotional nuance inherent in sports reporting, a domain traditionally reliant on human interpretation and insight.

The chapter also delves into the historical evolution of sports journalism, tracing its origins from the 19th century to its contemporary form. The relationship between sports and the media has been longstanding, with newspapers first recognizing the

commercial appeal of sports coverage during the Industrial Revolution. As sports journalism developed, particularly during the "Golden Age" in the early 20th century, it became a dominant feature in newspapers, contributing significantly to their financial success.

In the digital era, sports journalism has expanded through online platforms, which offer immediacy and interactive engagement with audiences. Digital sports reporting retains many of the traditional journalistic principles, including objectivity, but the shift toward market-driven content and the influence of algorithms, particularly on platforms like Bleacher Report, has sparked debates about journalistic integrity and editorial independence.

Ultimately, while automated journalism offers numerous benefits in terms of efficiency, scalability, and reach, the role of human journalists remains indispensable, particularly in areas requiring emotional depth, investigative reporting, and critical analysis. The integration of NLG technologies into sports journalism, though promising, necessitates a careful balance between automation and the need for personalized, contextually rich, and engaging content.

Chapter 3 examines the evolution of journalism from digital formats to the use of Artificial Intelligence (AI). It begins by addressing the ongoing challenge of defining "digital journalism," with terms like "cyber journalism" and "online journalism" used interchangeably. Key features such as hypertextuality (non-linear linking of texts), interactivity (user engagement with content and journalists), and multimedia (combining text, audio, and video) are highlighted as essential elements.

Digital journalism's history is traced from the early 1990s with the first online newspapers, through the rise of Web 2.0 in the 2000s, which introduced user-generated content and social media platforms like Facebook and Twitter, challenging traditional journalism models.

The chapter then explores AI's impact on journalism since the 2010s, particularly in automating tasks like data analysis and content generation. Case studies from countries such as Spain and Brazil showcase AI tools in sports journalism and audience engagement. AI is seen as enhancing journalism by automating routine tasks, enabling journalists to focus on deeper, investigative work, and reshaping media production and consumption.

Chapter 4 analyzes how Artificial intelligence (AI) has proven effective in fields like science, technology, and journalism, though development costs and specialized expertise remain challenges, especially due to competition with major tech firms. AI's integration into journalism has significantly expanded, particularly in automating news content through systems like those from Narrative Science and Automated Insights, which generate articles for outlets such as the Associated Press. This trend,

starting in the U.S. around 2010, has spread globally, with European companies using AI for stock and sports news updates.

The rise of AI in journalism has raised concerns about job automation, echoing fears from the Industrial Revolution. Experts argue that while AI can assist journalists by improving efficiency and fact-checking, it also poses risks related to information control by large tech firms and the blurring of human vs. machine-generated content. Additionally, AI-driven news customization may create “echo chambers” that polarize public opinion.

Despite AI's ability to automate tasks and support fact-checking, ethical concerns about accuracy, content prioritization, and the role of human creativity persist. A balanced approach that maintains essential journalistic qualities such as creativity and critical thinking is crucial for the future of journalism.

In conclusion of this thesis will be explained how AI and journalism can meet without any fear and doubt.

Chapter 1

AI and NLG

1 What is Artificial Intelligence?

Defining Artificial Intelligence is challenging. Artificial Intelligence (AI) is a branch of computer science that examines the theoretical underpinnings, approaches, and techniques that enable the design of hardware systems and software program systems that can perform tasks on a computer that, to the untrained eye, seem to be limited to human intelligence. The definition of a crucial scientific field is widely acknowledged due to its strategic importance. Its objective is to go beyond the existing limits of information technology in particular areas, with the scientific aim of improving and broadening computer capabilities. It's important to note that artificial intelligence (AI) is an exploratory field, wherein an AI system is considered adequate only when its intended performance can be quantified (Somalvico 1987).

According to Somalvico(1987) the objective of AI is not to directly replicate human intelligence, but rather to imitate specific aspects of it, such as problem-solving using inferential methods. This imitation can be achieved through mechanisms different from those used by humans, enabling equal or even superior performance. The scope of AI is not limited to a predefined set of subjects, but rather it continuously expands to address cutting-edge challenges. Abilities that are currently seen as exclusively human may, in the future, be achievable by artificial systems, reshaping our understanding of AI's capabilities. Additionally, AI encompasses both scientific inquiry and practical application: it aims to accurately model and assess intelligent behavior while also utilizing this knowledge to develop advanced artificial systems. Furthermore, AI is a field that thrives on innovation and adaptability, constantly evolving as new discoveries are made. When considering artificial intelligence, it's important to note that it doesn't depend on a fixed set of principles. Instead, it constantly tackles new challenges, pushing its limits and abilities. As a result, a capability that currently seems unique to human intelligence and is considered part of AI's definition might not be seen that way in the future. This change could lead people to believe that such performance can be achieved by an artificial system, potentially removing it from the AI category. The point to consider is that AI encompasses both a scientific and an engineering aspect (Nilsson 1998). It is considered a science because, by emulating specific intelligent behaviors in artificial

systems, humans can attain a thorough and objective modeling and experimental process that leads to promising outcomes, thereby allowing for undeniable progress in studying human intelligence using scientific methods. Additionally, it is viewed as engineering when machines achieve behaviors that were previously considered exclusive to humans, thereby contributing to the enhancement of human life. At the end of this chapter, I'm going to focus on what explains. Starting from the origin of AI to the development of AI application, furthermore I will argue about NLG(Natural Language Generaction), which tool is handy for journalistic reason.

1.2 The history of AI

1.2.1 The origin

The origin of AI started in the 19th century. Still, a lengthy series of steps have led to the development of the computer and what is now known as artificial intelligence over more than two centuries. If there is a beginning, it is with Joseph Marie Jacquard's invention of the "Jacquard loom," which was patented in 1804 and used a series of punched cards with instructions for fabric designs. Jacques Vaucanson in 1740, Jean Baptiste Falcon in 1728, and Basile Bouchon in 1725 were the inventors of earlier inventions on which it was based.

In the mid-19th century, Charles Babbage (1792–1871) conceptualized two computing devices, both of which remained unfinished. The Difference Engine was designed to generate mathematical tables for use in engineering and scientific endeavors. In contrast, Babbage's Analytical Engine represented a more groundbreaking initiative; it featured addressable memory, utilized stored programs derived from Jacquard's punched card system, and incorporated conditional branching. This machine is recognized as the first capable of universal computation.Later, Babbage built the Analytical Engine, a more potent and adaptable variation of the Difference Engine. The analytical machine's(Figure 1) design is similar to that of contemporary electronic computers. Punch cards, an idea from the Jacquard loom used to weave intricate textile patterns, could be used to program it. Formally speaking, Babbage's analytical machine, which could operate on data and carry out arbitrary instructions, complies with the later definition of a computer.

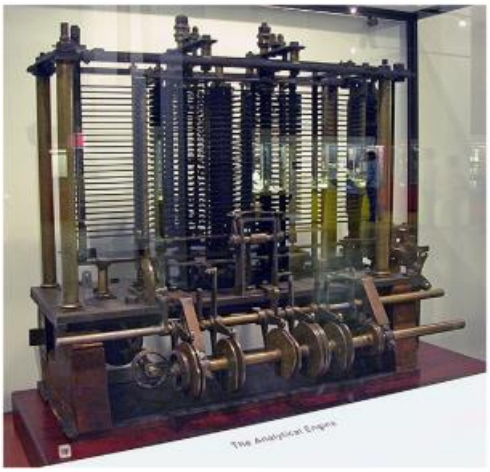

Figure 1 The Babbage machine was created by Charles Babbage was the first concept for a general-purpose, fully automated mechanical digital computer, capable of performing any calculation. Babbage was the first to conceive of and attempt to build such a device. (Britannica)

Moreover, the influence of Babbage's was strong on Ada Lovelace. In 1833 she participated in a gathering arranged by Charles Babbage to showcase his Difference Engine. Her admiration for Babbage's creation led her to engage in a 27-year correspondence with him. Moreover, she was Louis Menabrea's translator and added several annotations to enhance it. Within these annotations, was an explanation of utilizing the Analytical Engire for computer Bernoulli numbers. The initial computer program ever developed was this description, marking the inaugural execution of a algorithm for a mechanical computer in history. Ada devised it even in the absence of the actual implemented machine, as the Analytical Engine was merely a hypothetical machine that Babbage never constructed physically.

The specific field known as the "science of operations" was only named and recognized several years after its existence. Consequently, Ada Lovelace discussed Computer Science a century before it was formally acknowledged. Nonetheless, Ada Lovelace is widely acknowledged as the first computer programmer in history due to her contributions to the field.

1.2.2 The Turing Machine

Suppose Ada Lovelace was the precursor of computer science. In that case, Alan Turing is considered the father of Theoretical Computer Science and AI , because of its pioneering contributions, which gave him access to his theoretical machine.. But before this he worked for the “Turing machine” and in 1936, Turing created a machine to tackle an important question regarding Hilbert's decision problem, which inquires about the possibility of developing an algorithm to determine the universal validity of a first-order logic formula. Alonzo Church also studied this issue at the same time, but he approached it from a perspective entirely different from Turing's, although it was essentially equivalent in practice. Turing's proposed machine was purely theoretical, as he did not physically construct it.

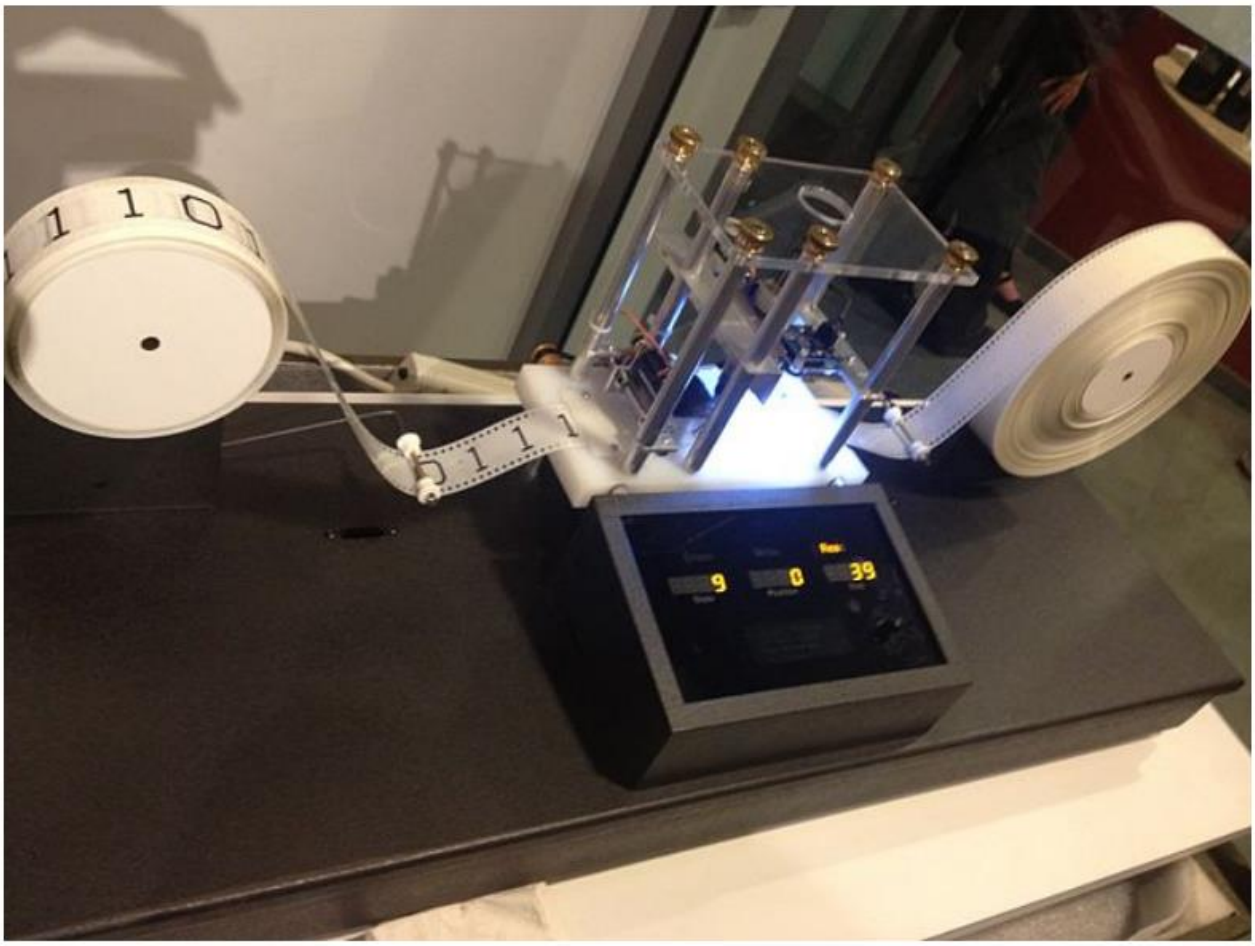

Figure 2 A physical Turing machine model. A true TuriFigura 3ng machine would have unlimited tape on both sides; however, physical models can only have a finite amount of tape. (Britannica)

Figure 2.

Because it demonstrated how well machines could mimic human intelligence, the Turing Test had a significant influence on research into artificial intelligence. But it has also brought up significant philosophical—and, more importantly, ethical—questions concerning the nature of intelligence and the possibility of machine consciousness. The ethical dilemmas posed by the Turing Test have implications beyond artificial intelligence, raising questions about the responsibilities of creating machines that could potentially possess self-awareness. These debates touch upon the fundamental question of what it means to be "intelligent" or "conscious."

On a more technical level, foundational models in computing, such as the Turing machine, provide a framework for understanding how machines process information. A Turing machine is a type of state machine that operates in one of a finite number of states at any given moment. The machine's behavior is governed by a set of instructions that dictate state transitions based on specific conditions. It utilizes an infinite, one-dimensional tape, which is conceptually laid out in a left-to-right orientation. This tape is divided into cells, each capable of holding a symbol, typically ‘0’ or ‘1’. The machine interacts with the tape via a read-write head, which scans a single cell at a time and can move left or right along the tape to access other cells.

The operation of a Turing machine is fully determined by three factors: its current state, the symbol in the cell currently being scanned, and a set of transition rules, which function as the machine’s program. Each transition rule is represented as a 4-tuple: <State0, Symbol, Statenext, Action>, meaning that if the machine is in state *State0* and the current cell contains *Symbol*, it will transition to state *Statenext* and perform *Action*—either writing a symbol to the tape or moving the read-write head left or right.

The machine halts if there is no applicable transition rule or if multiple transition rules apply simultaneously. Conceptually, the tape functions as the machine's memory, while the read-write head acts as the interface for accessing and modifying this memory. Importantly, the tape is assumed to be infinite, implying an unlimited memory capacity. Additionally, a function is considered Turing-computable if there exists a set of instructions that enables the machine to compute it, regardless of the time required, effectively assuming infinite time for computation.(Barker Plummer 2004)

1.2.3 The Birth of AI

During the years of the war, the first recognized work in artificial intelligence (AI) was by Warren McCulloch and Walter Pitts in 1943(McCulloch & Pitts 1943). They were inspired by Pitts's advisor Nicolas Rashevsky and drew on knowledge of brain physiology, propositional logic, and Turing's theory of computation. They showed a model of artificial neurons, which could be "on" or "off," depending on the stimulation from neighboring neurons. They demonstrated that any computable function could be replicated by a network of these neurons and that logical operations could be implemented through simple neural networks. They also suggested that these networks could learn. Later, Donald Hebb proposed a rule for modifying the connections between neurons, known as Hebbian learning, which remains influential in AI research today.

As interest in these early models of computation and neural networks grew, the foundation was laid for further collaborative research efforts to explore the potential of artificial intelligence more deeply. In 1955, John McCarthy of Dartmouth College convinced Minsky, Claude Shannon, and Nathaniel Rochester to help him bring together U.S. researchers interested in automata theory, neural nets, and the study of intelligence. They organized a two-month workshop at Dartmouth in the summer of 1956. There were 10 attendees in all, including Allen Newell and Herbert Simon from Carnegie Tech.

Although the workshop generated significant excitement and fostered collaboration among these pioneering thinkers**,** the Dartmouth workshop did not lead to any breakthroughs. The most developed work was introduced by Newell and Simon, a mathematical theorem-proving system known as the Logic Theorist (LT). Simon asserted they had created a computer program capable of non-numerical thinking, thus resolving the longstanding mind-body problem. This program, unlike LT, was initially created to mimic human problem-solving protocols. It was discovered that within the restricted category of puzzles it could manage, the program's approach to considering subgoals and potential actions was comparable to how humans tackled the same problems. The "thinking humanly" approach was likely first exemplified in the GPS program. The success of GPS and similar programs in mirroring cognition prompted Newell and Simon (1976) to articulate the well-known physical symbol system hypothesis, which asserts that "a physical symbol system possesses the necessary and sufficient tools for general intelligent action." Essentially, they argued that any intelligent system, whether human or machine, must function by manipulating data structures comprised of symbols. We will later explore how this hypothesis has been questioned from various angles.

Meanwhile,in 1958 John McCarthy significantly advanced the field of artificial intelligence through two key contributions. In the inaugural MIT AI Lab Memo No. 1, he introduced the high-level programming language Lisp, which subsequently

emerged as the primary language for AI development for the following three decades. In his seminal paper titled "Programs with Common Sense," McCarthy proposed a conceptual framework for AI systems grounded in knowledge and reasoning. This work outlines the concept of the Advice Taker, a theoretical program designed to possess comprehensive world knowledge and utilize that information to formulate actionable plans.

At the same time, another pivotal event unfolded in the AI community the year 1958 was significant as it marked Marvin Minsky's relocation to the Massachusetts Institute of Technology (MIT). However, his early partnership with John McCarthy was short-lived. McCarthy focused on representation and reasoning within the framework of formal logic, while Minsky prioritized the practical functionality of programs, ultimately adopting a perspective that was critical of traditional logic. Despite their diverging paths, both made substantial contributions to the field of AI in 1963, McCarthy established the artificial intelligence laboratory at Stanford University. His ambition to create the ultimate Advice Taker through logical frameworks was significantly propelled by J. A. Robinson's introduction of the resolution method in 1965. The research conducted at Stanford focused on the development of general-purpose methodologies for logical reasoning. Notable applications of these logical frameworks encompassed Cordell Green's systems for question-answering and planning (Green, 1969), as well as the Shakey robotics initiative at the Stanford Research Institute. Meanwhile, at MIT, Minsky's approach was distinctively different, emphasizing the application of AI to specific, well-defined problems**.** Under his guidance, a group of students chose particular challenges that appeared to require intelligent solutions, laying the groundwork for what would later be known as microworlds. For instance, James Slagle's SAINT program, developed in 1963, successfully addressed closed-form calculus integration problems commonly encountered in introductory college courses. Similarly, Tom Evans's ANALOGY program, created in 1968, tackled geometric analogy challenges typically found in intelligence quotient assessments(Evans 1968). But the most important microworld is blocks world, which is widely recognized as the most prominent microworld, characterized by a collection of solid blocks situated on a tabletop, or more frequently, a simulated tabletop environment. A common objective within this microworld involves rearranging the blocks in a specified configuration, utilizing a robotic hand capable of lifting one block at a time. This domain has been the foundation for several significant projects, including David Huffman's vision project (1971), David Waltz's work on vision and constraint propagation (1975), Patrick Winston's learning theory (1970), Terry Winograd's natural language understanding program (1972), and Scott Fahlman's planning system (1974). At least the foundational research that expanded upon the neural networks conceptualized by McCulloch and Pitts experienced significant growth. Notably, the contributions of

Shmuel Winograd and Jack Cowan (1963) demonstrated that a multitude of elements could collaboratively embody a singular concept, thereby enhancing both robustness and parallel processing capabilities. Furthermore, Bernie Widrow advanced Hebb's learning principles through his development of adalines (Widrow and Hoff, 1960; Widrow, 1962), while Frank Rosenblatt introduced perceptrons in 1962. The perceptron convergence theorem, articulated by Block et al. (1962), asserts that the learning algorithm is capable of adjusting the connection weights of a perceptron to align with any given input data, contingent upon the existence of such a match.

1.2.4 The Hard Times

From the outset, researchers in artificial intelligence exhibited a strong sense of optimism regarding the capabilities of their field, frequently making audacious forecasts. In 1957, Herbert Simon notably asserted that machines with the ability to think, learn, and create would soon match human intelligence. He expected that within a decade, a computer would emerge as a chess champion and resolve significant mathematical theorems; however, these achievements were not realized until approximately 40 years later. Simon's excessive confidence was largely influenced by the early successes of AI systems in performing straightforward tasks, yet these systems encountered considerable difficulties when faced with more intricate problems for two primary reasons: The foundational approach of early AI was more predicated on assumptions regarding human problem-solving than on comprehensive analyses of tasks and their specific requirements. There was a significant lack of insight into the computational complexity associated with numerous problems. Initial methodologies depended on exploring various combinations to identify solutions, which proved effective in limited, uncomplicated contexts but faltered in more complex situations. Researchers initially misjudged the difficulties involved in scaling these approaches to address larger, more challenging problems. In parallel to these issues with symbolic AI, another major challenge emerged in the domain of neural networks**.** The advancement of intelligent behavior within neural networks encountered significant obstacles due to inherent limitations in their foundational architectures. The seminal work "Perceptrons" by Minsky and Papert (1969) illustrated that while perceptrons to could learn any function they could represent, their representational capabilities were severely restricted, particularly in their inability to distinguish between two separate inputs. Although these limitations did not extend to more sophisticated multilayer networks, there was a notable decline in funding for neural network research. Paradoxically, the back-propagation learning

algorithms, which would later rejuvenate interest in neural networks during the 1980s and 2010s, had already been conceived in the early 1960s.

1.2.5 The Rebirth

After a period of darkness, AI started to gain momentum again with a series of innovative programs. The DENDRAL program, conceived at Stanford University by Ed Feigenbaum, Bruce Buchanan, and Joshua Lederberg in 1969, stands as one of the pioneering instances of artificial intelligence applied to the challenge of deducing molecular structures from mass spectrometry data (Buchanan 1969). The program takes as input the molecular formula alongside the mass spectrum, which details the masses of various molecular fragments generated through electron bombardment. Initially, DENDRAL produced all conceivable molecular structures that aligned with the provided formula and simulated the corresponding mass spectra to compare with the actual data. However, this approach quickly became computationally impractical for molecules of even moderate complexity. To enhance computational efficiency, the DENDRAL team integrated insights from analytical chemists, who recognized patterns within the mass spectrum indicative of prevalent molecular substructures. For instance, a specific heuristic for identifying a ketone (C=O) group involves detecting particular peak patterns in the mass spectrum, which significantly reduces the range of potential molecular structures. This incorporation of chemical heuristics into the program facilitated a more efficient and manageable process for identifying molecular structures. This strategy of leveraging domain-specific knowledge marked a significant shift from purely computational methods to a more collaborative and interdisciplinary approach in AI research. The researchers involved in the DENDRAL project worked in conjunction with analytical chemists and found that these chemists determine molecular structures by identifying recognizable patterns of peaks within a spectrum, which signify common substructures. For instance, they apply specific criteria to detect the presence of a ketone (C=O) subgroup.And the authors assert that DENDRAL's strength lay in its incorporation of mass spectroscopy knowledge, not through fundamental principles, but rather through effective "cookbook recipes" (Feigenbaum et al., 1971). DENDRAL's importance is underscored by its status as the inaugural successful knowledge-intensive system, with its expertise stemming from a multitude of specialized rules. In 1971, Feigenbaum and his colleagues at Stanford initiated the Heuristic Programming Project (HPP) to explore the potential applications of the emerging expert systems method in various domains.

The following of the DENDRAL project in utilizing domain-specific knowledge laid the groundwork for the next major advancement in expert systems, which aimed to replicate and even enhance human expertise in fields beyond chemistry**.** The subsequent significant initiative was the development of the MYCIN system,

designed to diagnose blood infections. Comprising approximately 450 rules, MYCIN showed diagnostic capabilities comparable to those of certain experts and significantly surpassed the performance of less experienced physicians. Notably, MYCIN exhibited two principal distinctions from the DENDRAL system. First, in contrast to the rules of DENDRAL, there was no overarching theoretical framework from which the MYCIN rules could be derived; instead, these rules were planned through comprehensive interviews with medical experts. Second, the rules were required to encapsulate the inherent uncertainty present in medical knowledge. To address this, MYCIN integrated a framework known as certainty factors, which, at the time, appeared to align effectively with how physicians evaluated the influence of evidence on their diagnostic processes. In 1981, the Japanese government unveiled the "Fifth Generation" project, a decade-long endeavor aimed at creating highly parallel, intelligent computing systems utilizing Prolog. The financial commitment for this initiative was projected to surpass $1.3 billion in contemporary terms. In reaction, the United States established the Microelectronics and Computer Technology Corporation (MCC), a consortium intended to maintain national competitiveness. In both instances, AI was integrated into a wider strategy that encompassed chip design and human-computer interaction research. Meanwhile, in the United Kingdom, the Alvey report reinstated funding that had been withdrawn following the Lighthill report. However, none of these initiatives ultimately achieved their ambitious objectives regarding advancements in AI capabilities or their economic repercussions. Despite various challenges, the artificial intelligence sector witnessed significant growth and fervor during the 1980s. This era saw a remarkable increase in market value, escalating from several million dollars in 1980 to billions by 1988. This surge led to the establishment of many enterprises focused on the development of expert systems, vision systems, robotics, and tailored software and hardware solutions. Nevertheless, this swift growth was accompanied by substantial obstacles. Subsequently, the industry faced a downturn referred to as the "AI winter," characterized by the failure of many firms that had made ambitious promises but failed to deliver. The complexities involved in constructing and sustaining expert systems for intricate domains became increasingly evident, particularly as these systems encountered uncertainty-related challenges and their inability to learn from past experiences. The AI winter refers to periods of reduced funding and enthusiasm for AI research, which occurred in the late 1980s and early 1990s following the crash of the specialized AI hardware market. The increased affordability and power of desktop computers from companies like Apple and IBM rendered expensive AI machines obsolete. In addition, early expert systems, which were expensive and hard to maintain, failed because they couldn't learn and adapt.

1.2.5 Probabilistic reasoning

The brittleness of early expert systems led to a shift in artificial intelligence (AI) towards a more scientific approach, prioritizing probabilistic reasoning over Boolean logic, using machine learning techniques instead of hand-coded rules, and relying on experimental data over philosophical arguments(Cohen 1995). Researchers started building upon established theories and concentrating on rigorous theorems and experimental methodologies, anchoring their work in practical, real-world applications rather than theoretical or artificial examples. The creation of standardized benchmark datasets, such as those from the UC Irvine repository, ImageNet, and the International Planning Competition, facilitated the assessment of progress across different AI subfields, including machine learning, speech recognition, and natural language processing. During this period, AI's evolution also involved a reintegration with other scientific disciplines, countering previous tendencies towards isolationism. This reintegration recognized the relevance of integrating insights from fields like information theory, stochastic modeling, classical optimization, and formal methods into AI research. For instance, the prevalence of hidden Markov models (HMMs) in the 1980s in speech recognition showed the significance of leveraging a strong mathematical foundation and large training datasets, leading to more dependable and widely applicable technologies. While not representing human cognition, HMMs offered an effective framework for the computational analysis of speech, advancing the field and enabling broader industrial and consumer use.

1.2.6 Big data era

As AI research continued to evolve, another key factor that emerged was the dramatic increase in the availability and accessibility of data. The increase in computational power and the rise of the Internet have made it possible to create and store large amounts of data, often referred to as "big data." These datasets contain a wide variety of information, including trillions of words, billions of images, hours of speech and video, as well as extensive genomic, vehicle tracking, clickstream, and social network data. The availability of such massive amounts of data has led to the development of machine learning algorithms specifically designed to make the most of these vast resources. A significant portion of these datasets is not labeled. For example, in Yarowsky's (1995) influential study on word-sense disambiguation, instances of the word "plant" are not marked to distinguish between meanings such as 'flora' or 'factory.' However, with sufficiently large datasets, effective algorithms can achieve high accuracy, surpassing 96% in distinguishing intended word senses. Furthermore, Banko and Brill (2001) showed that increasing dataset sizes by several orders of

magnitude can lead to more significant performance improvements than fine-tuning the algorithms themselves. A similar pattern can be observed in computer vision applications as images in painting. Hays and Efros (2007) presented an original approach to filling in missing parts of an image by incorporating pixels from comparable images. They discovered that this method was not effective with a limited database containing thousands of images, but it showed significant improvement with millions of images. The subsequent release of tens of millions of images in the ImageNet database (Deng et al., 2009) signaled a transformative phase in computer vision. The emergence of big data, coupled with a shift toward machine learning, played a crucial role in reigniting the commercial potential of artificial intelligence (AI) (Havenstein, 2005; Halevy et al., 2009).

1.2.7 Deep learning era

Because of these advancements, the AI landscape shifted dramatically. Researchers could now leverage vast datasets and improved computational power, which paved the way for a new paradigm in AI development: deep learning. Deep learning represents a subset of machine learning characterized by the use of multiple layers of interconnected processing units, commonly referred to as neural networks.Deep learning represents a subset of machine learning characterized by the use of multiple layers of interconnected processing units, commonly referred to as neural networks. The roots of deep learning can be traced back to the 1970s, with initial experiments and the advent of convolutional neural networks (CNNs), which gained significant traction in the 1990s, particularly in applications like handwritten digit recognition.(LeCun et al 1995) However, it was not until 2011 that deep learning truly began to flourish, first showcasing its capabilities in speech recognition and later in visual object recognition. A transformative event for deep learning transpired during the 2012 ImageNet competition, where a model created by Geoffrey Hinton's team at the University of Toronto dramatically outperformed prior systems that depended on manually crafted features(Krizhevsky et al., 2013). Since that time, deep learning frameworks have achieved performance metrics that exceed human proficiency in specific visual tasks, although they continue to encounter challenges in other areas. Similar advancements have been noted in domains such as speech recognition, machine translation, medical diagnostics, and gaming. A notable example is ALPHAGO, which triumphed over elite human Go players, largely due to its implementation of deep networks for complex evaluation functions. These extraordinary accomplishments have rekindled interest in artificial intelligence (AI) across a multitude of sectors, including academia, industry, and government, as well as among the public and media. There are frequent reports of novel AI applications

that either match or surpass human capabilities, generating both enthusiasm for swift progress and apprehension regarding the potential for another AI winter. The efficacy of deep learning is significantly reliant on sophisticated computational hardware. While a conventional CPU can perform around 10^9 to 10^10 operations per second, deep learning algorithms executed on specialized hardware such as GPUs, TPUs, or FPGAs can achieve between 10^14 and 10^17 operations per second, primarily through highly parallelized matrix and vector computations.

1.3 What is deep learning?

Deep learning (DL), a subset of machine learning (ML), represents knowledge as a hierarchy of concepts with increasing levels of abstraction, mimicking the human brain's multi-layered processing(Goodfellow,Bengio 2016). Unlike traditional ML models, which typically employ a simple two-layer structure, DL models use multiple layers to progressively transform input data(Murphy 2012). Each hidden layer in a deep neural network (ANN) combines information from the previous layer to learn more complex functions, ultimately predicting the output through successive transformations. This approach eliminates the need for explicit human intervention in defining all the necessary knowledge, allowing the system to build complex concepts from simpler ones. Visualizing these layers as a stacked graph reveals a "deep" structure, reflecting the numerous layers involved. DL is inspired by fields such as linear algebra and probabilistic statistics, and its advances have facilitated the development of foundation models. These models, trained on broad datasets using self-supervision, can be adapted to diverse tasks, as seen in large language models (LLMs) used in natural language processing. Despite recent progress in DL, artificial neural networks remain significantly smaller than those found in the human brain or even in simpler animals like frogs. Leading scholars, including Goodfellow, Bengio, and Courville, predict that ANNs will not match the scale of the human brain until at least the 2050s, barring breakthroughs in scaling technology. Significantly, recent progress in contemporary deep learning (DL) has been realized through artificial neural networks (ANNs) that possess a markedly lower number of neurons than those found in the human brain. Although these ANNs are considered computationally extensive by today's technological benchmarks, they are still less complex than the neural networks observed in relatively uncomplicated organisms, such as frogs. Prominent figures in the discipline, including Goodfellow, Bengio, and Courville, project that ANNs are unlikely to reach the neuron count of the human brain until the 2050s, unless technological innovations enable a more accelerated scaling process. Despite the notable success of deep learning (DL) models, they still face significant challenges in terms of reasoning and explainability. Many current DL methods,

particularly those based on artificial neural networks (ANNs), lack interpretability. This issue extends to foundation models and large language models (LLMs). The complexity arises from the hierarchical and nonlinear structure of ANNs and the core DL concept of connectionism(Alpaydin 2016). In this framework, individual artificial neurons perform simple tasks independently and contribute only partially to the final output(Goldberg 2017). However, when a vast number of these neurons operate in parallel within hidden layers, the overall output becomes difficult to interpret, making DL models prone to producing unexplainable results. Moreover it appears that it is not feasible to ascertain which specific artificial neuron influenced a particular segment of the output, nor is it possible to comprehend the processes occurring within the intermediate (hidden) layers of the artificial neural network (ANN). In essence, extracting any fundamental rules that may be inherent in the deep learning (DL) model is unattainable. This limitation persists even in deep learning algorithms that employ supervised learning techniques, as these algorithms are unable to learn effectively without being provided with accurate sample patterns. Even when an artificial neural network (ANN) has been effectively trained to meet its objectives, the numerous numerical values of the weights generated by the model lack interpretability for the supervisor. While the model is defined by these weights, the methodology behind their calculation and the degree to which different input variables influence the results remain ambiguous. ANNs can be dynamically updated as new data is introduced into the network. This process leads to modifications in the weights, which are adjusted based on experiential learning. Such updates pose additional challenges concerning the interpretability of deep learning (DL) methodologies. DL is particularly adept at processing intricate sensor data, such as that obtained from cameras and microphones, which has historically been challenging for traditional computational techniques. This capability is especially relevant for cognitive tasks, including natural language processing, speech recognition, and facial recognition, which will be elaborated upon later. Current investigations in DL are exploring the possibility of decoding speech directly from brain activity. These methods involve monitoring cortical activity to extract features of the speech produced. Advanced deep neural network models are believed to enhance the overall accuracy of speech reconstruction from neural recordings in the human auditory cortex. The immediate aim of these research initiatives is to assist individuals who are unable to communicate due to injuries or neurodegenerative conditions by developing a synthesized version of their voice that can be manipulated through the activity of their brain's speech centers.

1.4 The NLP

Natural Language Processing (NLP) encompasses a theoretically grounded array of computational methods designed to analyze and represent texts that occur naturally. These techniques operate at various levels of linguistic analysis, aiming to facilitate human-like language processing across a diverse set of tasks and applications (Liddy 2001). Given its complexity and wide-ranging applications, NLP draws upon multiple areas of study to enhance its effectiveness and expand its capabilities**.** Natural Language Processing (NLP) is informed by a variety of intellectual disciplines, ranging from formal linguistics to statistical physics. This section offers a concise overview of the relationship between NLP and these adjacent fields, highlighting the interdisciplinary nature of the domain. Given the multidisciplinary nature of NLP, its applications span a wide range of areas within AI and beyond. From its theoretical foundations to its practical implementations, NLP has evolved into a crucial tool for tackling complex language-related tasks. Natural Language Processing (NLP), a branch of AI, focuses on developing algorithms capable of understanding and generating human language (Joseph et al., 2016). It plays a significant role in tasks such as sentiment analysis, where expressed opinions in various forms of public discourse, including social media, are classified (Ponce, 2014). Additionally, NLP facilitates automated customer interaction through chatbots, which are capable of conducting numerous personalized conversations simultaneously by recognizing language patterns and issuing appropriate responses (Long & Connelly, 2017). Although the core steps in NLP, such as exploratory analysis, preprocessing, feature extraction, model training, and validation, mirror those used for numerical data, text preprocessing and the transformation of language into numerical vectors present unique challenges (Collobert et al., 2011). Algorithmically deconstructing human language for text data processing (Collobert et al., 2011) can evoke a sense of apprehension and inadequacy among business leaders (Collobert et al., 2011; Gandomi & Haider, 2015).

In light of these challenges, NLP research has evolved to focus on several key areas, each of which addresses different aspects of human-computer interaction and language understanding.

Nlp mainly has three areas of research:

- Communication with Humans: Natural language is more convenient for human-computer interaction than formal languages, like first-order predicate calculus.
- Learning from Existing Knowledge: Much of human knowledge is recorded in natural language, such as in Wikipedia, so AI systems need to understand natural language to access this information.

- Scientific Advancement: Understanding natural language helps advance our knowledge of languages and language use, integrating insights from AI, linguistics, cognitive psychology, and neuroscience.

1.4.1. Deep learning for NLP

The fundamental issues are ingrained in every computational linguistic system. In order to carry out tasks such as translation, text summarization, and image captioning, there needs to be a certain level of comprehension of the underlying language. This comprehension can be categorized into at least four key areas: language modeling, morphology, parsing, and semantics. Language modeling can be looked at in two ways, as it determines word sequences and, by extension, the meaning of words, which are only weakly meaningful on their own. Morphology involves studying how words are constructed, including their roots, prefixes, suffixes, compounds, and other devices used within words to convey tense, gender, plurality, and other linguistic structures. Parsing examines how words modify each other to form constituents and create sentence structure(Eisenstein 2018). Semantics focuses on the meaning of words, considering individual word meanings, their relationships with other words, the contexts in which they appear, and some level of world knowledge or "common sense." Now we see the most important use of Deep learning in NLP:

Language modeling

One of the most important tasks in natural language processing (NLP) is language modeling. It plays a key role in nearly all NLP applications by creating models that can predict words or basic linguistic elements based on previous text. This function is especially useful for applications involving user input, as it improves text input speed by providing predictive text. The power and adaptability of language modeling stem from its capability to implicitly grasp syntactic and semantic relationships among words or linguistic elements within a sequential context, making it essential for tasks like machine translation and text summarization, where producing coherent and contextually suitable sentences is crucial.

Evolution of Language models

Although neural networks have achieved significant progress in language modeling, quantifying these improvements presents a challenge. It is important to be able to assess language models independently of their specific applications. Various metrics have been suggested, but a perfect solution has not yet been identified .The most commonly utilized measure is perplexity, which represents the inverse probability of a test set normalized by the word count. Perplexity is a suitable benchmark for language models trained on the same datasets, but its significance diminishes when comparing models trained on different vocabularies. Fortunately, several standard datasets exist in the field, enabling meaningful comparisons.

Word embeddings

In the field of natural language processing (NLP), it is important to have a way to represent words that doesn't rely on manual feature engineering but still captures the connections between words, whether they are related syntactically, semantically, topically, sentimentally, or in other ways. Simple one-hot encoding, which assigns a unique binary vector to each word in a dictionary, doesn't capture the nuanced relationships between words. Another approach, inspired by John R. Firth’s idea that "You shall know a word by the company it keeps," involves representing words based on the counts of n-grams of phrases they appear in. However, this method isn't feasible due to the vast dimensionality and sparsity of n-gram vectors. A more efficient method involves utilizing word embeddings, which are dense, low-dimensional vectors that encode words in a continuous vector space. These embeddings can be learned automatically from data, capturing semantic similarities and other word relationships in a compact form. Using word embeddings instead of one-hot vectors is advantageous in nearly all deep learning applications for NLP. Pretrained embeddings, like those provided by WORD2VEC, GloVe, and FASTTEXT, are commonly used as they reduce the time and effort needed for training. However, embeddings can also be fine-tuned or trained from scratch for specific tasks, allowing them to emphasize word relationships most relevant to the task's requirements.

Morphology

The study of morphology involves the identification of the elements present in individual words, including roots, stems, prefixes, suffixes, and in some languages, infixes. These affixes are utilized to alter stems in order to express grammatical categories such as gender, number, and person. Luong et al. (2013) developed a language model with morphological awareness using a Recursive Neural Network (RvNN) to capture morphological structures, upon which a neural language model was constructed. They trained two models on the WordSim353 dataset (Finkelstein et al 2001) using Morfessor for segmentation (Creutz & Lagus,2007): one context-sensitive and one context-insensitive. The context-insensitive model often misclassified words with similar stems but different meanings, such as antonyms, while the context-sensitive model performed better by recognizing the relationships between stems and other morphological features like the prefix “un.” This model surpassed previous embedding models across various datasets (Miller and Charles,1991) Morphological analyzers play a critical role in many natural language processing (NLP) tasks. Belinkov et al. (2017) investigated the capability of various neural machine translation (NMT) models to learn and utilize morphology. They developed several translation models from English to languages like French, German, Czech, Arabic, and Hebrew, using LSTM-based models and character-aware CNNs. The models were trained on the WIT3 corpus (Cettolo, Girardi and Federico,2012) and the decoders were replaced with POS and morphological taggers. The study revealed that attention mechanisms reduced encoder performance but improved decoder performance. Additionally, character-aware models exhibited better morphology learning abilities, while the complexity of the output language's morphology negatively impacted encoder performance. Morita et al. (2015) introduced a novel morphological language model tailored for unsegmented languages such as Japanese. Their RNN-based model, equipped with a beam search decoder, was trained on both automatically and manually labeled corpora, and was evaluated on the Kyoto Text Corpus (Kawanara et al,2002) and the Kyoto University Web Document Leads Corpus (Hangyo et al,2012), outperforming all baseline models in tasks such as morphological analysis, POS tagging, and lemmatization. A current focus in morphology research is universal morphology, which explores the connections between the morphologies of different languages, to create a universal morphological analyzer. Deep learning has had limited application in this field so far, except for one study that used it as a supplemental task for universal parsing (Dehouck & Denis,2018). Datasets such as those from a CoNLL shared task (More et al,2018) are available for those interested in this area. Furthermore, the development of morphological embeddings could facilitate multilingual processing, particularly among cognate languages, and could be advantageous for specialized domains such as biomedical literature. With the increasing importance of deep learning in NLP,

advancements in handling morphological components could significantly enhance model performance.

Parsing

Analyzing the structure of a sentence by looking at how words and phrases are connected is known as parsing. There are two main types of parsing: constituency parsing, which focuses on extracting hierarchical phrasal units from a sentence, and dependency parsing, which studies the direct relationships between pairs of individual words. Recent developments in parsing have mainly concentrated on dependency parsing, which can be further divided into graph-based and transition-based approaches. Graph-based parsing involves creating multiple parse trees using formal grammar to find the best tree, often using generative models. On the other hand, transition-based parsing has gained popularity due to its simplicity. It constructs a single parse tree by manipulating a buffer of words and a stack with the ROOT label, forming dependencies between words through a series of actions until the buffer is empty. Transition-based parsing methods are governed by three main approaches: the arc-standard approach, which connects all dependents to a word before connecting the word to its parent; the arc-eager approach, which links words to their parents as soon as possible; and the swap-lazy approach, which adapts the arc-standard method to allow for the swapping of positions on the stack, accommodating non-projective edges.

Semantic

The process of semantic processing involves understanding the meanings of individual words, phrases, sentences, or entire documents. Word2Vec and GloVe are techniques used to capture word meanings based on the distributional hypothesis of meaning. These techniques create vector representations that can be interpreted as having semantic meaning when processed through neural networks, especially for phrases or sentences. Research in neural semantic processing can be divided into two main areas: comparing the similarity in meaning between different sections of text

and capturing and transferring meaning within larger text units, particularly sentences.

1.4.2 The application of NLP using deep learning

Deep learning has several function for NLP like: word embeddings, sentiment analysis, information extraction, topic modeling, text generation ,and information retrivial; but here we focus on some function

-Information Retrievial: The goal of Information Retrieval (IR) is to assist users in locating the most pertinent information in a user-friendly format whenever necessary(Kenter et al,2017). Deep learning models in IR primarily function by comparing query texts with document texts to compute scores for relevance. These models can be classified into two primary approaches: representation-focused, which establish strong representations of texts for direct matching, and interaction-focused, which specifically model the connections between words in the query and document to understand matching patterns.

-Information Extraction: has several functions like

NER:

Named Entity Recognition (NER) involves the identification of proper nouns and specific details such as dates and prices. Initial models, like feedforward networks and LSTMs, encountered challenges due to computational limitations and the absence of advanced word vector models. Later models, including CharWNN and bidirectional LSTMs combined with CRFs, significantly enhanced performance by utilizing character-level and word-level embeddings. Recent progress using pretrained models, such as bidirectional character language models, has achieved state-of-the-art results across multiple languages without the need for hand-engineered features or lexicons.

Event Extraction:

Event Extraction focuses on identifying and categorizing events within texts, including the recognition of event triggers and the roles of different participants. Early methods relied on CNNs with max-pooling but were limited in capturing multiple events within a single sentence. Subsequent approaches integrated RNN-based encoder-decoder frameworks and latent variable neural models, improving performance in identifying event triggers and their roles.

Relationship Extraction

Aims to identify various types of relationships, such as possessive, antonymous, or familial, between entities within the text. Initial deep learning approaches utilized simple CNNs and outperformed previous methods. Further advancements included the use of bidirectional LSTMs combined with CNNs and attention-based GRU models, introducing new mechanisms for comprehensive information extraction. The utilization of pretrained models like BERT has recently established new benchmarks in specific domains, such as clinical temporal relation extraction. As we shift from information extraction to other key NLP applications, we can see the importance of these foundational tasks in supporting more complex analyses, such as text classification is an essential use of natural language processing (NLP), which entails assigning unstructured text documents to predetermined categories. This method has broad applicability across diverse fields. Kim (2014) was the first to utilize pre-trained word vectors in a convolutional neural network (CNN) for classifying sentences, demonstrating that a basic CNN structure, comprising a single convolutional layer followed by a dense layer with dropout and softmax output, could achieve superior performance on various benchmarks with minimal hyperparameter adjustment. This method enhanced the leading results in four out of seven sentence classification tasks, such as sentiment analysis and question classification. Further research by Conneau et al. (2017) demonstrated the effectiveness of deep networks with multiple convolutional layers for document classification. Progress in text classification has also been achieved with hybrid designs. Jiang et al. combined a deep belief network with softmax regression. The deep belief network, composed of layers modeled after restricted Boltzmann machines, was trained using unsupervised learning to optimize data dimensionality. The model underwent multiple iterations of forward and backward propagation to minimize an energy-based loss, independent of the labeled classification task. Once the network's segments were pre-trained, they were amalgamated and fine-tuned using backpropagation and quasi-Newton techniques. Additionally, Adhikari et al. utilized BERT to attain superior results on

several document classification datasets, emphasizing the potential of deep learning methods in NLP(Devlin et al,2018). However, deep learning does not universally outperform all text classification tasks. For example, Worsham and Kalita found that gradient boosting trees surpassed neural networks, including CNNs and LSTMs, in the classification of lengthy texts such as books by genre. This indicates that although deep learning shows promise for numerous NLP applications, there are still notable challenges, and alternative methods may be more appropriate for specific tasks.

Text generation

Another significant area that demonstrates the versatility and power of NLP techniques is text generation.Text generation is necessary for many NLP tasks. One way to do this is by using summarization and machine translation, which convert text from one form to another in a sequence-to-sequence (seq2seq) manner. Additionally, tasks like image and video captioning, as well as automatic weather and sports reporting, involve transforming non-textual data into text. Some tasks generate text without any input data, or with only minimal input used as a topic or guide. Text generation has also :

Data-to-text generation: In the realm of data-to-text generation, the input can vary significantly, encompassing both structured data, like a weather forecast description, and unstructured perceptual information, such as raw images or videos. The resulting output can range from a concise statement, such as an image caption, to a comprehensive multi-paragraph discourse. Despite the wide array of input types, all data-to-text systems encounter several common challenges (Reiter and Dale, 2000), including:

- identifying which elements of the data to articulate;
- strategizing the presentation of this information;
- converting the data into appropriate words and phrases;
- arranging these words and phrases into coherent sentences and paragraphs.

Moreover the initial data-to-text generation systems had a modular structure, using separate software components for different tasks in the generation process. Artificial intelligence planning algorithms were important in organizing the information and structuring individual sentences to effectively achieve communication goals (McKeown, 1992; Moore and Paris, 1993). The surface realization process could

involve the use of grammars or templates that link specific data types with corresponding words and phrases. For more complex linguistic needs, such as morphological inflections, these adjustments might be made as a post-processing step. An example is represented in figure 3, an example of data-to-text generation . An important challenge in surface realization is creating diverse referring expressions to avoid repetition. The NITROGEN system is an example of a rule-based and statistical approach that transforms abstract meaning representations into different sentence plans. To deal with the numerous possible realizations for a given meaning, the sentence plans are consolidated into a finite-state acceptor, and a bigram language model is then used to assign probabilities to these arcs, optimizing for the most likely surface realization. In contrast, newer data-to-text systems use unified models trained end-to-end through backpropagation and share similarities with machine translation in addressing the alignment problem.

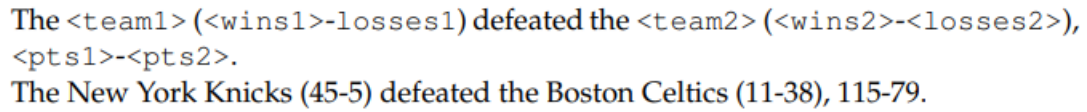

Figure 3

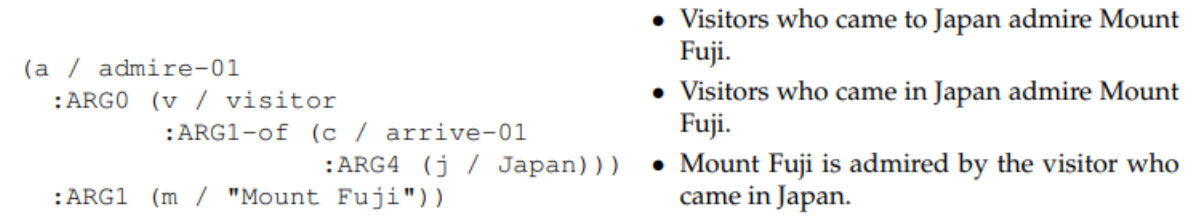

Figure 4

Text-to-text generation

Text-to-text generation encompasses tasks such as summarization and simplification, which can be exemplified by the following scenarios:

• analyzing a novel and producing a concise summary that captures the essence of the plot;

• reviewing a collection of political blog posts and generating a bullet-point list that outlines the various issues and viewpoints presented;

• examining a technical research paper on the long-term health effects of kombucha consumption and creating a summary that is accessible to a lay audience.

These tasks can be approached in two distinct manners: utilizing the encoder-decoder architecture previously discussed, or by directly manipulating the input text.

1.4.3 NLG

Text-to-text and data-to-text generation represent distinct categories within the domain of Natural Language Generation (NLG). A seminal survey conducted by Reiter and Dale (1997, 2000) articulates NLG as "the subfield of artificial intelligence and computational linguistics concerned with the construction of computer systems that can produce understandable texts in English or other human languages from some underlying non-linguistic representation of information" (Reiter & Dale, 1997). This characterization is more pertinent to data-to-text generation than to text-to-text generation. In their work, Reiter and Dale concentrate on data-to-text generation, providing an in-depth examination of the rule-based techniques that dominated the field during that period.

Defining Natural Language Generation (NLG) presents a challenge because of the diverse nature of its input, which can encompass semantic representations, numerical data, structured knowledge bases, and visual inputs like images or videos. The distinction between different NLG approaches is often unclear, with text summarization, for example, being considered a text-to-text task but employing techniques similar to data-to-text generation when generating new sentences from extracted opinions. Conversely, data-to-text systems may use text-to-text techniques to creatively express data. Applications such as generating spoken dialogue blur these boundaries further, often overlapping with dialogue management. This survey underscores that what sets data-to-text generation apart is its reliance on non-linguistic inputs, which represents the primary challenge faced by these systems. Unless specified otherwise, the term "Natural Language Generation" or "NLG" in this survey will refer to systems that produce text from non-linguistic data.

However, it is essential to note that NLG has not remained static since Reiter and Dale's initial survey. Over the past 15 years, the field of natural language generation (NLG) has undergone a significant transformation, marked by the development of applications that produce customized reports for specific audiences, as well as the advent of text-to-text and vision-to-text generation systems, which often employ statistical methods rather than traditional data-to-text approaches. These advancements have not been addressed by Reiter and Dale (2000). Furthermore, their work notably lacks discussion

on applications that extend beyond conventional, 'factual' text generation, such as those incorporating personality, affect, or creative elements like metaphors and narratives. Additionally, Reiter and Dale (2000) do not cover evaluation methodologies, an area that has only recently garnered systematic attention due to various shared tasks conducted within the NLG community.

Given these gaps and the evolving nature of the field, the main emphasis of this survey, particularly in its initial section, will focus on the generation of text from data, building on the research of Reiter and Dale (2000). While this survey will not cover all recent advancements in various applications of generating text from text, other surveys have explored these areas. For instance, Mani (2001) and Nenkova and McKeown (2011) have covered summarization, Androutsopoulos and Malakasiotis (2010) have examined paraphrasing, and Piwek and Boyer (2012) have addressed automatic question generation. Additionally, the survey will investigate the connections between generating text from data and generating text from text because the boundaries between them are often unclear. It will also explore how text-to-text systems have utilized data-driven frameworks that are now becoming popular in generating text from data, resulting in the development of hybrid systems that combine rule-based and statistical techniques.

Given the evolving landscape of NLG and the introducting of hybrid systems that blend rule-based and statistical techniques, it is also valuable to revisit the foundational components of NLG tasks. These tasks, which have long underpinned the generation of text from data, provide a structured framework for understanding how information is transformed into coherent text.

NLG Tasks: Traditionally, the NLG problem of converting input data into output text was addressed by splitting it up into several subproblems. The following six are frequently found in many NLG systems (Reiter & Dale, 1997, 2000):

Content Determination

Deciding which information to include in the text under construction, moreover it involves deciding which information from data should be included in the text and which should be excluded. Often, data contains more detail than necessary for text, so information must be filtered and abstracted into preverbal messages. This process varies by domain; for example, a soccer report might exclude minor events, while a neonatal care report would summarize sensor data. Recent research explores data-driven techniques for content determination, using methods like Hidden Markov

Models and clustering to model topic shifts and align data with text. This research also examines learning alignments between data and text in multiple domains, such as soccer and weather, often using weakly supervised techniques;

Text Structuring

The process of determining the order in which messages are presented to a reader is called text structuring. In domains like soccer reporting, it's common to start with general information (such as the game's location, date, and attendance) before describing specific events like goals. Researchers have explored using machine learning to automate document structuring and content selection, aiming to find an optimal order for presenting information. These methods are also applicable to multi document summarization, where sentences from various documents are organized based on their relevance to the summary.

Sentence aggregation

Rewording each sentence without changing the meaning, and not adding new sentences can cause in making the generated text more fluid and readable. Aggregation is challenging to define overall and has been understood in different ways, encompassing redundancy elimination and a combination of linguistic structure. Reape and Mellish (1999) provide an initial overview, differentiating aggregation at the semantic level. An example came from football

Sadio Mane scored three times for Southampton in less than three minutes.

Figure 3

Here we can see an example of aggregation

Lexicalisation

The number of lexical options and contextual restrictions influences the complexity of the lexicalisation process in natural language generation (NLG) systems. To introduce variety, systems might randomly select from different options, but they must ensure that the selections are stylistically appropriate. Lexical choices can convey attitudes or emotional positions. While some domains easily map domain concepts to words, lexicalisation often involves choosing between similar words or addressing ambiguous ideas, necessitating consideration of comparisons. Difficulties also arise when expressing gradable properties and presenting numerical information. Psycholinguistics offers insights into these processes, such as how speakers select words and make mistakes. NLG research, unlike cognitive modeling, views lexicalization as part of surface realization, combining conceptual representations with grammar rules.

REG

The generation of Referring Expressions (REG), as described by Reiter and Dale (1997), comprises the selection of appropriate terms or phrases to uniquely identify entities within a domain. This process shares similarities with lexicalization but differs by focusing on distinguishing a specific entity from others within the domain, a task that involves discrimination (Reiter & Dale, 2000). REG has attracted considerable attention in the field of automated text generation due to its applicability across diverse domains, allowing for independent solutions (Mellish et al., 2006; Siddharthan, Nenkova, & McKeown, 2011). An important aspect of REG is the determination of referential form, such as the choice between pronouns, proper names, or descriptions, based on the salience or focus of the entity (Poesio et al., 2004). This decision-making process has been the focus of shared tasks like the Generation of Referring Expressions in Context (GREC), which involves the classification of the most suitable referential form (Belz et al., 2010). While research has often concentrated on pronoun usage, there is an increasing exploration of the generation of proper names (Siddharthan et al., 2011; van Deemter, 2016; Castro Ferreira et al., 2017). It is important to determine the specific content being referred to, especially when using descriptive language in areas with multiple entities of the same type. In such situations, it's crucial to choose distinguishing characteristics that will ensure accurate identification by the reader or listener. Algorithms used for content determination in REG typically involve searching for the best combination of properties that set the referent apart in its context. Maintaining a balance of information is crucial, as too much can be redundant, while too little can make identification difficult. Many REG strategies are based on the

concept of Gricean maxim, which emphasizes providing sufficient information without going overboard (Grice, 1975). Different algorithmic approaches are used, such as exhaustive searches for the most concise descriptions (Full Brevity), incremental selection based on eliminating distractions (Greedy Heuristic), and preferences specific to the domain (Incremental Algorithm) (Dale, 1989, 1992; Dale & Reiter, 1995). Furthermore, research on dialogue references indicates that expressions may contain redundant information to achieve communication goals, like confirming a previous statement (Jordan & Walker, 2005) or tailoring descriptions based on user preferences (White et al., 2010).

Linguistic Realisation

Arranging the elements of a sentence correctly and generating the appropriate morphological forms, such as verb conjugations and agreement, where relevant, are part of the description. It also involves the addition of function words (like auxiliary verbs and prepositions) and punctuation. One main challenge is ensuring that the output includes various linguistic components that may not be present in the input, turning this into a process of generating between non-matching structures.

Numerous suggestions have been put forward like:

a) Templates: In scenarios where application domains are limited in size and there is anticipated to be little variation, bringing the realization to fruition is a relatively straightforward undertaking, and desired results can be defined by utilizing templates (e.g., Reiter, Mellish, & Levine, 1995; McRoy, Channarukul, & Ali, 2003).

```
$player scored for $team in the $minute minute.
```

This example has three variables in the code can be populated with the player's name, team, and the minute they scored a goal.

```
Ivan Rakitic scored for Barcelona in the 4th minute.
```

Templates offer the advantage of providing complete control over the output's quality and preventing the generation of ungrammatical structures. Contemporary versions of template-based approaches incorporate syntactic information within the templates, along with potentially intricate rules for populating the placeholders (Theune et al., 2001), thereby complicating the distinction between templates and more advanced methods (van Deemter, Krahmer, & Theune, 2005).

b) **Grammar Based-System**: General-purpose, domain-independent realization systems offer an alternative to templates. Most of these systems are grammar-based, meaning they base their choices on the grammar of the language being considered. The grammar can be manually written, as seen in many classic off-the-shelf realizers like fuf/surge (Elhadad& Robin, 1996), mumble (Meteer, McDonald, Anderson, Forster, Gay, Iluettner, & Sibun, 1987), kpml (Bateman, 1997), nigel (Mann & Matthiessen, 1983), and RealPro (Lavoie& Rambow, 1997). However, hand-coded grammar-based realizers often require very detailed input. For instance, kpml (Bateman, 1997) is based on Systemic-Functional Grammar (sfg; Halliday & Matthiessen, 2004), and realization is modeled as a traversal of a network where choices depend on both grammatical and semantico-pragmatic information. This level of detail makes these systems challenging to use as simple 'plug-and-play' or 'off the shelf' modules (e.g., Kasper, 1989), which has led to the development of simpler realization engines that provide syntax and morphology apis while leaving choice-making up to the developer (Gatt et al., 2009; Vaudry & Lapalme, 2013; Bollmann, 2011; de Oliveira & Sripada, 2014; Mazzei, Battaglino, & Bosco, 2016). One challenge for grammar-based systems is how to make choices among related options, which makes hand-crafted rules with the right sensitivity to context and input difficult to design.

Ivan Rakitic scored for Barcelona in the 4th minute.

For Barcelona, Ivan Rakitic scored in minute four.

Barcelona player Ivan Rakitic scored after four minutes.

c) **Statistical approach**: Recent approaches to natural language generation (NLG) seek to create probabilistic grammars from large datasets, reducing manual labor while expanding linguistic coverage. There are two main methods for integrating statistical data into the generation process. The first method, developed by Langkilde and Knight in their work on halogen/nitrogen systems, involves a two-tier strategy: a small manually created grammar generates different outputs as a forest, from which a stochastic re-ranker, informed by n-grams or more advanced statistical models, selects the best candidate. The second method avoids the computationally intensive generate-and-filter approach by directly including statistical data into generation decisions. An example of this approach is Belz's pcru system, which uses a context-free grammar to produce the most probable sentence derivation based on corpus data. In both methods, the base generator is manually constructed, and statistical information is mainly used to improve outputs. An alternative strategy involves using statistical data for the base-generation system itself. Grammar-based methods have been developed by extracting grammatical

rules from treebanks, as shown by the OpenCCG framework, which uses Combinatory Categorial Grammar (CCG) and statistical language models to achieve broad coverage in English surface realization. Similar principles underlie other generation approaches using different grammatical formalisms, including Head-Driven Phrase Structure Grammar (HPSG), Lexical-Functional Grammar (LFG), and Tree Adjoining Grammar (TAG). These systems typically use a variant of the chart generation algorithm to gradually put together input specifications into final structures, which are then ranked. The development of stochastic realisers with extensive grammars has shifted the focus to finer linguistic decisions, such as avoiding structural ambiguity or managing complementizer insertion in English. In a related domain, Gardent and Perez-Beltrachini proposed a statistical approach to microplanning that jointly models surface realization, aggregation, and sentence segmentation. Other approaches employ classifiers to enhance output quality. For example, Filippova and Strube describe a linearization process using Maximum Entropy classifiers to determine sentence structure, while Bohnet et al. use Support Vector Machines within a cascade framework to sequentially decode semantic input, linearize syntax, and achieve appropriate morphological realization.

Chapter 2

Automated journalism and sports journalism

2.1 NLG in journalism

Natural Language Generation (NLG) technology in journalism has increased significantly, especially in sports reporting and financial news. These systems are built to process structured data inputs and generate text that is coherent, precise, and customized for the intended audience(Oh & Choi,2020). One advantage of integrating NLG in journalism is the potential to save time and resources in the article creation process, allowing journalists to focus on more complex tasks like in-depth analysis and interpretation. Additionally, NLG systems can help cover events that may be difficult or impractical for human reporters to attend, such as simultaneous sports events. However, incorporating NLG systems into journalism comes with challenges, particularly in ensuring the accuracy and reliability of the generated content, as inaccuracies or misinformation could have serious consequences(Leppanen & al,2017). An additional challenge is ensuring that the produced text is engaging and easy to understand, as poorly written or uninteresting text is unlikely to capture and maintain the interest of its readers. Natural Language Generation (NLG) systems have utilized various approaches, including rule-based frameworks, machine learning methods, and hybrid models. Furthermore, these systems have been evaluated using multiple assessment techniques, such as human judgment, readability evaluations, and quality measurements like informativeness and fluency. Now, let's analyze several well-known and widely used NLG systems that are utilized in the journalism industry:

Wordsmith: developed a commercial system called in 2011. This system has various uses such as creating news articles, sports reports, and financial reports. It uses templates, statistical methods to produce written content, and a collection of customizable pre-built templates. A mix of human assessment and automatic algorithm checks are employed to evaluate the generated text. Human reviewers assess and modify the content as necessary, considering its quality and accuracy. On the other hand, the system incorporates multiple internal quality checks and metrics to ensure the generated content meets specific standards, including readability and grammar.

ArriaNLG: Arria NLG Studio3, developed by Arria NLG and first released in 2016, is a platform designed for custom natural language generation (NLG). It enables users to create personalized NLG templates via a drag-and-drop interface, allowing for the

customization of generated text structures and content. The platform also includes pre-built templates tailored to specific industries and employs a combination of grammar-based systems and statistical methods for text generation. Its performance is assessed through human and automated evaluations, which measure the quality and relevance of the output.

In contrast, OpenAI's GPT series, while also designed for text generation, functions as autoregressive transformer language models. These models generate text sequentially, one word at a time, by predicting the next word based on prior input, using neural networks and a transformer architecture. This architecture allows GPT models to efficiently manage long-term dependencies in text, improving their text generation capabilities compared to earlier models.

GPT(Generative Pre-Trained Transformer): The initial model in this line of products was introduced in 2018(Radford and Narasimhan,2018). It is mainly utilized for tasks like producing news articles, stories, and quick replies. The model underwent pre-training on an extensive dataset, implying that it had acquired specific language patterns and features before being fine-tuned for particular tasks. It relies on statistical methods to grasp the structure and patterns of language from the training data rather than using pre-made templates or manually coded grammar-based systems. This capability enables it to generate new text that closely resembles the style and content of the training data. GPT-1 can be assessed automatically using metrics and subjectively by humans through evaluations of the generated text. Then were released, GPT-2 and GPT-3(Radford & al,2019). The most recent is ChatGPT, which uses GPT-3. Autoregressive models pre-trained consist of natural language processing models that undergo training using extensive text data. This data includes books, articles, web content, and human conversations. Models like GPT-3 are specifically created to generate text based on prompts and instructions independently. The primary aim of language models is to produce text that is deemed equivalent to human-generated text. As the evolution of these models continued, their capabilities grew exponentially, garnering both attention and debate. GPT-3 writes automatically and autonomously texts of excellent quality, on demand. Seeing it in action, we understood very well why it has made the world both enthusiastic and fearful.

2.2 The automated journalism

Carlson (2015) characterizes automated journalism as the use of software or algorithms that can transform structured data into news articles suitable for publication, all without the need for human involvement. Building on this definition,

Automated journalism is part of post-industrial journalism, a term that describes the technological hurdles impacting journalism. It entails the utilization of AI (such as software or algorithms) to automatically produce news articles with no human input, excluding the programmer(s) who created the algorithm. Automated journalism relies on natural language generation (NLG) technology to produce text-based news articles from structured digital data. Early applications of NLG technology for automating journalism have mostly focused on short texts in specific areas, but they have demonstrated impressive quality and quantity. The text generated is often indistinguishable from human-written content, and the volume of generated documents exceeds what is feasible through manual editorial processes. Automated journalism can function by autonomously creating and publishing news articles without journalist involvement, or by collaborating with a journalist who oversees the process or contributes input to enhance the article. A perfect example is The Washington Post (WP) used Heliograf software, an advanced application of AI in news writing, to cover the 2016 Olympic Games. This instance highlights how AI can efficiently manage large volumes of data to generate accurate and timely reports. Yet, as we explore the landscape of automated journalism, it's essential to understand the foundational elements required for its successful implementation.

2.2.1 The prerequisite of automated journalism

When determining the need for automating news, we look at the software development process, specifically focusing on the requirements analysis phase, as an example of "recontextualization."(Linell 1998). Fundamentally, the process of recontextualization is a practice of making sense. First, the key elements of the source domain (like the objectivity norm outlined later) must be identified. Then, figure out the most effective way to incorporate those elements into the automated text domain. In news automation, when transforming human-written news text, the genre conventions are applicable throughout the entire process of creation. These conventions encompass everything from superficial elements like font and formality to guiding principles that impact the prioritization of specific attributes of the software.

There are six requirements for automated journalism

Accuracy

Similar to journalists, a natural language generation (NLG) system used for journalistic purposes must adhere to fundamental standards of journalistic accuracy. The content generated should be backed by verifiable data and should refrain from conveying any false or deceptive information (Wright, 2015; Smiley et al., 2017). It might be suitable to include a disclaimer regarding the accuracy or validity of the data (Smiley et al., 2017).

Fluency

The need for any news automation system goes beyond being factually accurate—it also requires producing understandable output. If the system generates confusing text, it is unsuitable for journalistic use. Therefore, there is a necessity for a basic level of textual fluency. The system's level of fluency in its output is not immediately clear. For instance, one might try to establish a threshold based on whether the reader comprehends the messages conveyed by the system. However, messages with minimal fluency are unlikely to be well received by general audiences. It is reasonable to allow a news automation system targeting journalists, such as one that produces newsroom-internal "news alerts," more leeway in determining how fluent the produced texts are.

Transparency

Deuze (2005) has defined transparency as the "methods through which individuals both within and outside of journalism are provided with an opportunity to observe, scrutinize, critique, and even interfere with the journalistic procedure." Concerning automated journalism, transparency may involve disclosing the procedures used to handle the data, the analysis code, the model or conclusions derived from it, software utilized, or data origins (Diakopoulos and Koliska, 2016). This transparency, in turn,

upholds the objective of media accountability (McBride and Rosenstiel, 2013; Stark and Diakopoulos, 2016). However, this requirement may conflict with the potential business advantages of keeping the algorithmic processes concealed (Diakopoulos, 2015).

Modifiability and transferibility

The system for automating the news should be capable of easy and specific modifications to rectify errors occurring during the generation process. The need for adjustability arises from another aspect of the transparency standard, specifically a requirement for the journalistic process to be open to intervention (Deuze 2005). The modifiability can also be expressed as a need for transferability across various sectors of news production. The software must be adaptable it can be converted into a system that operates using a different dataset on an alternate subject. Many current and past news automation systems have been criticized for not meeting this requirement, considered a weakness(Linden 2017).

Data availability

The presence of (structured) data is necessary for all NLG systems, and for newsrooms, it impacts various aspects. It influences the subjects and timeliness of the generated news and the efficiency of content creation (Dorr, 2016). Furthermore, the data must be of a newsworthy nature and of high quality to ensure accurate reporting. The frequency of availability also determines how frequently an NLG system can generate news. As a result, newsrooms must first identify the data sources available before investing in NLG systems. This leads to a broader challenge: the availability of suitable data. Addressing the issue of data availability remains a significant hurdle for Natural Language Generation (NLG) systems, in spite of advancements in technology and language capabilities. Companies and clients often acquire data via open APIs without revealing the sources. NLG systems demonstrate improved performance with access to highly structured and detailed data. These systems prove to be particularly impactful in specialized domains such as sports and finance, where reliable and structured data formats (e.g., .xls, .csv, .xml, .json) are readily accessible.

However, challenges emerge when dealing with broader, more ambiguous topics or when handling unreliable or contradictory data sources, as NLG systems struggle with fact-checking and independent interpretation.

Topicality

The velocity of publication is a critical factor in digital journalism, as online content can be perpetually refreshed when live data is accessible, thus circumventing conventional production processes (Neuberger 2003). Currently, structured data sources for Natural Language Generation (NLG) are limited, primarily confined to specific domains. The concept of topicality encompasses both the rapidity of content creation within media organizations and the significance of the subject matter (Meier 2003). Consequently, access to databases—whether public or private, offline or online—reduces the time required for content generation, since NLG functions autonomously following its initial programming. The capacity of NLG to produce a substantial volume of texts at any given moment is contingent upon the availability, quality, and relevance of the data, and the technological capabilities of both the provider and the client. This increased efficiency enhances the visibility of news, particularly in terms of search engine indexing. However, it is important to acknowledge that NLG systems are not infallible; they can generate inaccuracies due to erroneous coding or flawed data sources (Dorr 2016). This situation raises significant ethical concerns, exemplified by the Associated Press (AP) ceasing to monitor each text generated by Automated Insights for their earnings reports, as the process has proven to be excessively time-consuming.

2.3 NLG tools on sport journalism

NLG has a core task on the automation of sports journalism, an important task has the tools like:

PASS: is designed to create real-time sports reports in Dutch using a template-based approach. It can tailor reports for specific audiences through a customizable and modular framework. One of its key capabilities is to adjust the tone of the reports based on the audience's emotional connection of the audience to the teams involved

in a match. For instance, if the audience's favorite team loses, the report will express disappointment or frustration, while if the team wins, it will adopt a more upbeat tone. The aim is to replicate the emotional language found in human-written reports while maintaining a professional writing style.

GoalGetter: is a system that combines speech and language generation techniques, evolved from the existing DYD (Dial Your Disk) system. It aims to demonstrate the portability of the original system across different domains, languages, and speech outputs. GoalGetter is designed to create spoken summaries of football matches using data from Teletext pages containing information about one or more matches and the teams and players involved. The system's output is a accurately pronounced Dutch text conveying information about a match from the Teletext page. GoalGetter comprises two primary modules: the text generation module (TGM) and the speech output module. One notable feature that sets GoalGetter apart from most other NLG systems is its non-use of a pipelined architecture. The TGM's general architecture consists of two modules, namely Generation and Prosody, and three data resources: 1) a series of syntactically enriched templates, 2) the Knowledge State, and 3) the Context state. By leveraging these components, GoalGetter ensures a seamless integration of text generation and speech synthesis, allowing it to effectively deliver accurate match summaries. Now, let's delve into how the system processes and utilizes specific game data to generate these outputs. The information about the outcome of a specific football game, as well as the details about the teams and their players, serves as one portion of the input for Generation. Generation also relies on a set of internal syntactic templates, which are tree structures containing placeholders to be filled with the relevant information from the input data. Certain conditions are articulated as requirements on the Knowledge State, which tracks the data that has been presented to the user and the data that has not yet been presented. The Context State stores the segments of the text that have already been created, and it encompasses a Discourse Model that maintains a record of previously mentioned discourse objects for subsequent reference expression generation.

The Multilingual Cricket Summary Generator is a system that creates summaries of cricket matches in both English and Bangla. It uses plain text files of cricket match scorecards and follows the Reiter and Dale architecture. The system works by first selecting important information (content selection) and organizing it (document plan). Then, it combines the selected information into sentences (aggregation) before using templates to turn it into natural language. Finally, a post-processor applies language-specific rules to produce the final summary. The system uses two types of templates (sentence and phrase), which can be easily expanded for more variety in the output.

Finnish ice hockey: in October 2019, Kanerva presented a system for creating Finnish ice hockey news articles using structured data. To train They gathered a dataset from

over 2000 game reports covering 20 years of ice hockey information to train the system. The authors noted the challenge of using real journalistic content due to the blend of information readily available in the statistics and details inferred from the statistics, like background knowledge, game insights, and player interviews. Relying solely on a limited number of authentic news articles for system training can cause the model to generate information unsupported by statistics. The news dataset was meticulously refined by rephrasing and removing text sections not directly substantiated by available game statistics. The authors determined that the resulting system was capable of generating text closely resembling what would be considered a viable product by human journalists while considering the word error rate. The majority of these errors were attributed to a few categories, such as inadvertently copying names from the input, event types, and time references, which would require attention in future research.

According to Chen (2008), a commentator system to learn vocabulary from sportscasts of simulated football matches, much like how children acquire language through exposure to linguistic input and perceptual experience. This system uses of a simulated environment that emulates a dynamic world with multiple agents and actions, resembling a real soccer game, but without the complications of robotics and vision.

Similarly, GazelLex, a prototype that automates the creation of soccer articles by combining deep learning with data from KGs, such as DBpedia, leverages structured data to generate soccer match reports and multimedia content (videos, audio) using RDF triples, allowing journalists to review and refine the output. GazelLex aids this process by automatically generating draft articles. Knowledge Graphs (KGs) play a key role by representing real-world entities and relationships, offering structured data that is valuable for generating news content. The GazelLex prototype utilizes Neural Machine Translation (NMT) with LSTM networks to transform RDF data into natural language, allowing journalists to edit and refine the drafts(Cremaschi & al,2019).

However, the system relies heavily on structured data and requires manual reviews to maintain article quality. Despite these challenges, tools like GazelLex show great potential for enhancing journalism by automating routine tasks, enabling journalists to focus more on critical areas such as investigation and fact-checking.

2.4 Automated journalism and sport a perfect duo?

Algorithms increasingly play a part in the newsroom process by affecting content selection, distribution, and creation of content, while automation technology aids in quickly creating articles. (Tandoc et al., 2020; Thurman et al., 2019). The advancement of advanced software that uses natural language generation and sophisticated data analysis has expanded (Haim and Graefe, 2017). However, the role of human journalists remains crucial, especially in the selection of data and the creation of templates (Waddell, 2018). In this context automating routine event coverage at high speeds and volumes is one advantage of integrating automation into sports journalism (Galily, 2018). Additionally, automation makes it possible to cover events like amateur matches that would otherwise go unreported (Van Dalen, 2012).

The particular coverage may target certain groups, like athletes and their circles, but these people will likely show deep loyalty to the media providing this coverage (Kunert, 2020). Many sports often use pre-designed formats to simplify this procedure (Gunasiri and Jayaratne, 2019; Kanerva et al., 2019). There are differing opinions among journalists regarding automated sports journalism. Some journalists have criticized the formulaic style and lack of complexity in automatically generated texts (Van der Lee, Verduijn, Krahmer et al., 2018; Thurman et al., 2017). Additionally, critics point out that datasets lack the emotional component that is often crucial in sports reporting. As one reporter remarked, "the data might only present '10% of the story'" (Thurman et al., 2017: 1247–1248). Initially, journalists felt that technology was working against them in the early stages of automated news reporting (Van Dalen, 2012). However, recent studies indicate that journalists have a more positive view of automatically generated texts and see them as a valuable tool in achieving journalistic goals, such as covering all amateur leagues and allowing more time for in-depth reporting (Kunert 2020).

2.5 Sport journalism

Journalism is the creation and dissemination of reports on the relationships between people, ideas, facts, and events that are the "news of the day" and, to the best of its accuracy. There are many types of journalism, each with a distinct target audience. Journalists are said to play the part of a fourth estate, keeping an eye on how the government operates. Numerous journalistic genres can be found in a single publication and they can all be presented in various ways. Sports journalism has ever faced a conflicting, nearly schizophrenic view within the sector. The sports department is often dismissed as being run by enthusiasts rather than professional journalists (Boyle 2006), while on the other hand, it is recognized as a valuable contributor to the financial success of news organizations, allowing coverage of important but less popular topics (Boyle 2017; Perreault and Bell 2022; Waisboard 2019). The fact that sports departments have often been labeled as "the toy department" of the newspaper is well known, which is a derogatory term used to separate the perceived serious journalism of the news desk from the enjoyment and competition of sports (Rowe, 2007). Moreover, McChesney (1989) argued that newspapers began prioritizing sports coverage because it is considered ideologically safe, not offend people, fosters civic pride, and contributes to the perceived well-being of a community. However, this ideological safety contradicts the perceived role of traditional news journalism. Because of this, news journalists often perceive sports journalism as mere entertainment, not fulfilling the role of "real journalism" (Anderson, Shirkey & Bell, 2013). While news journalism is rooted in the concept of being the fourth estate and serving as a public watchdog on public officials, the roots of sports journalism are more promotional. Michener (1976) noted, "One of the happiest relationships in American society is between sports and the media." During the 19th century, the media played a crucial role in legitimizing sport as a social institution and as a popular commercial entity. Newspaper sports journalists, through their coverage and promotional efforts, assisted in standardizing and codifying the rules of horse racing, baseball, and college football (Bryant & Holt, 2006). Furthermore, television coverage from the late 1950s and early 1960s contributed to the growth of the NFL as the country's most popular sport (MacCambridge, 2004). As the media’s relationship with sports evolved, so did the nature of sports journalism itself. The distinction between "real" journalism and "sports" journalism is a topic of ongoing debate within the field. As Choitner (2014) pointed out, there are two categories of sports journalism: coverage of important issues related to sports and society, and light-hearted game coverage, which Choitner stated serves only one purpose: "bringing joy to sports fans." Recent instances of what is considered serious sports journalism include reporting on player safety in professional football (Marx, 2011), the use of performance-enhancing drugs (Fainaru-Wada & Williams, 2007),

the rights of college athletes (Branch, 2012), and discussions about the inclusion of openly gay athletes in men's sports (Ziegler, 2012; Hardin, 2009). However, in many respects, coverage of such significant issues is viewed as the unusual, rather than the norm, in sports journalism. The majority of sports journalism is concentrated on reporting on games and issues directly related to a team's performance on the field (Boyle, 2006). The ongoing debate over the legitimacy of sports journalism, particularly its role in covering more serious societal issues versus its focus on entertainment, reflects deeper tensions within journalism. This divide becomes even more pronounced when examining the professional challenges faced by sports journalists. The challenges faced in sports journalism are particularly pronounced. While journalism asserts its credibility primarily through adherence to the principle of objectivity, sports journalism frequently finds itself excluded from this standard. The term 'toy department' is often used within the journalistic community to belittle sports journalists, who are accused of compromising the professional integrity upheld by their news counterparts (McEnnis 2020). McEnnis (2020) contends that the term "sports journalists" implies criticism for failing in normative, cognitive, and evaluative aspects, resulting in a lack of objectivity, journalistic competence, and professional distance (Boyle et al. 2012). While previous research has indicated that sports journalists themselves strongly refute these accusations (English 2017), the majority of empirical evidence supports the notion that sports journalism lacks the rigor seen in other journalistic fields, justifying its questionable reputation (Oates and Pauly 2007). Investigative work by sports journalists is rare, with sports media content mainly focusing on anticipating, describing, and reflecting on sporting events, often overlooking critical issues beyond daily sports such as doping, discrimination, match-fixing, corruption, and sports politics (cf. English 2017; Horky and Nieland 2013; Rowe 2007).

On the other hand, sports journalism follows commercial growth, it is claimed by many that sports play a key role in the commercialization and globalization of the media. For many years, sports supplements in the written press have significantly grown in number and size and are read almost more than any other newspaper supplement. Indeed, for many, the sports supplement is the central reason for purchasing the newspaper, despite the technological developments discussed in this paper. Research in sports media has therefore focused a great deal on the reciprocal relations between media and sports. Emphasis has been placed on media content, while somehow neglecting the two poles of the transfer of information axis: Information producers (responsible for the organizational structure, the processes of recruitment and the production of sports information distributed by them) and the reading public. This dynamic has broader implications when considering the commercial landscape of sports journalism on an international scale.

In the field of sports journalism, the pursuit of economic capital is dominant. Although this is not surprising given the commercial nature of the analyzed media and previous research in the industry (see English, 2013b; Hamilton, 2004; McManus, 2009; Picard, 2004; Schultz-Jørgensen, 2005), it is unexpected to find distinct differences in the goals of companies in India and the United Kingdom. The placement of quality organizations within each media system results in significant gaps between the two newsrooms in India and the United Kingdom, positioning them relatively close to each other above the mid-point line of total capital and on the economic capital side. The United Kingdom has slightly more capital in both areas due to the high levels of journalistic and economic value when the media organizations are combined.

2.6 The spread of sport journalism

The sports journalism has a long tradition. It started in the 19th century, especially in Western Countries. A leading country is the United States of America, in fact during the time of the Industrial Revolution, the increase in urbanization and advancements in technology led to a decrease in the cost of collecting and publishing news, causing a significant rise in newspaper circulation. It was during this period of yellow journalism that the allure of attracting readers made sports coverage a logical focus. This resulted in a substantial expansion of newspaper sports coverage. Joseph Pulitzer's New York World made history by establishing its own sports department in 1883, while William Randolph Hearst's New York Journal introduced the first dedicated sports section in 1895, signaling a new era in sports coverage within newspapers. In the late 19th and early 20th centuries, there was a continuous rise in the significance of sports journalism. According to Schlesinger (1933), American newspapers allocated a mere 0.04 percent of their content to sports coverage in 1880. However, by 1920, this had increased to 12-20 percent of the newspaper's overall coverage. McChesney (1989) observed that by the mid-1920s, almost all newspapers in the United States featured a section specifically devoted to sports.. McChesney also stated that this period marked the emergence of sports journalism as a distinct form of journalism, and it became an essential part of everyday newspaper content. This time is often referred to as the Golden Age of Sports Journalism (Boyle, 2006; Bryant & Holt, 2006), during which renowned reporters like Grantland Rice and Damon Runyon provided coverage of sports for newspapers. The establishment of sports journalism as a distinct genre has shaped the practices of sports journalists. According to popular accounts, sports journalists traditionally attended games, took notes, and conducted interviews with coaches and star players either in press

conferences or locker-room settings, before meeting deadlines to write their stories (Walsh, 2006; Wilstein, 2002; Vecsey, 1986). Throughout the 20th century, sports remained heavily mediated, with newspaper sports journalists adapting their work routines in response to technological advancements such as radio and television broadcasts of games and events. Vecsey (1986) noted that in the 1960s, reporters sent their stories via Western Union telegram when covering road games, while by the mid-1980s, they were using portable word processors and computers. Walsh (2006) described his daily work as a sports reporter in the 1990s and early 2000s, covering games, practices, and breaking news at newspapers in Florida, Arizona, and Wisconsin. A notable shift in the profession of sports journalism was brought about by the expansion of game broadcasts, initially on radio and later on television. Newspaper journalists were forced to refocus their coverage due to game broadcasts. "Sports writing has become a secondary medium to television, with its primary role now being to uncover the story behind the story". (Oriard, 1993). Sports journalists started utilizing their game stories to provide more analysis, color, and perspectives from players and coaches rather than relying primarily on play-by-play descriptions.

In the 1920s and 1930s, a reaction to radio broadcasts marked the beginning of this phenomenon (Bryant & Holt, 2006), which continued to evolve with the expansion of television coverage in the 1950s and 1960s (McChesney, 1989). The use of quotes became an integral aspect of sports journalists' work, serving as a means to set themselves apart from other media outlets (Vecsey, 1986), and reporters, much like other journalists, were evaluated based on the quality of their sources (Boyle, 2006). According to McChesney (1989), the advent of TV coverage brought about changes in the approach of newspaper sports journalists. Stories became less focused on game recaps and increasingly relied on statistics, analysis, and background information. By the mid-1980s, a dedicated 24-hour sports coverage cable channel, ESPN, emerged, and sports occupied 25 percent of the daily space in the nationally circulated print newspaper, USA Today, compared to 12-20 percent in most local newspapers (McChesney, 1989). Despite this growth, the fundamental approach of newspaper sports coverage remained largely unchanged, a trend that persisted into the digital media era.

As the United States was developing its brand of sports journalism, similar trends were unfolding across Europe. The role of the press within the social structure of various nations, as well as its evolution in the coverage of football, may exhibit slight variations; however, a consistent trend of growth is evident throughout the examined period. Notably, English, French, and Spanish newspapers demonstrate comparable patterns in their subject matter. Initially, in the early twentieth century, coverage primarily consisted of straightforward reporting of facts. This gradually transitioned to more analytical discussions regarding team selections, while still maintaining a

focus on match-related topics. Over time, the prose became increasingly inventive, evolving from descriptive reporting to interpretative journalism. This shift encompassed a wide range of football-related interests, including serious business, financial implications, social issues, and even lighter topics such as trivia and gossip. Consequently, across the entire dataset, it is apparent that later years feature a significantly greater volume of material compared to earlier periods. The coverage of football in newspapers at the beginning of the century was markedly less extensive than that observed a century later. In the early 1900s in France, writers in Le Vélo, L'Auto, and L'Echo des sports had a biased view of sports and paid little attention to football (Wahl 1989). Rugby was favored by the press before the Great War, as it was the game played by editors and journalists. For example, L'Auto frequently featured rugby on its front page but never football before 1914, despite the existence of over 350 registered football clubs in France in comparison to about 140 rugby clubs by the 1900s (Hare 2003). It wasn't until the inter-war years that football and sports in general became prevalent in general newspapers, albeit overshadowed by cycling, rugby, and motor sports (Berthou 1999). The rise of football's importance in France was evident through the establishment of specialized football publications such as Football (founded in 1910), Football et Sports athlétiques (1909, bi-weekly), and France Football (founded in 1923 and still active today as a bi-weekly) (Berthou 1999: 9; Wahl 1989: 352). The well-known sports daily, L'Equipe, made its debut relatively late in 1946. Following the professionalization of football in 1932, the French press started dedicating more space to football (Wahl 1989: 314). However, some publications, like L'Auto, didn't cover the first professional league matches and instead prioritized covering the Italian Grand Prix (Berthou 1999: 11). Albert (1990: 163) observed that the French press in the inter-war years had to adjust to new trends and interests of the readers, including travel, cinema, radio, and, of course, sports. To adapt to the changing economic environment, newspapers increased their daily pagination significantly by the late 1930s and also incorporated more photo journalism. Charle (2004) considers this early twentieth-century period as the golden age of the French press, where France excelled in circulation figures and journalistic quality. However, economic challenges during the inter-war years, coupled with the rise of competing media, led to a decline in the national daily press's fortunes, although the sports press continued to have significant influence. According to Seidler (1964), by the late 1950s and early 1960s, France was one of the countries where the sports press held the most importance. During the 1980s, football in France experienced a marked shift in public perception, particularly reflected in its increasing prominence within the 'quality' daily press. While some journalists, such as one writing in *Le Monde* on 3 July 1982, expressed concern over football's dominance, referring to it as an "obsession" and a "tyranny" that overshadowed other topics during major tournaments, others acknowledged its cultural significance.

Football was compared to theatre and viewed as more culturally impactful than other forms of art, such as music. Though it may embody aggressive "male, warlike values," it was perceived as doing so in an acceptable manner, where players acted as symbolic warriors representing their club or nation. Importantly, football had come to replace traditional institutions like religion and the military as a key marker of collective identity, with footballers now occupying the role of contemporary heroes. As such, its increasing coverage in the press was justified. This shift was further evidenced by *Le Monde's* evolving approach, including the occasional use of photographs and expanded coverage of events like the 1986 World Cup in Mexico, where football reports took up entire pages, signaling its growing importance.

In Spain, sports coverage in the general press began in the late 19th century, though specialized sports publications had already emerged earlier. According to Altabella (1987), *El Cazador*, published in Barcelona in 1856/57, was the first such publication, while *El Mundo Deportivo*, also based in Barcelona, is the oldest surviving sports paper, established in 1906 as a weekly and becoming a daily in 1929. However, the first daily sports publication was *Excelsior*, printed in Bilbao between 1924 and 1931. During this period, Spaniards with an interest in sports accessed information from generalist dailies, specialist dailies, *Hojas del lunes* (special Monday editions), and magazines focused on individual sports (Jones & Baró i Queralt, 1996).

Between 1910 and 1920, a surge in leisure time and mass media led to the creation of approximately 30 specialized sports titles in Catalunya, many of which had Catalan nationalist leanings. However, most of these publications did not survive the Spanish Civil War (1936–39) due to paper shortages and limited sports activity.

Throughout the first half of the 20th century, sports coverage in the Spanish national press gradually expanded, with football becoming increasingly prominent. By 1925, football had overtaken other sports, except for bullfighting, in media coverage. By the 1930s, football dominated the sports pages, earning the title *el deporte rey* ("the king of sports"), although coverage remained limited and focused primarily on factual information.

Today, *Marca* leads the Spanish sports press, accounting for over 60% of sales, followed by *As*, *Sport*, and *El Mundo Deportivo*. Established in 1938, *Marca* moved to Madrid in 1942 and competes with *El País* as one of Spain's most widely read newspapers.

As the landscape of traditional sports journalism continued to evolve, a significant transformation was on the horizon with the advent of the digital age. This shift, which began in the final decades of the 20th century, brought about new challenges and opportunities for sports journalists. The rise of the internet and digital platforms led to

a dramatic change in how sports news was consumed and produced. No longer confined to print media and television, sports journalism expanded into a vast digital space, allowing for real-time updates, multimedia content, and interactive fan engagement. This transition paved the way for a new chapter in sports journalism—one characterized by immediacy, accessibility, and a more personal connection with audiences.

2.7 The origin of digital journalism

In the last decade of the 20th century, there is the spread of digital sport journalism. New media, particularly the Internet, digital television, and mobile telephony, are creating new platforms and services for the distribution of sports content. These services are defined by three key processes: digitization, convergence, and interactivity. Digitization occurs at multiple stages of sports-media production, encompassing both hardware and software used for capturing and editing audio-visual signals. From image capture to editing, the entire process has been digitized. Moreover, the global reach of sports information via digital networks is transforming the way breaking sports news is gathered, selected, and disseminated, significantly increasing the speed at which journalists can share and publish information. This acceleration is reshaping audience expectations for sports news consumption. However, digitization introduces significant challenges for protecting copyright and intellectual property in sports, as the ease of reproducing and distributing digital content complicates regulatory and technical mechanisms aimed at controlling the flow of media(Boyle & Haynes,2002). The first website of ESPN, then known as ESPNSportsZone, was launched in 1995 and rebranded as ESPN.com in 1998 (Bryant & Holt, 2006). ESPNET SportsZone, launched in April, rapidly gained prominence as one of the Internet's most widely utilized platforms, attracting 120,000 daily users. The service offers a comprehensive range of sports-related content, encompassing 15,000 pages including news, real-time scores, statistics, features, interactive games, polls, and chat functionalities. In November, a groundbreaking broadband application was unveiled, promising significant improvements in audio and video quality, alongside a substantial enhancement in data transmission speed, scheduled for implementation in 1996. Similarly, other television networks such as Fox and CBS also introduced sports-focused websites during this period (Bryant & Holt, 2006). Websites like Fox have become critical strategic tools for sports organizations, offering a convenient bridge between fans and organizations, and are increasingly utilized to achieve marketing objectives and enhance customer satisfaction. Scholars in sports management have recognized this trend and have investigated various aspects of sports websites, including their potential roles,

revenue models, and user demographics (Kahle & Meeske, 1999; Caskey & Delpy, 1999; Delpy & Bosetti, 1999). As the population of online sports fans grows, sports organizations have increasingly adopted the internet as a marketing tool to maintain fan engagement. They provide features such as news, scores, blogs, online stores, and fan forums. Major US sports leagues, including the NFL and NBA, have revamped their websites to enhance visual appeal and integrate blogs and social networking platforms, allowing fans to share opinions and buy event tickets (Stories of the Year, 2006).

Research has also focused on content analysis (Smith, Pent & Pitts, 1999) and the integration of TV and internet services (Turner, 1999). More recent studies have shifted toward understanding users' motivations for engaging with sports websites (Hur, Ko & Valacich, 2007; Seo & Green, 2008) and examining virtual advertising strategies (Yu, 2007; Tsuji, Bennett & Leigh, 2009). Additionally, some scholars have explored how professional sports teams can optimize marketing through their websites (Carlson, Rosenberger & Muthaly, 2003; Ciletti et al., 2010), and how to attract fans to newspaper and entertainment websites (Butler & Sagas, 2007; Kitchin, 2006). However, there remains a gap in research applying theoretical models to explain sports fans' behavior on these websites, which this study seeks to address by focusing specifically on fans' interactions with sports web portals. In 1997, the internet search engine Yahoo launched a website dedicated solely to sports. In 2009, ESPN initiated a series of microsites devoted to sports coverage in specific cities, including New York, Chicago, Los Angeles, and Miami (Barnes, 2009). ESPN already operates local sports coverage in major cities, including New York, Chicago, Boston, Dallas, and Los Angeles, as well as blogs covering NFL conferences. King emphasized that all existing staff would remain, with 19 new hires planned, and offers extended for all but three positions. These operations will largely consist of single-person teams supported by editors and producers to distribute content across various platforms. The initiative aims to provide fans with more in-depth coverage of their favorite teams, posing a competitive challenge to traditional newspapers, particularly as ESPN’s coverage will remain free, in contrast to the growing trend of paywalls at many local newspapers; according to Halliday(2012) also to eliminate competition from the local press. In 2013, ESPN recruited writers to cover all 32 NFL teams. These writers are referred to as "bloggers" by ESPN (Beaujon, 2013), ESPN is extending its hyperlocal coverage to New York, Los Angeles, and Dallas. This move is seen as a significant shift in the sports media landscape, as ESPN aims to dominate local sports news, potentially challenging local newspapers and other outlets.

In Chicago, ESPN has already outperformed local sports sections, attracting a higher number of unique visitors. Their model integrates national resources with localized content, covering professional, college, and high school sports. ESPN’s strategy

involves leveraging existing assets, such as their radio stations and limited new hires, to control costs. This expansion aims to capture local advertising revenue, particularly as online ad markets grow.

While ESPN's foray into local markets has met some resistance from established news outlets, it is positioning itself as an additive service, with plans to deepen its reach through community engagement and social networking features. The company's long-term goal is to transform its localized efforts into a significant growth engine, despite initial challenges related to content control and cost management.

The digital sports journalism practiced on these news websites generally adheres to the traditional journalism principles of objectivity (Scott, 2012; Krueger, 2010). For example according to Krueger(2010) Yahoo! Sports has distinguished itself within the crowded sports journalism landscape by focusing on enterprise journalism, particularly investigative reporting, rather than competing directly with outlets like ESPN or AOL Fanhouse in multimedia content. According to Tim Franklin, director of the National Sports Journalism Center, Yahoo! Sports strategically chose to specialize in investigative stories, such as the high-profile USC football and Reggie Bush scandals, positioning itself as a leader in this niche. The decision to focus on long-term, in-depth reporting allows Yahoo! to capitalize on the decline of investigative resources in traditional media, which has been hindered by budget cuts and reduced newsroom staff.

Yahoo! Sports' investigative success is supported by a team with a background in traditional journalism, emphasizing the need for patience and resources. Gerry Ahern, Yahoo!'s lead investigative editor, highlighted that the platform has benefited from the shift in old media, building a team suited for such reporting. However, Yahoo! still faces challenges, as investigative reporting requires time and financial support. Charles Robinson, Yahoo!'s only full-time investigative reporter, works closely with other reporters, using their expertise to pursue different stories.

Despite limited resources, Yahoo! Sports applies rigorous journalistic standards, including verification and transparency, to ensure high-quality reporting. Ahern and Robinson emphasize that the online medium, though perceived as less strict, maintains stringent standards akin to traditional outlets. Additionally, the online format offers advantages, such as longer shelf-life for stories, broader reach, and the ability to provide supporting documents. Robinson noted that online audiences are versatile, consuming both brief social media updates and long-form investigative content.

Yahoo! Sports' commitment to investigative journalism has proven successful, earning both industry respect and a leading position in sports website page views.

Franklin commended the platform’s focus on high-quality, original content, asserting that this strategy is key to thriving in the new media landscape.

On the other hand, blogs are known for expressing their enthusiasm. While traditional sports journalism has relied on player access for quotes, insights, and information, many blogs take pride in their lack of access. Since its inception, Deadspin's motto has been "sports news without access, favor, or discretion" (Deadspin.com, 2014). Deadspin is a sports blog founded by Will Leitch in 2005 and based in Chicago. Deadspin regularly shared daily previews, recaps, and commentary on major sports events, along with sports-related anecdotes, rumors, and videos. Deadspin regularly shared daily previews, recaps, and commentary on major sports events, along with sports-related anecdotes, rumors, and videos. As its content became increasingly popular**,** Deadspin's trajectory, from its rise to its eventual decline, was shaped by its commitment to self-professed ideals, as explored through news coverage and interviews with sports media professionals. Despite limited scholarly attention (with exceptions such as Burroughs and Vogan, 2015; Eldridge, 2019), Deadspin had a significant impact on journalism. As one of the most popular sports blogs, it attracted millions of visitors, peaking at 41 million monthly page views (Janssen, 2019; Koo, 2020). Its success can be attributed to what has been described as "punk journalism."

Punk journalism draws from two key traditions within journalism studies. First, it aligns with alternative journalism, a tradition that empowers marginalized voices and challenges mainstream news norms, often critiquing the political economy of corporate media (Atton and Hamilton, 2008; Bagdikian, 2004; McChesney, 2004). This approach echoes the ethos of punk theory, which views corporate media as repressive and outdated. Second, punk journalism intersects with Eldridge's (2018) concept of "interloper" journalism, referring to nontraditional digital entities like Reddit, Breitbart, and Gawker. These outlets, including Deadspin, operate on the fringes of traditional journalism, challenging the boundaries of legitimacy while often being labeled as contrarian, problematic, or heretical. This characterization fits both punk journalism's broader ideology and Deadspin's particular role in media.

Another sports journal like Bleacher Report (B/R) is an American sports news website founded in 2007 by a group of friends, which has gained notable popularity, ranking 113th in U.S. web traffic and 339th globally according to Alexa.com. While it covers global sports, the website's primary focus is on American sports news. As a sports journalism platform, B/R tends to focus on popular topics, consistent with Rowe's (2005) analysis of sports journalism's orientation. However, as Oates and Pauly (2007) argue, sports journalism remains embedded within the broader journalistic culture, facing similar reporting and ethical challenges.

Initially, B/R followed a start-up model, launching in 2008 with a venture capital-backed funding structure that included four major rounds of funding by 2011 (Lacy, 2011). Its approach to journalism was unconventional, relying on a citizen journalism model and an open-source platform that allowed journalists, writers, and fans to collaborate in content production (Ostrow, 2008).

In contrast to traditional news outlets, B/R exhibits several distinct characteristics. First, it operates as an online-only news provider, although it has recently expanded into video and broadcasting for other outlets. Second, as a sports-focused platform, it blurs the lines between entertainment and news, navigating the boundary between hard and soft news more freely than mainstream journalism. Lastly, the company was founded by young entrepreneurs with no formal journalism training or experience, setting it apart from legacy media organizations.

Bleacher Report has implemented rules prohibiting writers from attempting to break the news and instead positions itself as a site "written by fans for fans" (Eskenazi, 2012). The article by Joe Eskenazi in SF Weekly delves into Bleacher Report, analyzing its rapid ascent to becoming the third-most-visited sports site globally, with a value of $175 million as estimated by Turner. The piece points out the company's heavy reliance on unpaid writers and data-centric content creation, serving as an extreme example of modern journalism. Bleacher Report utilizes an analytics team to predict popular topics and guide writers to tailor their stories accordingly. This method places emphasis on SEO and market demand rather than editorial independence, resulting in a content structure often starting with sensational or misleading headlines aimed at maximizing reader engagement. Despite its "for the fans" image, Bleacher Report operates in a top-down manner, where editors formulate headlines based on data trends and writers follow suit. The platform encourages writers to exaggerate or downplay their claims and to use eye-catching headlines like "LeBron James Signing Makes the Miami Heat the Best Team in NBA History" to attract readers. Writers are directed to focus on popular sports and culture stories and even take contrarian stances to provoke debate, as seen in the headline "Why Tom Brady Is the Most Overrated Quarterback in NFL History." While Bleacher Report introduced the Lead Writer program to enhance content quality, some lead writers are dissatisfied, feeling their work is being watered down for mass appeal. One writer questioned why top-tier talent was being paid to create content that could been produced by unpaid contributors. The site strategically employs well-known writers to enhance credibility while using low-quality material to drive traffic. Ultimately, the criticism goes beyond Bleacher Report itself, suggesting that many trends, such as sensationalism and market-driven content, are widespread in modern journalism. However, Bleacher Report's complete embrace of these practices

represents a troubling shift, reducing editorial decision-making to algorithmic predictions and undermining journalistic integrity.

What's new?

The intersection of automated journalism and sports journalism, focusing on the role of Natural Language Generation (NLG) systems in news reporting. NLG has demonstrated remarkable capabilities in automating sports reports, financial news, and other structured data-based journalism. While automated systems like Wordsmith, ArriaNLG, and OpenAI's GPT series have proven efficient in generating coherent and timely articles, challenges remain in terms of accuracy, fluency, transparency, and adaptability. Automated journalism also raises ethical concerns, such as maintaining journalistic integrity and ensuring data quality.

In sports journalism, automation has shown great potential, particularly in covering events that might otherwise go unnoticed, like amateur sports. Systems such as PASS and GoalGetter allow for real-time reporting and personalized emotional tones based on audience preferences. Despite the efficiency and scalability offered by NLG, there is ongoing debate within the journalism community about the formulaic nature of automated texts and their inability to capture the emotional depth often crucial in sports reporting.

Historically, sports journalism has evolved significantly, especially with the rise of digital platforms, which have redefined how news is consumed and produced. The chapter highlighted the tension between traditional journalistic standards and the commercial nature of sports coverage, as well as the growing importance of investigative sports journalism. The spread of sports journalism, both in print and digital formats, reflects broader societal shifts and technological advancements, with sports coverage becoming a central feature in modern media.

Ultimately, while automated journalism offers numerous advantages, the role of human journalists remains essential, especially in delivering in-depth analysis, investigative reporting, and capturing the emotional essence of sports. The integration of NLG into sports journalism presents a promising yet complex future for the field, balancing efficiency with the need for quality, personalized, and engaging content.

Chapter 3

The journalism's development from digital to AI

3.1 The definition of digital journalism

Defining digital journalism is not simple. It may be a surprise that, even though digital journalism has been around for more than 20 years, international scholars have yet to develop a common term for it. There isn't even consensus on what this kind of journalism is called or how to define it. Even now, experts and scholars disagree when it comes to deciding between terms like "digital journalism" (Kawamoto, 2003); It may seem strange that 30 years after the advent of journalism in digital media, there is still no standard name; many academics disagree when it comes to name choices such as 'digital journalism', 'cyber journalism' and 'online journalism'. But there are some points of agreement, such as:

Hypertextuality

Hypertext is understood as a group of non-linear computer-based texts (i.e. written texts, images, etc.) linked together with hyperlinks. The term was first coined by Nelson (1965), who described it rather loosely as 'a series of blocks of text connected by links that offer the reader different paths' (cited in Tsay, 2009). Most scholars researching hypertext in online journalism rely on what Aarseth calls an 'industrial rhetoric of the computer' (1997), i.e. an understanding of hypertext as a technological function (made visible by the electronic link) rather than as an observable practice of interaction between text and reader. Researchers interested in hypertext as a text-reader practice are more likely to call the object of study a practice of interactivity rather than a practice of hypertext. The general assumption of researchers interested in online hypertext journalism is that, if hypertext were used innovatively, it would provide several advantages over print journalism: no space limitations, the possibility of offering a variety of perspectives, no finite deadlines, direct access to sources, personalized paths of perceiving and reading the news, contextualization of breaking news, and simultaneous targeting of different groups of readers, both those interested only in the headlines and those interested in the deeper layers of information and sources.

In addition in hyper textuality it is possible add wikijournalism, that have a relevant impact on digital journalism and it was launched in 2004 to promote partecipatory citizien. Andrew Lih supported the idea of wikijournalism and participatory journalism including blogs. As Bowman and Willis argue define participatory journalism as “The act of a citizen, or group of citizens, playing an active role in the process of collecting”. Moreover wikijournalism presents pro and cons. The advantages, wikis provide news operations with a tool to cover complex issues by enabling broad collaboration. They facilitate contributions from the public on diverse topics such as local transport problems, events, or local guides, which would be challenging for a single journalist to cover comprehensively. Jay Rosen highlights that traditional newsrooms, being closed systems, can't easily leverage the collective knowledge of their audience, a limitation wikis can overcome.

Wikis also help news organizations manage intricate stories with multiple contributors, and enhance transparency by allowing the public to view the development process and participate in discussions. This open collaboration can lead to translations and wider accessibility of content, engaging hard-to-reach communities.

Economically, user-generated content from wikis can help news sites increase user engagement and time spent on their platforms, which attracts advertisers. This approach allows professional journalists to focus on unique, high-quality journalism, while community contributions fill other sections. However, managing wikis is not without costs. Effective operation requires significant maintenance, monitoring, and skilled management to ensure quality and cultivate community, contradicting the notion that they are cheap to run. Despite these challenges, wikis can foster strong community bonds and reader loyalty, offering both economic and competitive benefits to news organizations.

But wikijournalism has also negative aspects like inaccuracy and vandalism, this aspect is very recurrent and wikis like wikipedia use a “soft security” approach, making damage repairable rather than preventable and Editorial presence is recommended to maintain community trust and narrative coherence. For news organizations, the long-term commitment to building wiki communities and the short lifespan of news articles pose challenges. Additionally, issues around authorship recognition and remuneration in collaborative sites need to be addressed, as current disclaimers focus more on legal protection than on acknowledging contributors' ownership.

Interactivity

it is the process of engaging humans or the participation of machines in the process of informing and trying to share information;(this is what Kawamoto argues), also research on interactivity.

Interactivity is a slippery concept that is used to describe numerous processes related to communication in general and practices such as online journalism in particular. Based on a review of the 'history' of interactivity, Jensen arrives at this definition: interactivity is 'a measure of the potential ability of a media to allow the user to exert influence on the content and/or form of the mediated communication' (1998, p. 201). Jensen separates interaction from interactivity and his definition is thus primarily technological. Interaction refers to the social dimension of interactivity, and McMillan advocates the incorporation of this dimension as well. Accordingly, he identified nine different interpretations of interactivity along two different axes (McMillan, 2002, 2005). All these types of interactivity can be found in an online journal. However, the Human-Computer axis (or 'navigational interactivity', as Deuze (2001) calls it) is similar to what, in the previous section, I categorized as hypertext. Research covering this axis has therefore been included there. Of the six remaining notions of interactivity, only two seem to have significantly occupied researchers of interactivity in online journalism: human-human (both features and processes). This research is dominated by questions such as the degree to which users are allowed to interact with online newsrooms/journalists through email; the extent to which online news sites offer discussion forums; and whether users are allowed to comment on news or in other ways be involved in the production process.

Multimedia

Its actual development in online media is far more sophisticated and varied than hypertext. The popularity of online video, which in 2018 accounted for three quarters of all Internet traffic worldwide, according to various consulting firms, has prompted online publications to investigate multimedia narratives. In this instance, scholarly investigations have concentrated on characterizing the categories of multimedia packages in journalism and their impact on the public's news consumption. According to Deuze , there are two common understandings of what multimedia is in online journalism studies: when a news story package is presented using two or more media formats (text, audio, video, graphics, etc.) either as a news story distributed via a

variety of media (such as a newspaper, website, radio, television, etc.). Most of studies conducted on multimedia in online journalism focus on initial comprehension. Moreover, the term "multimedia" will be used in a slightly more practical sense, aligning with empirical studies on multimedia in online journalism. Multimedia will be defined as stories and websites that use more than two media, as an online news story with just text and a photo is generally not considered multimedia. Additionally, this term will refer to both the creation and the presentation of news.

3.2 The history of digital journalism

3.2.1 The birth in 1990s

The 1990s represented a turning point from the point of view of journalism, the entry of technology was a meaningful change. Journalism joins digital and the first online newspapers are born in this period. According to David Carlson (2003), the first news article on the World Wide Web debuted on January 19, 1994. The weekly newspaper Palo Alto Weekly, which is published in the San Francisco Bay Area, has the distinction of being the first in the history of Web-based journalism. But before this, Several major media companies, including America OnLine (AOL), Compuserve, and Prodigy, had already started their online versions on those commercial networks. The Chicago Tribune was the first to publish an edition on AOL, doing so in March 1992 (Díaz-Noci; Meso-Ayerdi, 1998). Large newspapers like The New York Times, The Washington Post, The Los Angeles Times, and USA Today came after this. The rapid speed digital media led to their instant emergence as the fourth media, competing with print, radio, and television (Bonnington, 1995). The rise of the web in the 1990s was of the world wide web facilitated the spread of digital journalism like accessibility and breaking news H24, and also economic and quickness. The major spread was in the year 1994-96 with the diffusion of web browser like Netscape, Windows and Internet Explorer. The United States has set the precedents for other nations to follow. The majority of nations saw the emergence of their first digital media outlets between 1994 and 1996; these outlets were nearly invariably propelled by newspapers. However, historical studies of digital media still have much to learn because few nations have published monographs detailing the inception and early development of online media within their borders. However, a large number of these kinds of studies covering nations worldwide have been published. Historical accounts of the development of online media are available in numerous nations, most of them in Western countries. On the other hand, it so happened that at about the same time, most of these countries saw the emergence of their first online media. By the mid-1990s, almost every country in the world had been affected by this wave, except for

the most economically and technologically backward nations, as well as those with authoritarian governments and harsh restrictions on press freedom. Online journalism had just begun, thanks to these trailblazing outlets that first appeared on commercial networks and then, soon after, on the World Wide Web.

3.2.2. The new millenium and and the rise of web 2.0

The widespread adoption of online journalism gives a new way to do journalism, in the new century we can see the explosion of interaction and the explosion of social networks. Defining Web 2.0 is not easy. Cormode and Krishnamurthy(2008) noted that "there is a clear separation between a set of highly popular Web 2.0 sites such as Facebook and YouTube and the "old web." The two scholars showed the main difference between Web 2.0 and web 1.0 lies in the interactive and accessible nature for users. The web 1.0 was characterized by static web pages and a lack of interactive content. Websites were primarily a collection of static HTML pages that provided information but offered little to no interaction for users, more over the user interaction were often read-only with limited interaction. The web 2.0 had a upgrade like a more interactive and user-centered web. It was characterized by dynamic web pages, user-generated content, and the ability to interact with the website and with other users. Key technologies include AJAX[1] and APIs[2], and core features often involve social networking and collaboration, and also users are very participating and sharing contents. Since the goal of this study was to investigate the new media platforms that sports journalists are using rather than to focus on any one platform or genre of new media, the term Web 2.0 was chosen as the overall framework. In addiction web 2.0 is useful for sports journalism and the social media platform. In particular Facebook, Twitter and fans forum.

Facebook was born in 2004 by Mark Zuckenberg exactly in Massachusetts, and then Facebook began adding American universities to its network shortly after it launched, starting with Stanford, Columbia, and Yale in February and March. Later, in March, it added New York University, MIT, Cornell University, and other universities. On Facebook's open contact page, students from other universities were asked to recommend additional network additions. By May of the subsequent year, eight hundred American universities had access to Facebook. Students in upper secondary schools could access it starting in September 2005, and in October of the same year, Facebook—whose name was changed in September 2005—began to open up to

[1] Asyncrhronous JavaScript and XML , is a set of web development techniques that uses various web technologies on the client-side to create asynchronous web applications

[2] API: application programming interface s a way for two or more computer programs or components to communicate with each other. It is a type of software interface, offering a service to other pieces of software.

educational institutions outside of the United States. It was not surprising that Facebook grew quickly: two months after its debut in April 2004, there were only 70,000 users; by December, that number had risen to one million, and by December 2005, there were over six million users. The network and the profile were the two primary components of Facebook. If privacy settings allowed, the personal profile was linked to a specific individual and visible to all Facebook users.

It might contain details about your contact details, political beliefs, relationship status, favorite songs, movies, and books, as well as a picture and information about your place of education. Furthermore, the profile page indicated if you were signed in and if you were using a location that was listed in Facebook's college database. The profile cannot have been made for anyone other than yourself, including dead people, places, animals, and fictional characters. As the owner of a profile, you were included in a large social network consisting of “the group of all users whose privacy settings allow you to view their information”. But you can also choose to link to other profiles — friends — via hyperlinks. This network of friends was shown on your profile page via lists with photos, either as a list of friends or a list of classmates. In addition, you could see your friends’ lists of friends, and when you visited somebody else’s profile, you could see which friends you share. The most fundamental type of sharing on Thefacebook was that user profile information was visible to the user's network because users were socially networking with each other. However, the sharing options were expanded approximately six months after the launch with the addition of a "Wall" on each profile page, allowing users to write on their walls and post messages for their friends. Messages could now be shared between users, and later on images could also be shared (starting in October 2005). A further changing point was in 2006 Facebook opened up to the world outside of education, for starters. And yet another is that a great deal of new features that drastically alter Facebook were introduced starting in 2006. Facebook ceased to be a network exclusively for students in May 2006 and became accessible to workers of specific companies, such as Apple and Microsoft. More over Facebook approached also journalism in fact the social shaping of technology is the dominant paradigm in the literature on digital journalism(Paulussen 2013). The concept of "mutual shaping" describes how journalism and technology interact: Researchers in the SST tradition look into the everyday social and cultural practices that influence how people use technology and, in turn, the potentially novel practices that result from these social interactions with technology, as opposed to presuming that technology has a linear impact on society. (Lievrouw, 2006). When the SST perspective is applied to the study of social media and journalism, it suggests that social media should not be viewed as merely enabling technologies, but rather as platforms that both shape and are shaped by how these new technologies are being used, giving rise to new communication patterns and practices. As "spaces for audiences to share, discuss, and contribute to the news,"

social media platforms are often seen by researchers as "networked public spheres" in this sense (Hermida et al., 2012). At first glance, Facebook's influence on journalists' news selection processes might seem limited. But it's important to understand that journalists use Facebook not just as a direct source for their stories but also as a vital tool for keeping an eye on the never-ending, round-the-clock news feed.

Facebook serves two purposes, which highlights how complicated its role is in modern journalism. To learn more about how journalists' use of social media influences their judgments of newsworthiness, more research is required. Particularly important is the impact that popular or viral subjects have on journalists' news values on social media (Tandoc, 2014). Furthermore, Hermida (2015) notes that, instead of using traditional professional gatekeepers, "ad hoc publics" on social media can elevate certain individuals to the status of reliable sources. The power dynamics between journalists and their sources may change as a result of this change, since these social media players are now able to exert influence "on specific issues at specific times, within specific contexts or domains." Comprehending these subtleties is essential to comprehending the wider consequences of social media on journalistic practices and news values and calls for additional scholarly investigation.

Similarly, Facebook's role transcends being a mere newsgathering tool or an additional platform for news dissemination. Its impact lies in its potential to reshape the flow of news, altering the cultural and commercial interactions between social and legacy media, as well as between the ‘producers/suppliers’ and ‘users/recipients’ of news. Singer (2014) emphasizes that users are becoming "secondary gatekeepers" of the content published on news websites, as they increasingly engage in the recommendation and "selective re-dissemination" of that content.

Together, these perspectives underscore a significant shift in the power relations within the news distribution process. Social media platforms like Facebook not only empower individual users and "ad hoc publics" to influence news dissemination but also challenge traditional journalistic gatekeeping, necessitating a deeper understanding of these evolving dynamics.

Twitter

Another social network that changed the internet is Twitter, which is a microblogging platform that lets users communicate with one another using features like "@ replies," which let users tag other users, and "hashtags," which let users add a hash (#) to a message a word or phrase makes it a link that can be searched[3]. After its

launch in 2006, Twitter announced in September 2011 that it had achieved the milestone of 100 million active users, with 50 million users accessing the service daily (Tsukayama, 2011). Twitter can be seen as a content-delivery system that increases the competitiveness of news operations. Either by making direct connections with users or by directing them to other resources. Twitter promises to grow audience, boost revenue, and distribute content more effectively.

Increasingly, traditional media are exploring new technologies to stay economically competitive. Tom Rosenstiel, the director of the Project in Excellence in Journalism, stated, "I don't know whether it's five years or 10 years or 15 years, but at some point, [the] old media will be fully online" (Ibbitson, 2008). This transition has included the rise of news blogging, now a common feature in newsrooms nationwide (Sheffer & Schultz, 2009). Many news organizations are also adopting Twitter. More than just a delivery system, Twitter serves as a communication platform that can link to online content. At its core, it enables digital communication, allowing users to instantly connect with others, limited to 140 characters per message. For longer posts, users can include links to additional content online. Tweets can be published from computers or mobile devices, reaching followers who opt to receive their updates. Through Twitter’s website, users can follow others' message streams and post their own messages (Johnson, 2009). Twitter's popularity has surged beyond expectations. The Wall Street Journal reported that Twitter had 32 million users in June 2009, an increase of about 2 million from the previous year, with predictions of a 50-100% monthly growth rate (as cited in McIntyre, 2009).

Studying the interactions between sports journalists, athletes, and fans is made easier by Twitter. With sports already having a sizable, devoted audience eager to speak with sports journalists, players, and coaches directly, all of the benefits that Twitter offers to media and journalism are amplified for the sport. According to Gregory (2009, 124), Twitter has the power to "change the athlete/fan interaction forever." Gregory (2009) cites Nielsen Media as stating that the majority of Twitter users are between the ages of 35 and 49, which is consistent with the age range of avid sports fans (Gantz & Wenner 1991; Perse, 1992). Throughout the interviews, the most frequently discussed topic was Twitter use, which is not surprising considering the growing body of literature on the subject (Ahmad, 2010; Farhi, 2009; Hermida, 2010; Lasorsa et al., 2011). What wasmore important was that even though just 20 journalists had personal Twitter accounts, 25 talked about using the platform for news monitoring and research. According to earlier studies, journalists primarily use Twitter for breaking and sharing news or sourcing information (Hermida, 2010). This is in contrast to the sport journalists' reported usage of the platform, which was to see

"what's happening." Sport journalists revealed who they followed on Twitter, which was a mixture of athletes, media outlets, other journalists, and sport organizations—people or organizations that offered the most helpful information. This further demonstrated the significance of Twitter for news research and monitoring.

Blogs

Another useful tools for web 2.0 was the blogs, which provide users the ability to text-based chat, create threads, and respond to messages with other fans. The evolution of sports journalism has seen a significant transformation, driven by advancements in technology and shifts in economic funding (Tandoc, 2014). This transformation is evident in the emergence of platforms that provide users the ability to engage in text-based chat, create threads, and respond to messages with other fans. These changes have broadened the scope of sports communication, blurring the lines between traditional journalism and other forms of media.

In 2017, the positioning of sports within the journalistic field was highlighted, showing its slight pull towards economic capital rather than pure journalistic capital. General news holds greater capital than specialized blogs, yet sports journalism is influenced by various fields, including strategic communication. This cross-pull results in team media writers, who might seem part of strategic communications, being perceived as part of the journalistic field due to their socialization, training, and ability to articulate a doxa fitting within journalism (Mirer, 2019).

The rise of unpaid sports enthusiasts/bloggers and team media for sports organizations has paralleled traditional sports journalism, challenging its boundaries (Hutchins & Boyle, 2017). This phenomenon has compelled sports journalists to assert their professional identity and emphasize their adherence to standards of truth and independence, reinforcing their journalistic doxa (Mirer, 2019).

Despite the overlapping work of in-house reporters and unpaid bloggers (McEnnis, 2017), sports journalists strive to distinguish their roles within the journalistic field, often to delineate their work from that of team media. This demarcation underscores their commitment to journalistic ethics and their perceived indispensability within their organizations, even if not within the broader journalistic field. Digital practices that do not fully conform to the journalistic doxa that is central to this field provide the foundation for the development of capital in journalism. Therefore, the insurgent in-house reporter or enthusiast blogger poses a greater threat to field stability than the digital journalist covering sports for a news organization.

This study's theoretical implications include how much neighboring fields can influence and empower rebels in a specialized area of journalism. This insurgency was made possible by the roles that digital sports journalists defined. While watchdog and advocate roles would be more challenging to carry out, roles like monitorial, storyteller, educator, and enrichment roles could all fall under the purview of team media expectations. In this context, the efforts of sports journalists to maintain their distinct identity and adhere to ethical standards are crucial in navigating the evolving landscape where traditional and digital practices intersect.

3.2.3 The new decade and the rise of AI

Web 3.0 is a new idea that has begun to gain traction among business Web developers and others in recent years. Though opinions on the current state of the Web remain hazy and unclear, modern business models emphasize the significance of going beyond Web 2.0 and developing fresh approaches to handle, arrange, and interpret the massive amounts of data generated by users (Funk, 2008; Harris, 2008; Tasner, 2010; Watson, 2009). In the minds of Web developers and business models, Web 3.0 is often linked to the notion of the Semantic Web. The concept was first proposed in 1999 by Tim Berners-Lee, the man behind the World Wide Web. Berners-Lee saw the potential for machines to be able to "talk to one another" and be able to comprehend and interpret semantic data (Berners-Lee in Floridi, 2009: 27). According to Floridi (2009), the Semantic Web as it is being portrayed by Tim Berners-Lee and the members of the W3C (World Wide Web Consortium) is not technically or socially feasible at this time. However, he contends that the Web is evolving in a number of ways that go well beyond the interactive processes of Web 2.0. The development of ever-smaller, mobile applications by companies and computer engineers that utilize crowdsourcing to organize data on the Web and function as databases is one example of this shift (Harris, 2008; Watson, 2009). During the 2010s, the spread of smartphones and tablets, along with the rise of mobile journalism, marked a significant technological leap from personal computers to mobile devices. This transition paved the way for mobile journalism, or mojo, which became one of the fastest-growing areas of research on digital journalism in the 2010s (Westlund, 2013). Much of this research on mobile communication and other technologies is based on uses and gratifications theory, attempting to explain why people replace certain devices with newer ones on the market. Mobile devices have significantly enhanced the capabilities of journalists, enabling them to report more efficiently from the field. These devices are used not only for mobile news platforms but also across the entire cross-media portfolio. With internet connectivity and advanced search functionality, mobile devices offer journalists powerful tools for quick fact-checking, accessing

databases, and contacting informants. Bivens (2008) highlights that journalists can now instantly verify information while interviewing politicians, making news reporting more efficient, especially as the number of journalists in traditional news organizations declines.

As digital journalism continued to evolve, a new wave of artificial intelligence emerged in the decade following 2010, there is a new wave of artificial intelligence. Newsrooms began experimenting with AI to automate repetitive tasks such as interview transcription, news classification and data analysis. Tools such as machine learning algorithms were used to improve the accuracy and efficiency of information gathering and processing. This period marked a notable shift in digital journalism, which has been characterized by continuous transformations since its inception.

Digital journalism has undergone a "metamorphosis" (Vázquez-Herrero et al. 2020), resulting in a novel communication scenario where opportunities and challenges converge around journalism and its practice. Technology has consistently been at the forefront of these changes, dictating the pace and direction of the media ecosystem's evolution. The advent of AI has further propelled this evolution, pushing the media towards a convergent, mobile, ubiquitous, and now intelligent environment. This intelligent transformation, driven by AI, underscores the ongoing dynamic interplay between technology and journalism, shaping the future of news media in profound ways. Since the 2010s, one of the biggest changes has been the automation and robotization of the stages and procedures of journalism. Since algorithms are the cornerstone that enable any system to function, they are currently regarded as the essence of artificial intelligence. It brings the concept of "smart media" to life (Zhang 2023), but the essential components of the high-tech journalism ecosystem that make a difference are data. (López-García and Vizoso 2021), whose are continues to widen as a result of cutting-edge digital tools, resources, and technologies such as drones, 360-degree video, immersive audio (binaural), virtual reality, augmented and extended reality, newsgames, holograms, and 3D printers (Pavlik 2019; Silva 2022; Lima and Barbosa 2022); additionally, machine learning, Natural Language Processing (NLP), and Natural Language Generation (NLG) extend automation to the processes of journalistic content production, reporting, editing, publishing, distribution, and consumption. Algorithmic journalism, which is described as follows, has emerged as a result of the introduction of AI in the media. The (semi)automated process of creating natural language through the following steps: choosing electronic data from public or private databases (input); determining which pre-selected or non-selected data characteristics are relevant; processing and organizing pertinent data sets into a semantic structure (throughput); and publishing the finished text on a platform that can be accessed online or offline with a specific audience (output) (Dörr 2016). The use of AI in journalism is being driven by a number of factors, including

the need to publish more news, cut expenses, free up journalists from tedious and repetitive tasks, and draw in readers (Diakopoulos 2019; Graefe 2016; LeCompte 2015). A new field that affects jobs and functions that once needed human intervention is journalism supported by AI (Bodó 2019).

Now the AI development in sports journalism is spread in several countries and we examine the case in Spain, Portugal, Brazil, U.S.A and Germany. In Spain the AI development and in this context, educational practices in journalism highlight the increasing importance of integrating AI training into curricula, as emphasized by media companies (Beckett, 2019). In Spain, AI is notably absent from journalism degree and master's programs, prompting calls for the inclusion of innovative educational approaches to incorporate AI in communication studies (Gomez-Diago, 2022; Ufarte-Ruiz et al., 2020). While not all journalism processes can be automated, journalists must update their knowledge and skills to utilize AI tools (EBU, 2019) effectively. Researchers from the Tow Center for Digital Journalism also emphasize the need for media companies to invest in specialized training to ensure journalists can use AI ethically and proficiently (Hansen et al., 2017). Despite the growing importance of AI, research into its uneven adoption in newsrooms is limited, and this disparity could present significant challenges, especially for smaller organizations (Wu, 2019; Beckett, 2019). In Portugal, the news agency Agência Lusa briefly utilized software to automatically generate news stories on stock market openings and closings, with journalists validating the content prior to publication (Nunes, 2019). This system functioned similarly to editorial algorithms, producing news with human oversight (Marconi, 2020). According to Luísa Meireles, former director of Agência Lusa, such technologies facilitated routine journalistic tasks (Nunes, 2019). However, the application is currently under further development. Another example is the sports news outlet Zerozero, which employs Prosebot, an algorithm that generates news based on sports competition results using natural language processing (NLP) techniques. Prosebot provides journalists with initial drafts, enabling quicker news production, especially for covering a high volume of results from national and local championships (Pires, 2021). In Brazil, studies suggest that the use of artificial intelligence (AI) by national news organizations remains limited and sporadic (Carreira, 2017; Araújo, 2017; Essenfelder & Sant'Anna, 2022). However, notable examples of AI integration in journalism have been documented (Pase & Pellanda, 2019; Essenfelder et al., 2019; Dalben, 2020, 2022; Santos, 2020; Pérez-Seijo et al., 2023), as confirmed by empirical research conducted for this article. The uneven adoption of generative artificial intelligence (GAI) in Brazilian media outlets can be attributed to global disparities in technological distribution, often tied to capital distribution. Consequently, Brazilian digital journalism occupies a peripheral position in terms of access to technological and financial resources. This marginalization influences the types and scope of AI initiatives implemented in the country. In

Germany and in the recent years, the discourse surrounding AI and its future has gained significant attention in Germany, driven by concerns about the nation's economic trajectory. As Europe's largest economy, Germany's industrial strength lies in traditional sectors like automobiles, machinery, and chemicals. However, it has lagged in adapting to the Digital Revolution, largely dominated by Silicon Valley tech companies. Central to the public debate is the notion that Germany must swiftly modernize its economy and workforce to become a global leader in AI, akin to its historic leadership in industrial sectors. This narrative, promoted primarily by the German government and media, emphasizes the urgency of technological advancement and frames AI as critical to shaping the country's future. These key actors play a significant role in shaping public understanding of AI and its implications for German society.

3.2.4 Spain

Spain has experienced significant achievements in sports in recent years, with successes like winning Euro 2024. Sports are highly popular in the country, but this development started after 1975, as civil society during Franco's era had limited access to sports. After Spain's democratic transition, there were debates about citizen participation in decision-making, with municipalities initially focusing on inclusive and recreational sports aimed at well-being rather than performance.

In the 1990s, Spanish sports policy shifted towards individualism, efficiency, and market values, leading to the rise of fitness-focused programs. The 1992 Barcelona Olympics were a turning point, boosting sports infrastructure and national pride. In recent years, sports have become a political priority due to their economic impact, contributing 2.7% to Spain's GDP, with the central government now overseeing sports at the highest levels.

The government's sports policy focuses on promoting sport as a social activity, integrating it into education, addressing social issues, and supporting elite sports. Spain is now among the top sporting nations globally, attracting significant private sector investment, especially in sponsorship.

The challenge of adjusting to the particularities brought about by the use of AI in newsrooms has fallen to Spanish digital media in the third decade of the twenty-first century. The need for innovation drives its application in the media, but it also through the pursuit of efficient productivity (López-García and Vizoso, 2021)

Journalism is experiencing a profound shift driven by the evolving networked society and advancements in technology. This transformation encourages the adoption of models leveraging 5G, automation, blockchain, and robotics, which are particularly effective for repetitive tasks and complex issues. Innovations in IT, virtual reality, artificial intelligence, and repurposed military technology like drones and bots are bringing new applications to communication and journalism, reshaping how information is produced and consumed.. The phenomenon of journalistic automation and robotization, which has a global impact and transversal character and shapes what has been called the seventh stage of digital journalism, has not escaped the attention of the Spanish media system (Salaverría 2021).

AI-based projects in both legacy and digital native media can be found in Spain. An important example is the case of AnaFut, and the case of sports journalism is very important because is one of the first professional fields to apply AI to the development of news product and the need to be fed with statistical data. Moreover, the nature of the competitions themselves has a direct bearing on how widely used this technology is in sports. These, with their recurring and cyclical nature as well as media coverage, have a robust statistical foundation that supports the management of organized data and enables the creation of educational programs. The usage of automated information-writing bots has increased, which has sparked discussion among professionals in recent years. The argument centers on defining the limits between human and machine functions and the effects of their application in a world of journalism that is more flexible, varied, and of a higher caliber. In the late 2017 was born this tool and it was introduced by El Confidencial Lab to cover lower division football competitions, a niche not typically addressed by the Sports section's journalists. By clearly delineating the bot's functions from those of the human journalists, the managers aimed to ensure a harmonious coexistence from the outset. This case highlights the potential for technology to complement rather than replace human expertise, contributing to a more diverse and high-quality journalistic landscape. Victor García, the chief editor of the Sports department, has brought attention to this, calling AnaFut's endeavor "throw-away work" because the match reports are typically only available online for a few hours at a time. These are extremely brief texts meant to satisfy users' curiosity about the most recent news, as near as possible to the updates given by the live match coverage that a lot of them watch on social media and television. Considering that this is essentially about the bot "doing other work so that journalists can be journalists, and are able to produce other stories that generate organic traffic" to the web, El Confidencial Lab director Alejandro Laso believes that "the important thing is for editors to start understanding this technology." "To save editors time when doing their job, to cover niches and areas that were not possible due to a lack of professionals, to reach new audiences, and to make Sports people more capable of acting and disseminating" was the

explicit goal of this Sports bot from the outset. Furthermore, robots perform the grunt labor that journalists typically have to perform; however, this is the kind of labor that keeps them from shining, according to Laso.

While Segunda Division B football games have served as a testbed for the bot, it remains unclear if this technology can be extended to other sports. Alejandro Laso thinks that "it can be applied to anything that has structured data as well as frequent, routine competitions" in this sense.

Thus, while this technology has many benefits, it also has obvious limitations when it comes to journalism. For this reason, the discussion also centers on how news writers' editorial oversight or intervention can improve artificial intelligence, enabling the creation of better texts through better bot preparation, or "training."

The foundation of AnaFut is a programming meta-language created in the Lab that allows any editor to experiment with every variable. Thus, the automated reports will have a wider range of lexical and statistical readings based on the volume of data entered through programming. El Confidencial is already attempting to increase the usage of the bot through the customization of the news offering, real-time results monitoring, and the introduction of dynamic graphics through data journalism.

El Confidencial already uses this technology outside of the automated Second Division B reports, for example, in stock market alerts, judicial news (monitoring of pardons, appointments to public positions - companies, politics), politics (monitoring of electoral data), and other services like weather reports and lottery results. This demonstrates a broader trend where leading journalistic organizations are leveraging Artificial Intelligence to enhance their operations. Our experts note that we are in a phase where AI tools and techniques are being implemented predominantly in technologically advanced countries and within innovative journalistic organizations. These organizations are pioneering new approaches to journalism, supported by technology and knowledge, as a means of progress and growth. As the number of AI-driven experiences increases, along with their quality and integration into media structures, content automation solutions are evolving. These solutions are no longer just executing programmed commands but are now capable of making complex decisions based on their learning abilities. This indicates a significant shift towards more intelligent and autonomous AI applications in journalism, promising a future where AI not only aids but also enriches journalistic practices.

One area where AI has had a longstanding impact is sports journalism. For over a decade, artificial intelligence tools and solutions have been used to enhance sports coverage, particularly through automatic text writing or natural language generation (NLG). This technology allows for the creation of specific news narratives from sets of data, making it a natural fit for sports journalism, which relies heavily on statistical

information. The repetitive and cyclical nature of sports events, such as races and tournaments, further supports the integration of AI into sports journalism, enabling efficient data management and streamlined information routines(el periodismo deportivo terreno de vanguardia para la aplicacion de la inteligencia artificial). In addition to generating match reports or player statistics, AI can now analyze vast amounts of historical data to provide predictive analytics and insights. For instance, advanced algorithms can predict potential match outcomes based on player performance, injury records, and team dynamics. This shift towards data-driven content has not only improved accuracy but also allows journalists to explore in-depth angles, such as player health forecasts or tactical trends over time.

AI-powered tools like chatbots and virtual assistants are being increasingly employed by media outlets to interact with audiences in real-time, providing instant match updates, summaries, or even answering fan queries. Moreover, machine learning and computer vision technologies are being leveraged to automate video highlights, creating compelling video summaries within minutes after a game ends. This enables media houses to publish engaging content faster than ever before, catering to the fast-paced, 24/7 demand for sports news.

Looking forward, AI is poised to further revolutionize sports journalism by personalizing content according to reader preferences. Through the use of machine learning algorithms, news organizations can tailor articles, summaries, and even video highlights based on an individual's interests—whether they want to focus on a specific team, player, or type of game. This shift toward hyper-personalization could enhance fan engagement, turning passive readers into more active and loyal followers.

With AI-driven tools becoming more sophisticated, sports journalism has emerged as a vanguard field for the application of artificial intelligence, setting a benchmark for how technology can transform content creation in media industries

3.2.5 Portugal

Portugal has a story of sport tradition and we can see the demographic influences and higher population and density, and larger territorial areas impact the number of athletes; moreover, the presence of more sport director affects the number of athletes. However, the number of sports clubs, coaches, and referees doesn't show a significant positive impact. This may be due to coaches leaving the sport after training or the ineffective use of assistant coaches. There are also financial factors

like for example increased spending on culture and sports contributes positively to the number of athletes. In addition capital expenditures and sports facilities sometimes have a negative impact, possibly due to ineffective investments or ethical issues requiring further supervision.

Another important point of view of AI development are Portugal and Brazil. Starting from Portugal and the quantity of mainstream newspapers (DN-Diário de Notícias, Correio da Manhã, JN-Jornal de Notícias, and Público) and daily sports newspapers (A Bola, Record, and O Jogo) is nearly equal. A further example comes from television, where there are four generalist channels (SIC, RTP, TVI, and CMTV) and three sports channels (SportTV, Eleven, and Canal11). It's also worth noting that sports, particularly soccer, are becoming more and more prevalent on the generalist channels, and that the biggest sports teams in the nation have their own TV channels. Given this abundance of media outlets and the vast amount of information they must manage, newsrooms are facing challenges such as an excess of information and a lack of human resources. To address these issues and maintain a steady flow of information despite cuts in staffing, journalism has turned to solutions like the automatic production of texts, also known as "algorithmic journalism." According to a report of LSE also in Portugal(Beckett 2019). AI can assist them in producing more and better information, which will support the media's economic sustainability and help them win back the public's trust. Despite this, the respondents acknowledge that the process is hampered by the significant investments required for the implementation of AI, both in human resources and emerging technologies. The first use of AI in Portuguese journalism occured in 2015 in a native online and the uses of AI and The applications of artificial intelligence and the rationale behind their use are as follows: machine translation is linked to automatic text production, time savings, and research (news trends and archives); automatic text correction is linked to journalists working more productively; and the creation of graphics is linked to the distribution of more pertinent content. When asked why artificial intelligence was adopted, the most common responses were that it would save time and make journalists' jobs easier.

When asked about the reasons for adopting Artificial Intelligence (AI), the most mentioned were saving time and making journalists’ work more efficient, confirming trends from previous studies. AI is used to increase productivity, with specific applications like machine translation for timesaving and research, automatic spelling correction for efficiency, and graphics production for delivering relevant content. Audience involvement in the news process is associated with nearly all AI uses. On a scale from 1 to 10, the importance of AI in supporting journalism averaged 7.2. The age group 31-40 most recognizes AI's importance, while the 51-60 group is the most skeptical, though the difference is not significant. In summary, in Portugal, the

primary concern regarding the implementation of AI in media centers on the lack of adherence to ethical standards. The application of AI without rigorous ethical and quality oversight has the potential to introduce numerous negative consequences, particularly in terms of journalistic integrity and content accuracy. Beyond these ethical concerns, the integration of AI into media presents a range of editorial challenges. One significant issue is the increasing reliance on technology companies that provide automation tools for news production (Pérez-Seijo; Barbosa; Vicente, 2023). This growing dependence on external entities raises questions about control, transparency, and the independence of the media. Moreover, there is an urgent need for newsroom professionals to acquire expertise in AI technologies (Deuze; Beckett, 2022). Journalists and editors must not only become proficient in the use of these tools but also develop the technical skills necessary to critically engage with and navigate today's complex information ecosystem. This includes understanding how AI can shape content production, distribution, and audience engagement, while also maintaining journalistic values and ethical practices. The confluence of these factors suggests that the adoption of AI in media, if not managed properly, could have profound implications for both the profession and the broader public discourse.

3.2.6. Brazil

Another Portuguese-speaking nation is Brazil. The sport development was started in 1960s with "sport for all" and the aim was to promote social rights and active sports habits, using print media extensively for marketing and public education. In Brazil, this initiative utilized various media channels to instill a 'sports mentality' among the populace. In the 1970s Brazil was undergoing major political, economic, and social changes under the Military Regime. The initial economic boom known as the 'economic miracle' (1968-1974) increased industrialization and the purchasing power of the middle and upper classes. However, the global oil crisis led to economic decline and political discontent, prompting the government to invest in participatory initiatives, including mass sports promotion.

Brazil aimed for greater autonomy in its foreign policy, establishing agreements with several countries, notably West Germany, leading to the 'Brazil/West Germany Technical Cooperation Agreement' which included a project to promote sports in Brazil. The government used sports to foster nationalism and unity, exemplified by the national football team's success in the 1970 World Cup.

Brazilian sports policy remained state-controlled, focusing on elite sports development. During General Emílio Garrastazu Médici's government (1969-1974), a significant study on physical education and sports was conducted to create a

National Policy for Physical Education and Sports. This policy aimed to improve physical fitness, elevate sports levels, democratize sports participation, and promote healthy habits and sports performance.

Key actions included the National Campaign for Sports Education, the 'Mexa-se' Campaign, and the Sport for All Campaign, which utilized media to encourage physical activity. The Brazilian Literacy Movement (Mobral) was tasked with organizing the Sport for All Campaign, highlighting the government's effort to integrate sports into broader social policies. In addiction there is a “Decalogue of sport for all”, that a set for 10 axes considered the code of ethics like: leisure, health, community development, social integration, civism, humanization of cities, valorization ofnature, adherence to sports practice, adhesion to organized sports, and valorization ofcommunity service (Documento Básico da Campanha 1977). Moreover it was necessary a changing of mentality in Brazilian sport, which was helped also with universities. However, as Brazil entered the 1980s, its sports policy began to evolve. The focus shifted from state-centric approaches to a more inclusive and democratic model, integrating both official and private community initiatives. New communication channels like the Comunidade Esportiva newspaper and the 'Rede Nacional do Esporte para Todos' network played a pivotal role in consolidating the movement, introducing new concepts into sports, and promoting its democratization. This period laid the groundwork for the inclusion of sport as a social right in Brazil's 1988 Federal Constitution.

Shifting from the historical development of sports in Brazil to another domain of technological advancement, journalism witnessed the emergence of AI in the 1990s. Although AI's integration into journalism began modestly, it wasn't until 2014 that it gained significant traction with the advent of online journals utilizing AI. This advancement dramatically enhanced the capabilities of not just print media, but also TV, radio, and the web. However, despite these technological strides, a major concern persists among Brazilian journalists: the fear of being replaced by AI, potentially losing their jobs to automation. This tension between innovation and job security echoes the broader societal impacts of technological progress, much like the evolution of Brazil's sports policies over the decades. This interplay between technological progress and societal impact is not confined to sports alone. Journalism, too, faces its own set of challenges and opportunities in this era of rapid innovation. As technological advancements reshape various industries, including journalism, the focus shifts to how these innovations redefine journalistic practices and the broader media landscape. Lopezosa et al. (2023) define "innovation" in journalism as the integration of new technologies, methodologies, and strategies, which includes novel formats such as immersive journalism, 360° video reporting, virtual reality, and AI-driven journalism. This innovation encompasses not only new technologies and

practices but also addresses challenges within the creative industry caused by societal, technological, economic, and political shifts (Bossio & Nelson, 2021). Furthermore, Franciscato (2014) emphasizes innovation as technological adoption that transforms media business models, workflows, and journalistic roles, while also influencing the development of new news formats and consumption modes. Barbosa (2014) highlights the ongoing nature of this digital innovation.

However, the integration of artificial intelligence (AI) into journalism remains uneven, particularly in the Global South, where financial and technological limitations hinder widespread adoption (Lima Santos et al., 2022; Simon, 2022). In Brazil, the use of AI in news organizations remains sporadic despite some notable instances (Carreira, 2017; Araújo, 2017; Essenfelder & Sant'Anna, 2022; Pase & Pellanda, 2019; Santos, 2020; Pérez-Seijo et al., 2023). This disparity stems from transnational inequalities in the distribution of technology and capital, positioning Brazilian digital journalism at a marginal level in terms of technological development and access to resources. Consequently, this affects the types of AI initiatives implemented in Brazilian journalism.

To sum up, in Brazil we can see the fear of losing a job and this kind is largely spread in the world and it is not involved the role of journalism. The role of journalism is watchdog but the biggest fear is the the great fear is that it may lose this monitoring task, in light of these concerns The shift from traditional media production (print, TV, radio) to the continuous output required by online platforms has led to a dramatic increase in content production, with a focus on generating more traffic, views, and clicks. This "streaming" model demands a constant flow of stories, often at the expense of the quality and depth of individual pieces. Media theorist Dean Starkman refers to this phenomenon as the “hamsterization of journalism,” where journalists are pressured to produce more stories in less time, adhering to the principle of “doing more with less.”

In this context, news organizations prioritize quantity over quality. While a well-researched article might attract thousands of readers, an alternative strategy of publishing numerous lower-quality pieces can collectively drive traffic, even if each story garners fewer readers. Consequently, speed and volume become the primary concerns for editors, who may prefer fast-working journalists or algorithms that can generate content rapidly over skilled writers. For example, the Associated Press adopted the Wordsmith service not because it produces superior writing, but because it does so more quickly and efficiently, highlighting the economic rather than qualitative motivations behind this shift.

Ultimately, if human journalists retain a place in the media industry, it will be due to social and political pressures to maintain employment rather than economic efficiency. In this way, the preservation of jobs in media reflects a broader trend where socialism, in terms of protecting labor, prevails over capitalism's drive for efficiency. The competition with algorithms is secondary to this larger social imperative(Miranchenko 2018).

3.2.7. USA

Sport in the United States reflects the country's unique historical trajectory of capitalism and constitutionalism. Under the influence of neo-liberal ideology and austerity-driven economic policies, which gained prominence during the Reagan era, sport at various levels—high school, college, university, and professional/Olympic—has increasingly aligned with market-driven and business principles, generating significant revenues and attracting considerable societal attention. However, following the 2007/2008 global financial crisis, mass participation in sport, particularly at non-elite levels, has been negatively impacted due to the economic strain on the middle class, resulting in decreased discretionary spending on sport. Despite these challenges, uncertainty persists regarding the future of sport for the broader U.S. population(Smolianov et al,2014)

The level of government regulation in the U.S. differs markedly from many other countries. The U.S. Constitution emphasizes individual freedoms and limits on government intervention, fostering a unique form of liberalism. However, despite these constitutional ideals, government involvement in daily life is extensive, with elected officials increasingly entangled in legislation that affects various sectors. While state intervention is inevitable in any political system, U.S. sports remain relatively insulated from direct governmental regulation.

Also, the American press started to use AI in the mid of 2010s, we can see like that The Washington Post is among the media outlets that have demonstrated the strongest dedication to innovation in covering the most recent international sporting events (World Cups and Olympic games) through story rewrites and, particularly, technological advancements.

Thanks to its innovative use of technology to improve journalism, The Washington Post has emerged as one of the most innovative news organizations and a model for the media sector as a whole. The Post has invested in creating its own technology platform, known as ARC Publishing, and has encouraged a culture of innovation and experimentation to transform its newsroom, moving from being a local to a national newspaper and reorienting its business to develop a digital subscription-based model since this newspaper was acquired by Jeff Bezos, the founder of Amazon, in 2013.

The Washington Post is one of the first newspaper publishing companies to use bots for automated journalistic writing in this cutting-edge technological wager. The Washington Post created its artificial intelligence program, Heliograf, to create headlines and summaries automatically, imitating the actions of the AP and other news organizations. In order to achieve this, the competition's liveblogging, Facebook Messenger, and Twitter were the selected platforms.

The 2016 Olympics could not miss the debut of Artificial Intelligence for the coverage of certain events, in fact, in the second half of the 2010s there was an increase in AI-related activities. And here we see the debut

The 2016 Summer Olympics in Rio de Janeiro marked the first use of this technology in a sporting event that evolved into a test run. In 2018, the so-called Post Oly Bot was utilized once more for Olympic coverage in PyeongChang. The Olympic experience inspired The Washington Post to employ this technology for coverage of the 2016 US presidential elections as well as college football league games.

The main goals of this study are to analyze the coverage of the 2016 and 2018 Olympics by The Washington Post's automated storytelling bot, identify the opportunities that automation offers to improve news coverage, and, finally, determine the degree of complementarity between algorithms and bots and human work in the news-making process. This is due to the uniqueness and relevance of the medium, as well as the pioneering nature of this technology and its application to the coverage of large sporting events and its subsequent use in other news areas.

There are two stages to the suggested study design. In the first, the coverage ofthe 2016 and 2018 Olympics by the automated storytelling bot is quantified and differentiated, with consideration given to the identification ofthe tweet typology, frequency of publication as well as how they use information, graphics, and audiovisual materials. The second phase is qualitative and aims to ascertain whether the automated publications required more or less human editorial intervention, as well as how much the bot integrated different elements or simply added information to the coverage made by special correspondents who published on the newspaper's website during both Olympic events.

The Olybot covered several events in Rio 2016 and the behaviour was structured in this way :

- Day’s schedules;
- Medal tallies;
- Reminder of start competition with the link of the coverage;
- Results

Daily cycles that follow a predetermined editorial schedule. This type of human involvement goes beyond the database that the bot uses to extract the data needed to write the tweets. Regarding the tweet typology, it should be mentioned that most of them followed the "one tweet, one sentence" approach and did not include charts in the results.
Additionally, each result is written in accordance with a pattern that is virtually always the same:
- The athlete's name plus the hashtag containing the nation's three-letter abbreviation
- A statement indicating the medals won, either in gold, silver, or bronze for third-place performances.

In addition to the results, the bot sent out tweets that included schedule and competition reminders as well as updates to the medal table—typically one update per day without charts. In reference to the latter, only a portion of the events—two or three per day—were recalled. Corrections to inaccurate data were also released. Once more, both before and after, human intervention in the form of editorial character is used to enhance the algorithm's performance.

Reminders about upcoming competitions are another kind of tweet that this bot produces. These reminders typically consist of a single sentence announcing the start of a final or an event with medals up for grabs (the editorial standard of news hierarchy and selection is also applied here). The automated messages are intended to supplement and reinforce the medium's overall coverage. They are typically published fifteen minutes prior to the start time and contain a link to the live coverage on the website.
Olybot was also used for the winter games in 2018, The Washington Post's 2018 Olympics coverage via its new bot included a few enhancements. These enhancements were a result of the graphics department's transversal work and the weight this medium places on design to improve the visual and informative quality of its coverage of major athletic events. In this instance, to depict the

competitions across all platforms, the pictograms of each sports modality and the athlete logos also surfaced in the GIF format.

In the PyeongChang Olympics, whose coverage by the Post Oly Bot included up to 18 tweets in a single day (February 24), the daily behaviour of this bot was structured in the following way:

1. Reminder of medal tallies
2. Reminder of the start of a competition
3. Podium results with link to the website
4. New result/announcement
5. Medal leaders at the end of the day

The results stand out in terms of the tweet typology, outperforming medal tallies and reminders, whose postings were made between 15 and 5 minutes before the event. In contrast to Rio 2016, the results were typically expanded to two sentences rather than one and addressed the following pattern: - Winner of gold medal, name of event, country, time, or points; - And in second sentence: name-country-silver medal and name-country-bronze medal.

In this instance, the editorial intervention was evident not only in the tweaks and schedule adjustments made to a few tweets, but also in the utilization of this account to advertise the media's coverage via the tweets' included link. In this sense, the data on the website also helped to put the results that were tweeted into context. Similarly, the bot only retweeted content from The Washington Post's sports section, and this particular retweet was created by a human.

The first important factor for content selection and hierarchy was the editorial criterion of the people in charge of the sports section, which included coverage of the results and reminders about the most notable events. This editorial criterion was also evident in the coverage of American athletes' accomplishments, which received less attention than in Rio 2016, when the publications produced by the bot featured a wider range of themes and characters.

In conclusion, the hybrid approach allowed journalists to focus on more nuanced and creative reporting, enhancing the overall quality of sports coverage. The study concludes that automated journalism and human journalists can coexist harmoniously, with bots handling rapid, data-centric updates while reporters focus on storytelling, research, and analysis. This division of labor can improve the profitability, efficiency, and quality of sports journalism.

3.2.8.Germany

In Germany sports journalism is viral, the sport is popular from football to track field, the central European country developed a strong sports tradition and also makes no distinction in sports journalism, which is supported by artificial intelligence. We can see the German case as a way to increase automated journalism. A notable example comes from Fussball.de, the largest source of automated reports on German amateur football is the German Football Association. Nonetheless, the German Football Association asserts that their only goal is to support their amateur leagues, and as such, they have no desire to become journalists and do not question the authority of journalists. There are also other online journals that use automated text, like FussiFreunde, which covers amateur football, and OVB24 Beinschuss, which covers amateur football in Bavaria. Focusing on major online journal, in this case Fussball.de; programming the software to "talk sports" required "a lot of manual work," such as adding "metaphors and idioms." A specialized project team of "sports journalists, machine learning specialists, software specialists, and linguists" was hired by fussball.de to handle this programming. This team initially set up the software and is still working on changes. Therefore, implicit trespassers are not only accepted but also invited to participate in the news process to establish the guidelines and meaning of automated sports coverage (Graefe, 2016). But in this instance, the German Football Association collaborates with other intruders to produce automated news, making it implicitly an intruder as well.

Team managers can edit the automatically generated articles on fussball.de as well. As a result, three reports about the same game could be written by "the software, the home team, and the away team." Eighty percent of teams don't use this option, which validates fussball.de's assertion that the quality of their articles is "excellent."

Fussball.de and Reports on amateur matches are provided by the German Football Association, for example, despite their assertion that they "don't call [their coverage] journalism, but rather text generation according to data," because the software is unable to "give context" or "judge" matches; therefore, "the machine won't replace journalists." For amateur clubs, who can use the articles on their websites and social media feeds, they view automated content as a service.

Notwithstanding these assertions, the comprehensive coverage they provide reveals a novel aspect of the automated news processing system. Even if it was not their original intention, implicit intruders have the potential to become explicit, as demonstrated by the aforementioned examples.

The German Football Association works in close collaboration with a diverse team of linguists, software developers, and machine-learning experts, significantly expanding the involvement of external specialists in the process. This collaboration illustrates how multiple players, including hackers and other experts, can work together to create automated news articles, leveraging advanced technology to streamline content generation. As a result, providers of software and data, who facilitate this process, could potentially also position themselves as intruders, raising questions about the ethical dimensions of such partnerships.

Furthermore, fussball.de operates not just as a website but also as a comprehensive app, offering extensive management of information relevant to amateur football players. Through the use of advanced AI, the platform excels in covering lower-tier leagues that often receive less attention in mainstream media. By providing real-time, detailed coverage of amateur football, fussball.de serves as a valuable resource for both players and fans. Its capabilities also prove advantageous for sports journalism, as the platform's AI-driven approach allows journalists to efficiently cover a wide range of leagues, thus expanding the scope of sports reporting and fostering a deeper understanding of grassroots football. The integration of AI enables greater accessibility to local and amateur sports, making it an essential tool for both the public and professional media outlets.

Some consideration

Analyzing various case studies of artificial intelligence applied to sports journalism, it becomes clear that several countries have embraced this technology as part of their research and development strategies. Sports journalism has emerged as a particularly suitable field for the application of AI

due to the highly structured nature of its databases, which allow for efficient information extraction, processing, and automation (Graefe; Bohlken, 2020). These databases are rich in statistics, player performance data, and historical records that can be seamlessly integrated into AI systems to generate accurate and timely reports. Furthermore, the fast-paced nature of sports reporting, where speed often takes precedence over in-depth analysis, makes it a natural fit for automation and algorithmic support (Kim; Kim, 2017).

One notable example of AI in sports journalism is the use of Natural Language Generation (NLG) tools to automate match reports and summaries. In certain sports leagues, AI is used to instantly create articles summarising key moments, statistics, and results, often faster than human journalists. This practice, widely adopted in countries like the United States and parts of Europe, shows AI's potential to enhance efficiency without compromising on the quality of information.

Additionally, AI-driven chatbots have begun to play a role in sports journalism (Lokot; Diakopoulos, 2016; Jones; Jones, 2019; Veglis; Maniou, 2019). These bots engage with readers through platforms like websites and social media, answering real-time queries about match updates, player statistics, and other sports-related information. Although the presence of chatbots in mainstream media has been somewhat limited to date, their ability to interact with audiences and provide instant responses to routine inquiries hints at the transformative potential of AI in audience engagement.

In conclusion, artificial intelligence can complement, rather than replace, the work of human journalists. While critics have expressed concern about the potential for AI to supplant human creativity and judgment, these fears may be unfounded in the immediate term. AI's greatest strength lies in automating repetitive tasks, freeing journalists to focus on more complex, interpretive, and investigative reporting.

However, AI will inevitably introduce new forms of journalism, revolutionising how stories are crafted, distributed, and consumed, even if this vision faces opposition. As AI continues to evolve, it will enable innovations that promote cost and process optimisation, improving workflow efficiencies in newsrooms. Specifically, in sports journalism, AI can generate dynamic, data-driven narratives that push the boundaries of traditional reporting formats. As a result, journalists, media organisations, and audiences must progressively adapt to a hybrid communicative landscape where human expertise and AI work symbiotically to enhance the overall journalistic experience.

Chapter 4

The pro and cons of AI

4.1 The use of AI in Journalism

The effectiveness of artificial intelligence has been demonstrated through numerous experimental studies, particularly in the fields of science and technology like ChatGPT, NLP, healthcare. Estimating the financial impact of creating and using an AI application is difficult without a thorough review of specific project details, but

many experts recognize that the expenses related to AI system development have decreased in recent years(de Lima & Santos 2021). However, the need for specialized expertise in AI remains a challenge, especially when competing with companies in Silicon Valley. Leading technology firms leverage their dominance in the AI industry by strategically acquiring startups pioneering AI solutions, aiming to strengthen their influence and reduce competition (Linden 2017). This trend extends beyond the tech industry, as news organizations are adopting AI technology in their editorial departments. Over the past few years, there has been a growing trend of automating news articles (Linden 2017). While AI's influence continues to reshape industries like technology and journalism, the integration of AI in media, specifically in automating news content, has seen remarkable advancements in recent years. This transformation illustrates the expanding scope of AI applications and how its adoption is not confined to high-tech firms. Automated news in the United States began to emerge around 2010, largely due to the efforts of companies like Narrative Science and Automated Insights, which have become dominant players in the market. Their software can generate large numbers of articles for various news organizations, including well-respected outlets such as the Associated Press and Forbes. The effectiveness of their automated news systems relies heavily on well-organized data sets. The types of articles that can be created depend on the availability of data. For example, Automated Insights uses financial data from Zacks Investment Research in Chicago to produce earnings reports for the Associated Press, while Narrative Science uses data collected by team leaders via the mobile application Game Changer for sports-related news. In contrast, automated news in Europe is still largely in the experimental stages. Companies like Retresco and Aexea in Germany, Syllabs in France, and Arria in the United Kingdom are currently offering limited services to specific media clients, focusing mainly on stock market updates, weather forecasts, and football news (Dorr 2015). In France, the interest in automated journalism has grown after Syllabs produced 34,000 articles for Le Monde during the departmental election night in March 2015. This successful initiative has led French newsrooms to see the potential of large-scale article production, with a data journalist from Syllabs noting that "it is not difficult to get newsrooms to see the possibility with personalized content."

This growing trend of AI integration in journalism demonstrates how newsrooms worldwide are beginning to explore the potential of automated content creation. As some countries embrace large-scale AI-driven reporting projects, others are still in the experimental phases. Nevertheless, the momentum toward automation is undeniable, signaling a broader transformation in how news is produced. The practice and legal frameworks of journalism have been significantly impacted by technological advancements. Concerns about the future of the profession have arisen due to the increasing automation of various journalistic tasks. Gartner (2017) states that the year

2020 was a crucial moment with the introduction of AI, drawing comparisons to job losses during the industrial revolution. While AI is predicted to remove 1.8 million jobs, it is also expected to create 2.3 million new ones, indicating a shift in professional requirements. These new roles demand greater technical knowledge, leading to a call for journalists to enhance their proficiency in technological tools to manage evolving editorial processes. Opinions on AI's impact on journalism vary, with some expressing optimism about the opportunities it presents, while others, like Amy Webb from the Future Today Institute, voice concerns about the control of information falling into the hands of a few tech giants. These concerns arise from the rapid automation of newsrooms and the lack of technical expertise among many journalists. As AI reshapes news consumption and production, the focus remains on maintaining a human-centered approach, with AI seen as a tool rather than a replacement. AI's advancing capabilities, particularly in language, translation, and text synthesis, are expected to blur the distinction between human and machine-generated content. Tools such as chatbots, designed to engage readers in discussions on current events, demonstrate how AI could complement rather than replace journalistic work. Nevertheless, there are worries about AI's potential to independently generate articles and videos, raising questions about the future of human-driven journalism. Experts like Olivier Ezratty emphasize the potential for AI to disrupt journalism if the industry does not adopt and invest in it. According to Ezratty, innovation thrives in large, solvent markets, and the media must prioritize AI to remain competitive with other industries like transportation, banking, and healthcare(Seyfodin & Tabrizi,2024). This urgency is reflected in modern journalism training programs, which increasingly integrate AI-related skills to prepare journalists for a hybrid future where they work alongside algorithms. In conclusion, while AI offers speed and efficiency, the distinct qualities of human journalists—such as creativity, critical thinking, and investigative depth—will continue to be crucial. A collaborative approach between journalists and AI, where technology serves as an aid rather than a threat, is seen as the way forward in ensuring the profession's sustainability in an era of rapid technological change.

4.2 The advantage of AI in journalism

A significant technological advancement that has profoundly impacted the field of journalism in recent years is artificial intelligence (AI) (Whitakker, 2019). Notably, with the rise of data journalism and the advent of big data tools (Sandoval-Martín & La Rosa-Barrolleta, 2023), news organizations have started to innovate by integrating AI applications into various stages of news production. This includes the automated generation of news articles (Rojas-Torrijos & Toural, 2019), for example the spanish

case of “El Confidencial”, and the customization of content informed by generated metrics (Chan-Olmsted, 2019; Kotenidis & Veglis, 2021), in fact ,scholars have expressed worries about the impact of automated news distribution and the tailoring of content. The rise of customized news has the potential to divide public opinion (Graefe 2016). This division can be linked to the creation of "news echo chambers," which are online spaces that mainly showcase perspectives in line with the preferences of the news consumer while excluding opposing viewpoints (Garrett 2009). The rise of customized news, while potentially dividing public opinion, also brings into focus the broader implications of AI on the journalism industry as a whole. As news consumption patterns shift and the boundaries between human and machine-driven content blur, it becomes essential to evaluate both the risks and advantages that AI technology introduces.

AI offers a range of significant benefits for modern journalism, revolutionizing the way news is collected, processed and disseminated. Here are some of the most important potential benefits of AI in journalism like:

Automation of Repetitive Tasks

Artificial intelligence has the potential to streamline monotonous and repetitive activities, including preliminary data gathering and fundamental reporting. This capability enables journalists to conserve time and concentrate on the more innovative and analytical dimensions of their profession. As AI continues to evolve, its influence on journalism extends beyond mere efficiency, prompting deeper discussions on its role in reshaping the profession. Experts in technology and journalism, such as Burrell and Adepoju(2023), have contemplated the implications of the release of GPT technologies,is Generative Pre-trained Transformer, reiterating both the opportunities and challenges associated with these models in alignment with the perspectives of our respondents. It is essential to acknowledge that these technologies will be experimented with and are likely to be integrated into content creation processes, serving both as a systematic means of automating story production and as a supportive tool for individual journalists in crafting narratives.

Enhancing current process automation platforms with AI can amplify the benefits of regular and on-demand information gathering for trustworthiness. Journalists' and editors' information-seeking behaviors can be analyzed using process mining to enhance and automate information gathering and analysis routines. For instance, if a journalist frequently searches a specific online community forum to validate and enhance football match reviews, event logs can be utilized to extract a process description from this pattern.

Fact-Checking

Facilitating fact-checking is one of the uses of AI, achieved through cross-referencing information with reliable sources and identifying inconsistencies or incorrect information. This, in turn, improves the accuracy of journalistic reporting.

Given the increasing prevalence of misinformation, the role of AI in enhancing this process cannot be overstated. It is crucial for trustworthy news to be based on factually accurate information. Assessing the truthfulness of claims is the task known as fact checking or fact verification(Thorne & Vlachos,2018). AI not only supports this process but also strengthens the foundations of journalism, where fact-checking plays a critical role. By assisting journalists in verifying information more efficiently, AI enhances their ability to maintain accuracy and uphold the standards of credible reporting. Fact-checking is often described as a 'discipline of verification,' distinguishing journalism from other forms of media like entertainment or propaganda (Shapiro et al., 2013). In fact, Shapiro (2010) highlights that verification, or "examination",examination involves assessing the verifiability and coherence of claimed facts should follow "discovery", in journalism, this corresponds to the activities and techniques used to identify a reporting idea or focus, akin to what other disciplines describe as formulating a research question. It also involves developing a reporting strategy, analogous to research methodology, and gathering relevant information, comparable to data collection separating the identification of facts from the testing of their accuracy. However, the concept of verification in daily journalistic practice is less structured. Although fact-checking in prominent, well-resourced magazines involves a thorough and systematic process that addresses definitional questions comprehensively. All claims presented as facts undergo verification, adhering to a structured set of methodological standards and established routines. This ensures the accuracy and credibility of published content.Verification are frequently regarded as synonymous, but some scholars have questioned the feasibility of verification as a practical goal in routine news reporting. Rosner (2009) argues that daily journalism would be "paralyzed" if journalists were required to verify all information before publication. In this view, follow-up investigative reporting serves as an alternative form of verification. Nevertheless, Rosner emphasizes that when the stakes are higher, more thorough verification is essential prior to publication.

recent research has aimed to delineate them as two separate yet interrelated processes (Silverman, 2016). According to Silverman, verification is fundamental to journalism and is progressively being adopted and utilized by various other professions. While Fact-checking is a specific application of verification in the world of journalism. The

establishment of fact-checking as a specialized editorial role emerged in response to the growing emphasis on accuracy within journalism, particularly following the Progressive Era the first dedicated fact-checking departments were introduced by *Time* magazine and *The New Yorker* in the 1920s, a practice that gradually extended to other national magazines as they sought to distinguish themselves from newspapers by offering reliable and comprehensive reporting (Angeletti & Oliva, 2010; Silverman, 2007; Sivek & Bloyd-Peshkin, 2018). By the late 20th century, fact-checking had become a standard procedure in the North American magazine industry, with estimates suggesting that major newsweeklies in the 1980s employed one fact-checker for every three or four correspondents (Shapiro, 1990; Sivek & Bloyd-Peshkin, 2018). As journalism enters the digital age, the role of artificial intelligence (AI) in content verification becomes increasingly relevant. The integration of AI in the media provides significant opportunities to counter disinformation and fake news, but it also raises ethical concerns regarding the selection, prioritization, accuracy, and dissemination of content (Ali & Hassoun, 2019). In this context, an initial analysis of international AI-driven verification institutions and media laboratories is presented. AI can be utilized for content verification, automation, personalization, and the distribution of both general and specialized content, including political information, through bots and chatbots (Sánchez & Sánchez, 2020). It enables capabilities such as character recognition, text processing, deep learning for detecting fake images and videos and enhancing search engine functions. A notable example is Duke Reporters' Lab, a journalism research center at Duke University’s Sanford School of Public Policy, established in 2014 with funding from the Knight Foundation, Facebook's Craig Newmark Journalism Project, and the Google News Initiative. The lab has identified 461 global fact-checking initiatives, 349 of which are currently active, and collaborates with major fact-checking organizations like The Washington Post, PolitiFact, and FactCheck.org. The Duke Lab works with Researchers and journalists engaged in combating misinformation now have access to a significant resource: a new, comprehensive fact-checking database. Fact-Check Insights, developed by the Duke Reporters' Lab with support from the Google News Initiative, consolidates structured data from over 180,000 political claims and online hoaxes. These claims, made by political figures and social media accounts, have been analyzed and rated by independent fact-checkers. This extensive collection aims to enhance efforts to counter misinformation by providing a valuable tool for fact-checking.

Personalization of Information

Reader preferences can be analyzed by AI algorithms, which then deliver personalized content to improve user experience by providing relevant information.

This trend of catering to individual interests is not entirely new. Personalization and customization in journalism have a longstanding tradition. Print media and news broadcasters have historically tailored localized editions in formats specific to their mediums, such as printed newspapers or television programs. Prominent providers, like The New York Times, have adapted by offering location-based printed editions and, more recently, personalized web homepages(Adar et al ,2017). News aggregators, such as Google News, further allow readers to customize their news feeds. This personalization benefits content producers by enabling them to utilize archival material and develop local or hyper-local editions, while readers experience reduced information overload and increased engagement. This increased ability to personalize content has not only transformed how news is delivered but also introduced new challenges, especially in the realm of algorithmic transparency. The algorithms for personalization, often called "black boxes" (Diakopoulos, 2013; Kitchin, 2016) but we can't manage. They are leveraging and adding to science and technology studies, new media studies, and software studies to analyze the essence of algorithms and their influence and functionality are not transparent but have a significant impact on operational decisions(Powers 2017). Google and Facebook illustrate this through their constantly evolving algorithms, the inner workings of which are not publicly known. However, various factors influencing news selection and prioritization have been recognized. Despite this growing academic attention, these systems are not transparent but have a significant impact on operational decisions. Google,utilizes multiple signals such as user location, search history, language, browser choice, and social connections to personalize search results (Pariser, 2011; Singhal, 2011). Similarly, Google News ranks news content based on factors like relevance, freshness, journalistic standards, and circulation. Editors, news outlets, and hired evaluators serve as gatekeepers in this process (Bui, 2009; Lohr, 2013), and users have some control over personalization by adjusting their settings (Google News Blog, n.d.). Facebook personalizes its News Feed by prioritizing content based on user interactions, preferences, and engagement metrics (Backstrom, 2013; DeVito, 2016). Content ranking is influenced by users' likes, shares, and time spent viewing posts. Facebook also allows manual customization of the News Feed, giving users the ability to determine the content they see (Oremus, 2016). The company has shifted between algorithmic and hybrid approaches, particularly in its trending news section, where it has faced criticisms of bias and misinformation, notably during the 2016 U.S. presidential election (Isaac, 2016; Silverman, 2016). Both Google and Facebook thus rely on a combination of automated and human gatekeepers in their content personalization systems. As these developments unfold, a critical area of focus has emerged around user awareness and perception of news personalization practices. Research on awareness of news personalization has predominantly focused on Facebook and Google. Rader (2014) explored users'

recognition of data collection practices by these platforms, finding that while users believed these companies collected first-party data, there was less certainty regarding tracking across other websites. Rader and Gray (2015) examined Facebook users' perceptions of their News Feeds, discovering that 73% of respondents did not believe they saw every post from their friends, and 65% thought posts were prioritized based on system knowledge of their preferences. Bucher (2017) investigated Facebook users' understanding of algorithms, revealing that despite limited explicit knowledge, users had complex theories about algorithmic functioning and often altered their behavior to influence their feed. Eslami et al. (2015) found that 63% of Facebook users were unaware of algorithmic filtering, attributing missing stories to user behavior rather than algorithmic intervention. These mixed findings suggest a gap in understanding among users and indicate a need for further research, particularly involving participants with varying levels of digital literacy.

4.3 The disadvantage of AI in journalism and its ethical implications

Although AI has brought tremendous benefits to journalism, it is also important to recognize its inherent limitations. One key area where these limitations become evident is in the realm of creativity. Artificial intelligence operates on a rational model, suggesting that any new information failing to diminish uncertainty "beyond the level of chance" is dismissed by artificial neurons. Nevertheless, the previously examined rational processes impose a ceiling on the level of creativity humans can achieve(Seyfodin & Tabrizi,2024). Since AI algorithms are confined to the conceptual frameworks established by their human designers, they cannot reach the pinnacle of creativity, which necessitates the ability to navigate into novel and unforeseen conceptual domains. While AI can process vast amounts of information and identify patterns that might escape human notice, its inherent limitations in

creativity remain evident. This stems from several factors, which become clear when we consider the distinctive qualities that set human creativity apart. The level of creativity that AI algorithms can produce is limited due to several reasons. Firstly, AI algorithms are unable to replicate the neurobiological processes of the human brain that are directly associated with creativity, as they are confined to syntactic symbols without inherent semantic meaning. Secondly, while AI algorithms excel in intelligence and knowledge expansion, their potential for creativity is not directly correlated with high IQ levels beyond a certain threshold. Thirdly, human creativity and discovery at a high level are closely linked to subconscious processes, intuition, dreaming, spontaneity, and emotional intelligence, all of which are connected to "irrational thinking" and are challenging for AI developers to incorporate into algorithms. Lastly, creativity is often tied to human adventure and risk-taking, which are also linked to the human survival instinct, making it difficult to program these human capabilities into an AI artificial brain(Latar 2018).

In journalism, AI raises ethical questions regarding transparency, accountability, and data privacy. Focusing on transparency is essential for the credibility of assessments, necessitating that the results, rationale, and supporting evidence are readily available to the audience. Consequently, innovative approaches are required to facilitate audience comprehension of the outputs generated by artificial intelligence methodologies. At the same time, the rise of AI in journalism has sparked deeper debates about the distribution of ethical responsibility. The need to rethink journalism ethics in the context of algorithmic systems requires an understanding of the extent to which such systems can be considered social actors, potentially responsible for aspects of editorial accountability. This involves the concept of distributed responsibility, where ethical responsibility is shared between human and non-human agents, such as AI, in the news production process. Some scholars, especially from robot ethics, argue that only humans should bear moral responsibility for the consequences of technology. These scholars emphasize the unique rational capabilities of humans and maintain that moral responsibility should rest with the human agents involved, such as designers or users.

In contrast, those who support the notion of distributed responsibility assign social and human-like traits to non-human agents, suggesting that ethical responsibility can be shared. To explore the role of algorithms in journalism, three ethical frameworks—agent-oriented, action-oriented, and patient-oriented—can be applied. Virtue ethics focuses on the agent’s character, while deontology and consequentialism examine the moral nature of actions. The patient-oriented approach, which aligns with the concept of distributed responsibility, focuses on the system as a whole rather than individual agents or actions.

Floridi's (2016) model of distributed moral actions (DMAs) emphasizes that moral responsibility in human-machine interactions is spread across networked agents, both human and artificial. This perspective contrasts with agent-oriented ethics, which emphasizes individual development, and instead prioritizes the well-being of the system and minimizing harm. Applied to journalism, a model of distributed moral responsibility (DMR) is proposed to account for how accountability and responsibility should be shared between humans and algorithms.

Traditional notions of responsibility, based on Aristotelian principles, focus on control over actions and knowledge of those actions, which implies humans bear the full weight of moral responsibility, as algorithms lack awareness or consciousness. However, in scenarios involving automation, such as self-driving cars or automated content generation, responsibility is complicated by the algorithms' limited ability to control their outputs.

Another factor in assigning responsibility is performance. Algorithmic systems often outperform humans in information processing and generation, creating a dynamic similar to a child-adult relationship. Like children, algorithms may lack the capacity to anticipate or control certain outcomes, suggesting a diminished level of responsibility compared to human agents. This "diminished anticipation" complicates the ethical framework, as news workers may not fully understand or anticipate algorithmic behavior, yet algorithms play an increasingly dominant role in news production. At this point, a shift toward a more collective understanding of responsibility becomes essential. Floridi proposes a shift from individual-based moral assessments to a broader, network-oriented approach. He argues for a model where all agents involved in producing a morally loaded outcome—regardless of their intentionality—are held equally accountable by default. This model draws from three concepts: backpropagation from network theory, strict liability from jurisprudence, and common knowledge from epistemic logic. In this framework, agents in a network are collectively responsible for an outcome and must work together to correct or prevent morally negative results. This "faultless responsibility" implies that even without direct fault or intent, each agent contributes to the overall moral status of the system.

Floridi critiques traditional ethics for its focus on individual intentionality, which often leads to overlooking the collective responsibility arising from complex systems. By attributing responsibility through the lens of distributed networks, Floridi's approach emphasizes the need for new ethical frameworks that reflect the interconnected, multi-agent nature of contemporary digital and social environments. His work seeks to adapt moral responsibility to fit these modern contexts, proposing a patient-oriented ethics that focuses more on the outcomes and impacts on the system rather than the intent behind individual actions.

Floridi’s call for a collective responsibility framework is particularly relevant in contexts where algorithms increasingly shape public discourse, such as in automated fact-checking systems. As these AI systems take on critical roles in processing and classifying information, ensuring that they are transparent and accountable becomes essential.

For instance, the classifications produced by automated fact-checking systems should be elucidated upon request. Explainable AI (XAI) represents a specialized area within artificial intelligence dedicated to rendering intricate AI models comprehensible to humans in a structured and interpretable fashion. This involves clarifying AI decisions through various means, such as transparency, interpretative techniques, and natural language generation methods. However, this endeavor becomes particularly challenging when decisions and evaluations stem from the integration of multiple AI-driven techniques and models. A significant concern is that the explanations provided by AI and machine learning solutions may be difficult for non-experts to understand, exacerbating the divide between those with varying levels of information literacy. Furthermore, the explanations regarding the rationale behind news content should be tailored to align with the audience's perceptions of truthfulness and reliability. Another potential issue lies in the accuracy of these explanations, which may also be derived from learning algorithms, emphasizing the need to prioritize correctness over mere believability. Given these complexities, it becomes imperative to consider the broader implications of AI in journalism. Responsible journalism and media practices are rooted in ethical considerations, as they assist journalists in finding a balance between the public's need for information and the potential harm that reporting can cause. Ethical considerations encompass the principles, norms, and guidelines that shape decision-making and guarantee that journalism upholds professional standards and honors the rights and interests of all involved parties (Plaisance, 2013). As the field evolves, these ethical challenges are further compounded by the impact of digital technologies. Journalism studies have been compelled to rethink the concept of journalism in light of the digital transformation, extending beyond traditional boundaries and institutions. This shift has led to various approaches (Steensen & Ahva, 2015) shaped by the current hybrid media ecosystem (Chadwick, 2013), and the ongoing struggle of journalists to maintain the distinctiveness of their professional work compared to other content creators (Eldridge II, 2019). Discussions regarding the boundaries of journalism (Negreira-Rey et al., 2023), intrusive media (Hujanen et al., 2022), journalism versus advertising/entertainment (Balint, 2021), and the impact of social media and digital society (van Dijck, Poell & de Waal, 2018) have fueled debates on digital journalism, which has emerged as a significant institution for generating knowledge in society (Ekström, Lewis & Westlund, 2020). Many of these discussions have centered on the ethical dimensions of digital journalism in today’s online society, with little change observed in the ethical codes of professional

associations and bodies, despite increasing pressure to adapt self-regulation mechanisms to better define the ethical boundaries of online journalism (Mateus, 2019). Studies examining the current codes have identified a gradual and growing integration in both Spanish and Latin American cyberjournalism (Parra-Valcarce, Real-Rodríguez & López-Talavera, 2017), highlighting the slow inclusion of references to the Internet and ICTs (DíazCampo & Segado-Boj, 2015). These studies have also noted the absence of important aspects of computational journalism, such as control and verification of software and techniques used, in recent years (Díaz-Campo & Chaparro-Domínguez, 2020). The delay in updating professional codes of ethics, resulting in a lack of unified guidelines for journalists to follow, is compounded by the introduction of AI in newsrooms, as the relationship between journalism and ethics cannot be overlooked (Rodrigo-Alsina & Cerqueira-da-Silva, 2018). The ethical codes combine the fundamental rules that determine what qualifies as newsworthy based on principles, and the regulatory rules that dictate how news should be produced in journalistic media (Ryfe, 2006), contributing to enhancing the legitimacy and ethical behavior of journalists (Hujanen et al., 2022). Journalists may encounter ethical concerns as they embrace innovative processes that emerge in newsrooms, necessitating the adaptation of ethical standards to accommodate this new reality (García-Avilés, 2021). The increasing use of AI throughout the news production process has sparked enthusiasm in some sectors; however, the implementation of this technology should be accompanied by well-defined frameworks and corresponding proposals, balancing changes with thoughtful and ethical decision-making (Kim, 2019). Due to the cautious approach to implementing AI in newsrooms, some suggestions for ethics in the algorithmic era propose a distributed responsibility model as a practical way to hold multiple actors, including humans and algorithms, accountable for ethical standards (Paik, 2023). This stage of reflection and debate calls for essential proposals to redesign the ethical standards of AI in journalism for the next decade. The ethical implications of AI are becoming increasingly urgent, as highlighted by Tzachor et al. (2020), calling for new technological perspectives and renewed journalistic approaches. Journalistic research has led to a growing body of work advocating for greater consideration of how AI will impact both people and journalism, emphasizing its potential for the betterment of society and the profession (Peña-Fernández et al., 2023). Information professionals are also joining the call for clear guidelines in their regular use and testing of AI, stressing that the absence of ethics undermines the quality of journalism. With the rapid advancement of AI and its integration into journalistic practices, there is a pressing need to update ethical codes to address the increasing concerns of ethical journalists striving to provide valuable and accurate information through cutting-edge AI tools (Thurman, Dorr & Junett, 2017). As automated journalism becomes more commonplace, journalistic organizations are increasingly concerned about the ethical

implications of automation (Porlezza & Ferri, 2022), as well as the regulatory frameworks that govern AI's impact on media and journalism (Porlezza, 2023). While these concerns remain at the forefront of the conversation, it is equally important to acknowledge the growing recognition of AI's potential benefits. The potential of AI technology is widely recognized, as reflected in various national AI strategies, and media coverage often highlights its benefits. Studies indicate that AI's advantages are more frequently discussed than its risks, with portrayals emphasizing AI's superiority over human capabilities and framing it as a beneficial "helping hand" rather than a threatening force (Chuan et al., 2019; Fast & Horvitz, 2017; Bunz and Braghieri, 2022). The more advanced AI becomes, the more positively it tends to be represented (Cools et al., 2022). This optimistic view extends to the media industry, where experts and journalists generally express enthusiasm about AI's efficiency and potential to free up time for investigative journalism (Beckett, 2019; Gutierrez Lopez et al., 2021; Porlezza & Ferri, 2022).

However, public discourse has grown more critical over time, with increasing concerns about AI's ethical implications, loss of control, and its impact on jobs (Nguyen and Hekman, 2022; Fast and Horvitz, 2017). Despite this, positive portrayals still dominate, with negative views often framed in dystopian terms (Porlezza, 2019). Importantly, how AI is depicted in the media can influence its implementation and regulation within the news sector. Brennan et al. (2022) argue that media representations of AI as a near-artificial general intelligence can shape political perceptions. This evolving narrative underscores the need for immediate and thoughtful action.It is imperative to address these issues promptly, as the lack of clear ethical guidelines only fuels public distrust in journalism. Meanwhile, AI lacks the inherent commitment to truthfulness that human journalists possess, further emphasizing the urgency of establishing specific decisions and measures. Any regulatory responses should be comprehensive, considering the multifaceted impact of AI on individuals, newsrooms, and media markets (Helberg et al., 2022).

4.4 The future?

The socio-political, legal, occupational, and psychological implications of automated news generation algorithms in journalism are significant for news organizations, journalists, and audiences (Montal & Reich, 2017). The authorship of these algorithms is a contentious issue that has not been thoroughly explored. Despite the increasing sophistication of automated reporting, journalists are expected to adapt while retaining their essential roles, given the enduring importance of journalism as a professional and ideological field (Linden, 2017). Although journalists may resist the transformation, the evolution of news organizations presents opportunities for

technological, economic, and political advancements (Bourdieu, 2005). Technology companies recognize the pivotal role of news organizations in advancing digital literacy, and other industries such as healthcare, manufacturing, and sports also stand to benefit from automation (Wu et al., 2019). The integration of digital transformation is essential across various sectors, including journalism (Alam et al., 2018; Kelly et al., 2022). Cetindamar Kozanoglu and Abedin (2021) introduce a new concept of digital literacy aimed at enhancing employees' involvement in these transformations, which is crucial for enterprises. The development of a national digital literacy system necessitates a comprehensive understanding of population demographics and their interaction with automated systems. This study lays the groundwork for conceptualizing digital literacy and provides insights into the sociodemographic factors influencing digital transformation in news organizations. As the industry grapples with these changes. The implementation of automation in the field of journalism has led to a decrease in employment opportunities, affecting the professional identity of journalists. According to Sherwood and O'Donnell (2018), the loss of employment weakens the identity of journalists, emphasizing the significant connection between their work and their sense of self. This change could have extensive implications for the media industry and its influence on society (Schapals and Porlezza, 2020). Transitioning from these foundational shifts in employment and identity, it is important to consider how journalistic practice has always been influenced, structured, and simultaneously limited by technological processes. The practice of journalism has always been affected and shaped by technological processes, which have influenced the quality and depth of news. Media transformation processes have disrupted traditional news production and consumption in both positive and negative ways. AI-driven news reporting is versatile, covering various sections such as politics, economy, business, technology, health, social issues, culture, and sports. The societal importance of AI is on the rise due to its ability to cover a wide range of fields. However, the implementation of automation in editorial offices has not progressed due to a lack of resources and insufficient technological skills among journalists. The future training of journalists with a focus on socialization and technology can enhance the status of journalism as a profession and increase its independence from external influences. The incorporation of advanced technologies into editorial practices requires extensive training for staff editors and the availability of relevant resources. Mastery of data handling, coding, and algorithmic techniques can enhance editorial effectiveness and improve competitiveness. While many journalists view automation as a supportive tool rather than a replacement, the emergence of automated content generation presents significant risks to job stability and overall industry security. Technology firms are expected to gain increasing influence, potentially challenging established notions of journalism. Therefore, it is vital to uphold the essential roles and skills of journalists,

along with the ideological coherence of editorial work. As automated processes become more integrated into journalism, developing new conceptual frameworks will be necessary to promote effective interaction between humans and machines.

Some consideration

Among its advantages, we can note the reduction in time and the ability to expedite the process of creating an article. These positive aspects make AI a valuable ally because, in an increasingly connected world, we see the speed of news and the desire to enhance the accuracy of information. One way AI assists is through fact-checking, as it is a useful tool for combating fake news, which has become more prevalent in recent times through the use of social networks.

However, AI also has negative aspects, such as: Fairness and inclusion, which are significant issues. The use of unbalanced data can lead to discriminatory decisions by AI systems. Additionally, excessive digitization could exclude certain social categories, exacerbating preexisting inequalities (Ntoutsi et al., 2019).

Freedom and control over decisions are threatened by the lack of transparency in AI's decision-making processes, which are often not explainable. This undermines trust and compromises cooperation between humans and AI, limiting AI's potential to support the development of human capabilities.

Another important consideration is the impact on personal growth and human innovation, particularly in the educational context. AI, capable of autonomously generating content, could hinder the development of cognitive and creative skills, reducing opportunities for individuals to acquire essential competencies for future innovation.

A further risk concerns the truth and quality of information. Systems like ChatGPT can generate false or inappropriate information, undermining trust and collaboration based on truthful data, with consequences at both the local and global levels. For instance, ChatGPT presents numerous relevant solutions and opportunities for journalism, particularly by streamlining routine activities and reducing time spent on simple tasks. Its advantages include writing in various formats, assisting in the selection of topics for media coverage, and adapting content for social networks. Although these tasks are more communication-focused than journalistic in nature, they allow journalists to engage in more in-depth research and complex reporting.

However, the tool also has significant limitations, the most critical being its occasional inaccuracy. ChatGPT can acknowledge its limited dataset, which only extends to 2021, or state that it is not connected to the Internet, resulting in potentially outdated or incorrect information. Worse, it can fabricate information without warning, risking the spread of false or inaccurate content, especially when used by inexperienced journalists. Although OpenAI has announced a version connected to the Internet, which will mitigate this issue, there remains the challenge of biased or dominant ideas shaping its output. This tendency to favor frequently repeated content over accurate or nuanced information can perpetuate traditional biases, similar to the "tyranny of the majority" in politics (Harper, 2017). These limitations raise ethical concerns about AI's role in journalism, particularly regarding its potential to replace human judgment in areas intrinsic to the profession. This necessitates a reevaluation of journalistic training to ensure that future journalists are equipped with ethical knowledge and critical thinking skills, enabling them to use AI tools while preserving journalistic independence and core values (Peña-Fernández, Meso-Ayerdi, Larrondo-Ureta, 2023).

Furthermore, artificial intelligence (AI) has become a pervasive presence with a profound impact on the global dynamics of contemporary society. This technology is already significantly influencing our daily lives and will continue to do so in the future. To avoid being passively dragged toward a future that may not reflect our principles, it is essential to undertake a path of awareness and collective cooperation, aimed at outlining a shared vision of the future we desire, as well as the acceptable paths for achieving it.

AI ethics play a crucial role in this global effort, promoting multi-stakeholder collaboration where all parties involved are called upon to contribute: companies, legislators, academics, media, civil society, and individuals, including consumers of AI-based services.

Each of us can contribute to this path by using technology consciously, responsibly, and respectfully toward others and the environment. We must not rely on technology passively, but use it to enrich our knowledge, explore new horizons, engage in dialogue with others, change our opinions, find compromises, and collaborate. In our daily lives, we must embody the principle of using technology to improve ourselves, not the other way around. We all have the duty and opportunity to contribute to creating a future where AI supports and reflects our values.

In conclusion, I believe that the future of journalism will be strongly influenced by AI, bringing new innovations. This is coupled with the integration of digital network technologies and the evolution of the relationship between producers and consumers. As consulting and entrepreneurial activities become integral parts of journalism, the boundaries and identity of the profession are blurring. The full implications of these changes are not yet fully understood, partly due to limited empirical research on journalistic work and a lack of interaction with broader socio-economic literature.

Conclusion

This study set out to explore the impact of AI in sport journalism and how can implement the journalism. According several studies many journalists risk their job, but it will be an ally because automated journalism can save time and avoid repetitive task. Moreover the AI has emerged as a transformative technology for journalism and media, increasingly relevant to both fields and their associated educational programs. This essay highlights AI's potential to produce content with a high degree of accuracy, grammatical correctness, and factual reliability, positioning it as a valuable tool for journalistic work. However, AI systems, such as ChatGPT, also exhibit

significant limitations, particularly in their depth of knowledge and ability to engage in critical or creative thinking. Despite these constraints, AI platforms demonstrate considerable knowledge of journalism, including historical figures, media scholars, and contemporary issues like "news deserts."

Generative AI systems, capable of mimicking human-like responses, may challenge the role of human journalists, especially amid shrinking newsroom budgets and increasing pressures for economic efficiency. While AI lacks sentience and self-awareness, its ability to process and present information effectively positions it as a potential asset for assisting human journalists, enhancing both the quality and efficiency of their work. In light of these developments, it is essential to consider how the implementation of AI technologies is reshaping journalistic workflows and redefining the roles of journalists in the digital age. The integration of Artificial Intelligence (AI) into journalism has introduced a transformative shift, primarily in enhancing operational efficiency and productivity. AI automates routine tasks such as data collection and analysis, which previously required significant time and effort. By managing these repetitive functions, AI enables journalists to focus on more intricate and nuanced aspects of storytelling, thereby enhancing the overall quality of journalism. This shift allows for deeper engagement with complex narratives, where human attributes such as insight, empathy, and investigative skills become more prominent. In high-pressure environments like breaking news or live reporting, AI proves invaluable, offering real-time data analysis and content suggestions that accelerate the news delivery process, ensuring audiences receive timely and relevant information.

Additionally, AI has revolutionized data analysis and reporting by processing large datasets quickly and efficiently, revealing trends and patterns often difficult for humans to detect. This capability has significantly impacted fields such as election reporting, financial journalism, and sports coverage. AI's advanced data analysis has deepened investigative journalism, enabling the discovery of stories embedded within complex data layers and offering insights grounded in robust, data-driven analysis.

AI has also improved the personalization of news content by analyzing user behavior and preferences, curating tailored news feeds and article recommendations. This personalized approach enhances user engagement and fosters stronger relationships between audiences and news platforms, leading to increased loyalty and platform usage. Furthermore, AI’s role in automated reporting and content generation, particularly in producing timely and accurate reports from structured data, has become essential in fields like finance, sports, and elections. This automation supports newsrooms in meeting the growing demand for rapid, consistent news dissemination in a 24-hour news cycle.

Overall, AI's integration into journalism has streamlined many processes, enriching both the content and the context of news reporting. It marks a significant advancement by augmenting human capabilities, enabling journalists to delve deeper into complex topics and contributing to the evolution of the field. The shift toward AI-driven automation is seen as a substantial progression that enhances both the efficiency and depth of journalistic practice.

AI has also negative impact The integration of AI into journalism, as explored by Nurelmadina et al. (2021), presents a multifaceted landscape of both advancements and challenges. Ethical concerns are prominent, particularly regarding bias and transparency in AI algorithms used to select and prioritize news. These algorithms, often opaque in nature, may unintentionally reinforce biases present in their training data, raising issues of impartiality and accountability in AI-driven journalism. Furthermore, questions of authorship and credibility arise as readers may unknowingly consume AI-generated content, leading to debates over trust and authenticity in the journalistic process(Verma 2021)

The potential for job displacement due to AI's growing capabilities in routine reporting and data analysis is another pressing concern. This development threatens not only employment but also the value of human journalistic skills, which could be overshadowed by automation. Qureshi & Tekin (2020) also highlight risks associated with AI's ability to produce realistic content, such as deepfakes, which could jeopardize the integrity of news through misinformation and manipulated media. This underscores the necessity for AI-generated content to remain accurate and transparent.

Additionally, the over-reliance on AI in news creation can foster echo chambers, where algorithms cater to users' existing beliefs, limiting exposure to diverse perspectives and hindering the development of an informed public. AI systems also struggle with understanding context, sarcasm, and cultural nuances, leading to potential inaccuracies, which emphasizes the continued importance of human journalists for ethical decision-making and nuanced reporting.

While AI offers benefits such as efficiency, enhanced data analysis, and personalized content, it also introduces significant challenges, including ethical dilemmas, job displacement, and misinformation risks. As AI evolves, the journalism industry must carefully navigate these issues to ensure responsible and ethical use of the technology in news production and dissemination.

At the end the ongoing advancements in AI technology will significantly transform journalism. However, it asserts that the role of human journalists remains indispensable, particularly in areas requiring complex interpretation, storytelling, ethical judgment, and investigative depth. The conclusion emphasizes the importance

of a symbiotic relationship between AI and human reporters, where AI's capabilities in efficiency and data processing enhance, rather than replace, the critical thinking and ethical oversight of journalists. It advocates for a balanced integration of AI to preserve journalism's credibility, reliability, and ethical integrity in an increasingly digital and AI-driven landscape.

Bibliography

Adar Eytan, Gearig Carolyn(2017) *PersaLog: Personalization of News Article Content*

Araújo Lucas Vieira, (2017) *Adoção de algoritmos, NLG e inteligência artificial na imprensa brasileira em âmbito nacional e regional. Estudos em Jornalismo e Mídia 14: 2*

Asraful Alam.Namhee Cho,Kyun Soo Kim,(2018) *The Role of News Media Literacy in Predicting News Personalization and News Engagemen*

Atton Chris, Hamilton James (2008), *Alternative Journalism*

Backstrom Lars (2013) *News Feed FYI.*

Barassi Veronica, Trerè Emiliano, (2012), *Does Web 3.0 come after Web 2.0? Deconstructing theoretical assumptions through practice*

Bateman, John A. (1997). *Enabling technology for multilingual natural language generation:the KPML development environment.Natural Language Engineering,3(1), 15–55*.

Beckett, C. (2019). *New powers, new responsibilities: A global survey of Journalism and Artificial Intelligence. London: The London School of Economics.*

Belinkov Yonatan, Durrani Nadir, Dalvi Fahima(2017), *What do Neural Machine Translation Models Learn about Morphology?*

Belz Anja, Kow Eric (2010). *Comparing rating scales and preference judgements in languageevaluation. InProc. INLG'10, pp. 7–15*

Berthou T. (1999) *Dictionnaire historique des clubs de football français, Créteil: Pages de foot.*

Block, H. D., Knight, B., and Rosenblatt, F. (1962). *Analysis of a four-layer series-coupled perceptron. Rev. Modern Physics, 34, 275–282.*

Bodò Balazs (2019) , *Selling News to Audiences – A Qualitative Inquiry into the Emerging Logics of Algorithmic News Personalization in European Quality News Media* https://doi.org/10.1080/21670811.2019.1624185

Bollmann Marcell(2011). *Adapting SimpleNLG for German. InProc. ENLG'11, pp. 133–138*

Bonington Paul, (1995), *The fourth media. Internet world*

Boyle Raymond (2006). *Sports journalism: context and issues. Sage*

Boyle Raymond, Haynes Richard (2010), *New Media Sport*

Bradshaw Paul(2007), *Are wikis the new blogs?*

Branch Taylor (2011). *The shame of college sports. The Atlantic, 308(3), 80-110*

Brugger Niels,(2015), *A brief history of Facebook as a media text: The development of an empty structure in*

Bryant, J. & Holt, A. (2006). *A historical overview of sports and media in the United States. In Raney, A. A., & Bryant, J. (Eds.). (2006). Handbook of sports and media. New York: Routledge, 22-45*.

Bui CamLy N. 2009. "*Examining the New Gatekeepers: News Portals' Inclusion and Rankingof Media and Events." Paper presented at the annual International Communication Association Conference, Chicago, Illinois*

Canavilhas Joao,Artificial Intelligence and journalism(2022). *Current situation and expectations in the portoguese sports media*

Carlson Matt, (2015), *The Robotic Reporter,Automated journalism and the redefinition of labor, compositional forms, and journalistic authority*

Carreira, Krishma A. C.(2017). *Notícias Automatizadas—A Evolução que Levou o Jornalismo a ser Feito por Não Humanos. Master's thesis, Universidade Metodista, São Paulo, Brazil*

Castro Ferreira Thiago, Wubben S, & Krahmer, E. (2017). *Generating flexible proper namereferences in text: Data, models and evaluation. InProc. EACL'17, pp. 655–664*

Cettolo Mauro, Girardi Christian,Federico Marcello,(2012)*WIT3 : Web Inventory of Transcribed and Translated Talks*

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

Collobert Ronan, Weston Jason, Bottou Leon, Karlen Michael, Kavukeuoglu, K., & Kuksa, P. (2011). *Natural language processing (almost) from scratch. Journal of Machine Learning Research, 12(1)*

Conneau Alexis,Schwenk Holger, Barrault Loic (2017) *"Very deep convolutional networks for text classification," in Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics, vol. 1, 2017, pp. 1107–1116*

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

Costa Pinto Moises, Oliveira Barbosa Suzana (2024), *Artificial Intelligence (AI) in Brazilian Digital Journalism: Historical Context and Innovative Processes*

Cremaschi Marco, Bianchi Federico, Maurino Andrea, Pierotti Primo Andrea(2019), *Supporting Journalism by Combining Neural Language Generation and Knowledge Graphs*

Creutz Mathias, Lagus Krista(2007), *Unsupervised models for morpheme segmentation and morphology learning*

Dalben, Silvia,(2020) *Jornalismo Automatizado No Brasil: Análise de três robôs no Twitter. Brazilian Journalism Research*

Dalben, Silvia,(2022) *Uso de Inteligência Artificial nas Redações Jornalísticas na Guerra Contra a Corrupção na América Latina. In Mobilidade e Inteligência Artificial. Os Novos Caminhos do Jornalismo*

de Lara-González Alice, García-Avilés Jose Alberto, Arias-Robles Felix(2022), *Implantación de la Inteligencia Artificial en los medios españoles: análisis de las percepciones de los profesionales,* https://doi.org/10.56418/txt.15.2022.001

De Lima Santos Felipe, Ceron Wilson (2022), *Artificial Intelligence in News Media: Current Perceptions and Future Outlook*

de Oliveira Rodrigo, Sripada Somayajalu (2014). *Adapting SimpleNLG for Brazilian Portugese realisation. InProc. INLG'14, pp. 93–94.*

Deng Jia, Dong Wei, Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). *Imagenet: A large-scale hierarchical image database. In CVPR-09.*

Derr Konstantin (2016), *Mapping the algorithmic journalism*

Deuze Mark (2001) *"Online Journalism: modelling the first generation of news media on the World Wide Web", First Monday 6(10), http://firstmonday.org/htbin/cgiwrap/bin/ojs/ index.php/fm/article/view/893/802*

Deuze Mark(2005). *What is journalism? Professional identity and ideology of journalists reconsidered. Journalism, 6(4):442–464, 2005.*

DeVito Michael A. (2016). *"From Editors to Algorithms: A Values-Based Approach to Understanding Story Selection in the Facebook News Feed." Digital Journalism. doi: 10.1080/ 21670811.2016.1178592*

Devlin j, M.-W. Chang, K. Lee, and K. Toutanova, (2018) *"BERT: Pretraining of deep bidirectional transformers for language understanding,"*

Diakopoulos Nicholas (2019) *Automating the news: how algorithms are rewriting the media. Harvard University Press*

Diakopoulos Nicholas, Koliska Michael (2016). *Algorithmic transparency in the news media. Digital Journalism, pages 1–20.*

Diaz Noci Javier, Meso Ayerdi Koldo(1998), *Desarrollo del periodismo electrónico". El profesional de la información, v. 7, n. 12, pp. 4-11.*

Dorr Konstantin, (2016), *Mapping the algorithmic journalism*

Eisensten Jacob, (2018), *Natural Language Processing*

Elhadad, M., & Robin, J. (1996). *An overview of SURGE: A reusable comprehensive syn-tactic realization component. InProcedings of the 8th International Natural LanguageGeneration Workshop (IWNLG'98), pp. 1–4*

English Peter(2015), *Mapping the sports journalism field: Bourdieu and broadsheet newsrooms*

Eskenazi, J (2012). *Top 5 ways Bleacher Report rules the world!*

Essenfelder Reneto, Sant'Anna Emilio(2022). *Inteligência Artificial Aplicada ao Jornalismo no Brasil: Experiências e Obstáculos à Adoção de Tecnologias Inteligentes. In Mobilidade e Inteligência Artificial. Os Novos Caminhos do Jornalismo.*

Evans, Thomas G. (1968). *A program for the solution of a class of geometric-analogy intelligence-test questions. In Minsky, M. L. (Ed.), Semantic Information Processing. MIT Press.*

Fast Ethan, Horvitz Eric, (2017), *Long-Term Trends in the Public Perception of Artificial Intelligence*

Feigenbaum Edward. A., Buchanan, B. G., and Lederberg, J. (1971). *On generality and problem solving: A case study using the DENDRAL program.*

Finkelstein Lev, Yossi Mathias, Gabrilovich Evgeniy, (2001), *Placing search in context: The concept revisited*

Floridi Luciano (2009) *Web 2.0 vs. the semantic web: A philosophical assessment. Episteme 6(1): 25–37*

Floridi Luciano, (2016) *Faultless responsibility: on the nature and allocation of moral responsibility for distributed moral actions.Phil. Trans. R. Soc. A 374: 20160112. http://dx.doi.org/10.1098/rsta.2016.0112*

Forja-Pena, Tania, García-Orosa Berta & LópezGarcía, X. (2024). *The Ethical Revolution: Challenges and Reflections in the Face of the Integration of Artificial Intelligence in Digital Journalism, Communication & Society, 37(3), 237-254. https://doi.org/10.15581/003.37.3.237-254*

Funk Tom (2008) *Web 2.0 and Beyond: Understanding the New Online Business Models, Trends, and Technologies. Ann Arbor: The University of Michigan Press*

Gailily Yair (2018), *AI and sport journalism : is it a sweeping change?*

Gandomi, A., & Haider, M. (2015). *Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007*

Gantz W, Wenner L.A, (1991), *Men woman and sports : Audience experiences and effects*

García-Avilés, J.-A. (2021). *Review article: Journalism innovation research, a diverse and flourishing field (2000-2020). Profesional de la Información, 30(1), e300110. https://www.doi.org/10.3145/epi.2021.en*

Gatt Albert, Krahmere Emiel,(2018) *Survey of the State of the Art in Natural LanguageGeneration: Core tasks, applications and evaluation*

Gatt Albert, Portet F., Reiter, E., Hunter, J. R., Mahamood, S., Moncur, W., & Sripada, S.(2009). *From data to text in the neonatal intensive care Unit: Using NLG technologyfor decision support and information management.AI Communications,22(3), 153–186*

Goldberg Yoav (2017). *Neural network methods for natural language processing. Synthesis Lectures on Human Language Technologies, 10.*

Gómez-Diago Gloria (2022*). Perspectivas para abordar la inteligencia artificial en la enseñanza de periodismo. Una revisión de experiencias investigadoras y docentes. Revista Latina De Comunicación Social, 80, 29-46.* https://www.doi.org/10.4185/RLCS-2022-1542

Goodfellow Ian, Bengio, Yoshua, and Courville Aaron (2016). *Deep Learning. MIT Press*

Graefe Andreas (2016) *Guide to automated journalism. Center for Digital Journalism*

Graefe Andreas, Bohlken Nina (2020). *Automated journalism: A meta-analysis of readers' perceptions of human-written in comparison to automated news". Media and communication, v. 8, n. 3, pp. 50-59,* https://doi.org/10.17645/mac.v8i3.3019

Graves Lucas, Amazeen Michelle, (2019)*, Fact-Checking as Idea and Practice in Journalism*

Green Cordell (1969). *Theorem-proving by resolution as a basis for question-answering systems. In Meltzer, B., Michie, D., and Swann, M. (Eds.), Machine Intelligence 4. Edinburgh University Press*

Grice, H. P. (1975). *Logic and conversation. InSyntax and Semantics 3: Speech Acts, pp.41–58. Elsevier, Amsterdam*

Gunasiri DY and Jayaratne KL (2019) *Automated cricket news generation in Sri Lankan style using natural language generation. European Journal of Computer Science and Information Technology 7(9): 42–56*

Gutiérrez-Caneda Beatriz,Vázquez-Herrero Jorge, López-García Xosé (2023). "*AI application in journalism: ChatGPT and the uses and risks of an emergent technology". Profesional de la información, v. 32, n. 5, e320514.* https://doi.org/10.3145/epi.2023.sep.14

Hansen Mark, Roca-Sales, M., Keegan, J. & King G. (2017*). Artificial Intelligence: Practice and Implications for Journalism. Brown Institute for media innovation and the tow center for digital journalism. Columbia Journalism School.* https://www.doi.org/10.7916/d8x92prd

Hardin Marie, Kuehn Kathleen M., Jones, H., Genovese, J., & Balaji, M. (2009). *'Have You Got Game?'Hegemonic Masculinity and Neo-Homophobia in US Newspaper Sports Columns. Communication, Culture & Critique, 2(2), 182-200*

Harper Tauel (2017). *"The big data public and its problems: Big data and the structural transformation of the public sphere". New media & society, v. 19, n. 9, pp. 1424-1439*

Harris Daniel (2008) *Web 2. 0 Evolution into the Intelligent Web 3. 0: 100 Most Asked Questions on Transformation, Ubiquitous Connectivity, Network Computing, Open Technologies, Open Identity, Distributed Databases and Intelligent Applications.*

Hermida Alfred(2015), *Power plays on social media in Journalism Practice*

Hermida Alfred, Fletcher Fred, Korell Darryl, Logan Donna(2012), *Share, Like, Recommend*

https://doi.org/10.5210/fm.v20i5.5423

Hur Youngjin , Ko jae Yong, Claussen Cathryn (2007), *Determinants of using sports web portals: an empirical examination of the Sport Website Acceptance Model*

Ibbitson John,(2008), *Race magnifies Internet's growing role in media*

Isaac Mike (2016). *"Facebook, in Cross Hairs after Election, is Said to Question Its Influence."*

Joao Canavilhas, Fabio Giacomelli(2023) *Intelligencia artificial en el periodismo deportivo estudio en Brasil y Portugal*

Jones Bronwyn, Jones Rhianne (2019). *"Public service chatbots: Automating conversation with BBC News". Digital journalism, v. 7, n. 8, pp. 1032-1053.* https://doi.org/10.1080/21670811.2019.1609371

Jones, D. and Baró i Queralt, J. (1996) *'La prensa', in D. Jones (ed.) Sport i mitjans de comunicació a Catalunya*

Jordan Pamela W., Walker Marilyin A. (2005). *Learning content selection rules for generatingobject descriptions in dialogue.Journal of Artificial Intelligence Research,24, 157–194*

Joseph Sethunya, Hlomani Hlomani, Letsholo Keletso, Kaniwa Freeson, Sedimo Kutlawo,(2016), *Natural Language Processing: A Review*

Kahle, L. & Meeske, C. (1999). *Sports marketing and the Internet: It's a whole new ball game*

Kanerva Jenna, Ronnqvist Samuel, Kekki Riina, Salakoski Tapio, Ginter Filip,(2019) *Template-free Data-to-Text Generation of Finnish Sports News*

Kasper Robert T. (1989). *A Flexible Interface for Linking Applications to Penman's SentenceGenerator. InProc. Workshop on Speech and Natural Langauge, pp. 153–158*

Kawamoto Kevin(2003), *Digital journalism: Emerging media and the changing horizons of journalism*

Kenter T, Borisov A, C. Van Gysel, M. Dehghani, M. de Rijke,and B. Mitra(2017), "*Neural networks for information retrieval," inProc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2017,pp. 1403–1406.*

Kim Daewon, Kim, Seongcheol (2017). *Newspaper companies' determinants in adopting robot journalism". Technological forecasting and social change, v. 117, pp. 184-195.* https://doi.org/10.1016/j.techfore.2016.12.002

Kim, H. (2019). *AI in Journalism: Creating an Ethical Framework. Syracuse University Honors Program Capstone Projects. 1083. Retrieved from* https://surface.syr.edu/honors_capstone/10

Krizhevsky Alex, Sutskever, Ilya, and Hinton, Geoffrey E. (2013). *ImageNet classification with deep convolutional neural networks. In NeurIPS 25*

Kunert Jessica(2020), *Automation in sports reporting : strategies of data providers, software providers, and media outlets*

Kunert Jessica, Duncan Sam,Karg Adam(2024), *Attitudes to automated and human written sport journalism*

Lacy S (2011*). As football season kicks off, Bleacher Report raises $22 million more*

Latar Noam Lemelshtrich, (2018), *Are AI's limitations creating new opportunities for human journalists?*

LeCompte Celeste (2015) *Automation in the newsroom. How algorithms are helping reporters expand coverage, engage audiences, and respond to breaking news.*

LeCun Yann, Jackel, L, Bottou, L., Brunot, A., Cortes, C., Denker, J., Drucker, H., Guyon, I., Muller, U., Sackinger, E., Simard, P., and Vapnik, V. N. (1995). *Comparison of learning algorithms for handwritten digit recognition. In Int. Conference on Artificial Neural Networks*

Leppanen Leo, Munezero Myriam, Granroth-Wilding Mark, Toivonen Hannu, *DataDriven News Generation for Automated Journalism. In Proceedings of the 10th International Conference on Natural Language Generation, pages 188–197, Santiago de Compostela, Spain, September 2017. Association for Computational Linguistics*

Liddy Elizabeth, (2001), *Natural Language Processing*

Lievrouw Leah A., *The handbook of new media*

Lima Lucillen, Barbosa Suzan (2022) *Mobilidade, imersão e experiência: entre ajustes e discordâncias conceituais no jornalismo imersivo. Estudos Em Comunicação 34:1–11*

Linden Carl Gustav(2017) , *Decades of Automation in the Newsroom Why are there still so many jobs in journalism?, https://doi.org/10.1080/21670811.2016.1160791*

Lohr Steve (2013) *"Algorithms Get a Human Hand in Steering Web." The New York Times*

Lokot Tetyana, Diakopoulos Nicholas (2016). *"News bots". Digital journalism, v. 4, n. 6, pp. 682-699, https://doi.org/10.1080/21670811.2015.1081822*

Long, M., & Connelly, E. (2017). *SportsPro smart series – Smart data. SportsPro*

Lopez-Garcia Xose, Vizoso Angel(2021), *Periodismo de alta tecnología: signo de los tiempos digitales del tercer milenio, https://doi.org/10.3145/epi.2021.may.01*

Luong Minh-Thang, Socher Richard, Manning Christopher(2013), *Better Word Representations with Recursive Neural Networks for Morphology*

Mani, I. (2001).Automatic Summarization. John Benjamins Publishing Company, Amsterdam

Marconi Francesco (2020). *Newsmakers: Artificial Intelligence and the Future of Journalism. New York, NY: Columbia University Press.*

Mazzei Alessandro, Battaglino Cristina, Bosco Crisitina (2016). *SimpleNLG-IT : adapting SimpleNLG toItalian. InProc. INLG'16, pp. 184–192*

McBride Kelly, Rosenstiel Tom (2013). *The new ethics of journalism: Principles for the 21st century. CQ Press.*

McChesney, Robert W. (1989). *Media made sport: A history of sports coverage in the United States. Media, sports, and society, 49-69*

McCulloch Warren S., Pitts Walter (1943). *A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5, 115–137.*

McEnnis Simon(2020) , *Toy department within the toy department? Online sports journalists and professional legittimacy*

McKeown Kathleen, (1992). *Text generation. Cambridge University Press*.

McMillan, Sally J. (2002) *"Exploring Models of Interactivity from Multiple Research Traditions: users, documents, and systems", in: Leah A. Lievrouw and Sonia A. Livingston (Eds), Handbook of New Media, London: Sage, pp. 162-182.*

McMillan, Sally J. (2005) *"The Researchers and the Concept: moving beyond a blind examination of interactivity", Journal of Interactive Advertising 5(2), http://jiad.org/ article58*,

McRoy Susan W., Channarukul Songsak, Ali Syed S. (2003). *An augmented template-based approachto text realization.Natural Language Engineering,9(4), 381–420*

Mellish Chris, Scott Donia, Cahill Lynn, Paiva, D. S., Evans, R., & Reape, M. (2006). *A ReferenceArchitecture for Natural Language Generation Systems.Natural Language Engineer-ing,12(1), 1–34*

Meteer, M. W., McDonald, D. D., Anderson, S., Forster, D., Gay, L., Iluettner, A., & Sibun,P. (1987). *Mumble-86: Design and Implementation (Technical Report COINS 87-87).Tech. rep., University of Massachusetts at Amherst, Amherst, MA.*

Minsky, Marvin L. and Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry. MIT Press.*

Montal Tal , Reich Zvi (2017) *I, Robot. You, Journalist. Who is the Author?, Digital Journalism, 5:7, 829-849, DOI: 10.1080/21670811.2016.1209083*

Moore, Johanna D.,Paris Cecile (1993). *Planning text for advisory dialogues: Capturing intentional and rhetorical information. Computational Linguistics 19(4), 651–694.*

Morita Hajime, Kawahara Daisuke, Kurohashi (2015), *Morphological analysis for unsegmented languages using recurrent neural network language model*

Murphy Kevin (2012). *Machine Learning: A Probabilistic Perspective. MIT Press*

Nelson, Theodor H. (1965), *Complex Information Processing: a file structure for the complex, the changing and the indeterminate", in: Lewis Winner (Ed.), Proceedings of the 20th ACM Annual Conference, New York: Association for Computing Machinery, pp. 84-100*

Nenkova Ani, McKeown, Kathleen R. (2011). *Automatic Summarization.Foundations andTrends in Information Retrieval,5(2-3), 103–233.*

Nguyen Dennis, Hekman Erik, (2022), *The news framing of artificial intelligence: a critical exploration of how media discourses make sense of automation*

Noain-Sanchez Amaya(2021), *Addressing the Impact of Artificial Intelligence on Journalism: the perception of experts, journalists and academics*

Norvig Peter, Russell Stuart,(2020), *Artificial Intelligence: A Modern Approach*

Oates Thomas P, Pauly John. (2007). *Sports journalism as moral and ethical discourse. Journal of Mass Media Ethics, 22(4), 332–347. https://doi.org/10.1080/08900520701583628*

Oh Changhoon, Choi Jinhan, Lee Sungwoo, Park SoHyun, Kim Daeryong, Jungwoo Song, Dongwhan Kim, Joonhwan Lee, and Bongwon Suh(2020). *Understanding User Perception of Automated News Generation System. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pages 1–13, Honolulu HI USA, April 2020. ACM.*

Oremus Will (2016) *"Who Controls Your Facebook Feed?"*

Ostrow A(2008), *Bleacher Report launches citizen journalism for sports; Raises Series A*

Otter Daniel, Medina Julian, Kalita Jugal,(2021), *A Survey of the Usages of Deep Learning for Natural Language Processing*

Paik, S. (2023). *Journalism Ethics for the Algorithmic Era. Digital Journalism. https://www.doi.org/10.1080/21670811.2023.2200195*

Paradigm Shift or Passing Fad? Twitter and Sport journalism, Mary Lou Sheffer, Brad Schultz, 2010

Pariser Eli (2011). *The Filter Bubble: What the Internet is Hiding from You. New York: The Penguin Press.*

Paulsen Steve, Harder Raymond, Johnson Michiel, *Facebook and news journalism*

Pavlik John (2019), *Journalism in the age of virtual reality: how experiential media are transforming news. Columbia University Press, New York*

Peña-Fernández Simón; Meso-Ayerdi Koldobika; Larrondo-Ureta Ainara; Díaz-Noci Javier (2023), *Without journalists, there is no journalism: the social dimension of generative artificial intelligence in the media*

Perse Elizabeth, (1992), *Predicting attention to local television news: Need for cognition and motives for viewing*

Piwek Paul, Boyer Kristy E. (2012). *Varieties of question generation: Introduction to thisspecial issue.Dialogue and Discourse,3(2), 1–9.*

Poesio Massimo, Stevenson Rosemary, Di Eugenio Barbara, Hitzeman Janet (2004). *Centering: A parametrictheory and its instantiations.Computational Linguistics,30(3), 309–363*

Ponce J. (2014). *Sony launches social network that will form centerpiece of 2014 World Cup sponsorship. SportTechie*

Porlezza Colin (2023). *Promoting responsible AI: A European perspective on the governance of artificial intelligence in media and journalism. Communications, 48(3), 370-394.* *https://www.doi.org/10.1515/commun-2022-0*

Porlezza Colin , Ferri Giulia (2022). *The missing piece: ethics and the ontological boundaries of automated journalism. #ISOJ 12(1), 71-98.*

Powers Elia(2017), *My News Feed is Filtered? Awareness of news personalization among college students in* *https://doi.org/10.1080/21670811.2017.1286943*

Puig Nuria, Martinez Joaquin and Garcia Borja, (2010) *Sport policy in Spain*

Radford Alec, Narasimhan Karthik(2018). *Improving language understanding by generative pre-training.*

Radford Alec, Wu Jeff, Child Rewon, Luan David(2019), *Language models are unsupervised multitask learners. 2019.*

Reape, M., Mellish, C. (1999). *Just what is aggregation anyway? InProc. ENLG'99,pp. 20–29*

Reiter Ehud, Dale Robert (1997). *Building natural-language generation systems.NaturalLanguage Engineering,3, 57–87.*

Rojas Torrijos Jose Luis (2018)**,***Automated sports coverage. Case study of bot released by The Washington Post during theRío 2016 and PyeongChang 2018 Olympics*

Rojas Torrijos Jose Luis, Bran Tournal Carlos(2019), *Automated sports journalism. The AnaFut case study, the bot developed by El Confidencial for writing football match reports*

Rojas Torrijos José Luis, de Santis Andrea (2024), *El periodismo deportivo, terreno de vanguardia para la aplicación de la Inteligencia Artificial in ComunicAI*

Rojas Torrijos Jose Luis, Nolleke Daniel(2023), *Rethinking sport journalism*

Rosenblatt, Frank (1962). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan.*

Rowe David(2007), Sports journlism still the toy department of the news media?

Salaverria Ramon(2019), *Digital journalism: 25 years of research. Review article". El profesional de la información, v. 28, n. 1, e280101.*,. https://doi.org/10.3145/epi.2019.ene.01

Sanchez Gonzales Hada (2022) , *Transformación digital y audiencia. Tendencias y uso de la inteligencia artificial en medios verificadores*

Sánchez H. M. , Sánchez, M. (2018). *Analysis of the functionality and usability of the Politibot's online visualizations. Icono 14, 16 (2), 14-39. doi: 10.7195/ri14. v16i2.1192*

Schapals Aljosha Karim, Porlezza Colin, (2020), *Assistance or Resistance? Evaluating the Intersection of Automated Journalism and Journalistic Role Conceptions*

Seyfodin Elahe, Tabrizi Seyed Javad (2024) , *Effects of Artificial Intelligence on the Future of Journalism*

Shapiro Ivo, Brin Collette, Bedard Brule Isabelle, Mychajlowycz Kasia, (2013), *Verification as a Strategic Ritual How journalists retrospectively describe processes for ensuring accuracy in : https://doi.org/10.1080/17512786.2013.765638*

Shapiro Ivor(2010), *Evaluating journalism, Towards an assessment framework for the practice of journalism in https://doi.org/10.1080/17512780903306571*

Shapiro, S. P. (1990). *Caution, this paper has not been fact checked! A study of fact checking in American magazines. New York, NY: Gannett Center for Media Studies*

Sherwood Merryn, Nicholson Matthew(2013), *Web 2.0 platforms and the work of newspapers sport journalism*

Sherwood Merryn, O'Donnell Penny(2018), *Once a journalist always a journalist?Industry restructure, job loss and professional identity*

Siddharthan Advaith, Nenkova Ani, McKeown, Kathleen R. (2011). *Information Status Distinc-tions and Referring Expressions: An Empirical Study of References to People in NewsSummaries.Computational Linguistics,37(4), 811–842*

Silva Fernando (2022) *Áudio imersivo em narrativas jornalísticas de Realidade Virtual, Aumentada e Estendida. Revista Eco-Pós 25(1):180–197*, https://doi.org/10.29146/ecops.v25i1.27848

Silverman Craig (2016). *"Here's Why Facebook's Trending Algorithm Keeps Promoting Fake News."*

Singer Jane B, (2014) *User-generated visibility: secondary gatekeeping in a shared media space.in New Media and Society*

Singhal Amit (2011) *Some Thoughts on Personalization*.

Sivek, S. C., & Bloyd-Peshkin, S. (2018). *Where do facts matter? Journalism Practice, 12(4), 400–421*

Smiley Charese, Frank Schilder, Vassilis Plachouras, and Jochen L Leidner. (2017). *Say the right thing right: Ethics issues in natural language generation systems. EACL 2017*

Smolianov Peter, Dwight Zakus, Gallo Joseph, (2014) *Sport Development in the United States: High Performance and Mass Participation*

Somalvico Marco,(1987) *Intelligenza Artificiale*

Stark Jennifer, Diakopoulos Nicholas(2016), *Towards editorial transparency in computational journalism. Computation + Journalism Symposium.*

Steensen Sten(2011), *Online journalism and the promises of new technology, A critical review amd look ahead*

Susan Jacobsen, (2012),*Transcoding the news: An investigation into multimedia journalism published on nytimes.com 2000-2008. New media & society*

Tandoc Edson(2014), *Journalism is twerking? How web analytics is changing the process of gatekeeping*

Tasner Michael (2010) *Marketing in the Moment: The Practical Guide to Using Web 3.0 to Reach your Customers First. Upper Saddle River*

Theune, M., Klabbers, E., de Pijper, J.-R., Krahmer, E., & Odijk, J. (2001). *From data tospeech: a general approach.Natural Language Engineering,7(1), 47–86*

Thorne James, Vlachos Andreas (2018), *Automated Fact Checking: Task formulations, methods and future directions*

Thurman Neil, Lewis SC and Kunert J (2019) *Algorithms, automation, and news. Digital Journalism 7(8): 980–992.*

Thurman, N., Dörr, K. & Kunert, J. (2017). *When Reporters Get Hands-on with Robo-Writing. Digital Journalism, 5(10), 1240-1259. https://www.doi.org/10.1080/21670811.2017.12898*

Tramuntolo Rodrigo Navarro, Marinho Mezzadri Fernando and Marcelo Moraes e Silva(2021), *The genesis of the sport for all campaign in Brazil as seen through the Journal dos Sport in the 1970s,*
Tsay, Ming-Yueh (2009) "*Citation Analysis of Ted Nelson's Works and His Influence on Hypertext Concept", Scientometrics 79(3), pp. 451-472*

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

van der Lee Chris, Verduijn Bart, Krahmer Emiel, et al. (2018) *Evaluating the text quality, human likeness and tailoring component of PASS: a Dutch data-to-text system for soccer. In: Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe Community Convention Center, Santa Fe, USA, 20-26 Aug 2018, 962–972*

Vecsey G (1989). *A Year in the Sun. New York: Crown*

Veglis Andreas, Maniou Theodora A. (2019). *"Chatbots on the rise: A new narrative in journalism". Studies in media and communication, v. 7, n. 1. https://doi.org/10.11114/smc.v7i1.3986*

Wahl, A. (1989) *Les Archives du football. Sport et société en France (1880–1980), Paris*

Waltz David (1975). *Understanding line drawings of scenes with shadows. In Winston, P. H. (Ed.), The Psychology of Computer Vision. McGraw-Hill*

Wanless Liz, Seifried Chad, Bouchet Adrien, Valeant Annie , Naraine Michael L.,(2022) *The diffusion of natural language processing in professional sport, https://doi.org/10.1080/14413523.2021.1968174*

Watson Mark (2009) *Scripting Intelligence: Web 3.0 Information, Gathering and Processing (Apress Series: Expert's Voice in Open Source). New York: Springer Publishing. Zimmer M (2008) The externalities of Search 2.0: The emerging privacy threats when the drive for the perfect search engine meets Web 2.0. First Monday 13(3). Available at: http://firstmonday.org/htbin/ cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2136*

Web 2.0 platforms and the work of newspaper sport journalist, Merry Sherwood, Matthew Nicholson, 2012
Westlund Oscar (2013). *"Mobile news: A review and model of journalism in an age of mobile media". Digital journalism, v. 1, n. 1, pp. 6-26. https://doi.org/10.1080/21670811.2012.740273*

White Michael, Clark Robert A. J., Moore Joanna D. (2010). *Generating tailored, comparativedescriptions with contextually appropriate intonation.Computational Linguistics,36(2), 159–201*

Whittaker Robert, (2019) *Tech Giants, Artificial Intelligence and the Future of Journalism*

Widrow Bernard (1962). *Generalization and information storage in networks of ADALINE "neurons". In Yovits, M. C., Jacobi, G. T., and Goldstein, G. D. (Eds.), Self-Organizing Systems. Spartan*

Widrow Bernard and Hoff, M. E. (1960). *Adaptive switching circuits. In IRE WESCON Convention Record.*

Wright Alex (2015). *Algorithmic authors. Communications of the ACM, 58(11):12–14*

Zhang Caixa, (2023), *On the Ethical Dilemma and Countermeasures of Algorithmic Recommended News*

Ziegler, C. (2012). *Pro sports are finally ready for a gay athlete. Outsports. retrieved from* *http://www.outsports.com*