

The impact of digital technologies and big data on the creation of a digital audience and the development of data journalism

Aigul Niyazgulova^{1,†}, Nurzhan Duzbayev^{1,†} and Saule Ashenova^{1,†}

¹ International Information Technology University, 34/1 Manas St., Almaty, 050000, Kazakhstan

Abstract

In the new digital age, humanity is experiencing profound changes in the perception, processing and use of information. The observations indicate the formation of a new social structure, viewed through the prism of a digital society and its close interaction with new technologies. The article examines the current state of the deep impact of digitalization on society and its subjects from the perspective of virtualization of mass media and the active development of computerization and informatization of society. The process of digitalization is considered as an intermediary of pre-existing changes, and as a causal factor of emerging changes in the media environment. The object of research in this case is the interaction between the subjects of society, a flexible area of communicative relations entering the digital space, and the technical process itself, which resulted in transformations in journalism. The combination of the specifics of data science, the emergence of a new digital personality and the possibilities of using new information technologies made it possible to modify and develop such an urgent field as data journalism.

Keywords

Digitalization, new technologies, digital society, audience, digital identity, databases, data journalism

1. Introduction

Digitalization today is not just a process of transformation in the media sphere, which has led to concepts such as convergence, new media and web journalism, but also the main trend in the development of the media market. Modern journalism is defined in the creative sector of the economy, which is gradually introducing advanced concepts such as robotics, artificial intelligence, and augmented reality into the production of journalistic content [1]. This has become a natural course of modern development of communication processes, which were created using special technologies. This refers to the use of information technology, the active use of promotion technologies in the Internet space and social networks. Information is changing, audience perception is changing, thinking is changing, and specific approaches to content are changing.

International studies conducted since the beginning of 2023 have revealed more than five billion (5.16) active Internet users and more than four billion active social media users worldwide, respectively, this is 64.4% of all inhabitants of the earth[2]. For example, in 2021, the figures were 4.66 billion and 62.5%. Despite the fact that digital technologies have long become an integral part of people's lives, user growth is still actively continuing. The social isolation and quarantine caused by the COVID-19 pandemic at one time further strengthened and expanded the penetration of digitalization around the world, the use of Internet services increased from 40% to 100% since the period preceding the quarantine.

These figures underline that the impact of digitalization on humanity is deeper than it seems. As a result of the extensive, progressive and deepening influence of the digitalization process on the world

DTESI 2024: 9th International Conference on Digital Technologies in Education, Science and Industry, October 16–17, 2024, Almaty, Kazakhstan

* Corresponding author.

† These authors contributed equally.

✉ a.nijazgulova@iitu.edu.kz (A. Niyazgulova); saule_ashenova@mail.ru (S. Ashenova); n.duzbayev@iitu.edu.kz (N. Duzbayev)

ORCID ID 0000-0002-7989-9463 (N. Duzbayev); 0000-0002-5301-5340 (S. Ashenova)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

and people, developing phenomena and growing evidence point to the possible emergence of a new form of the individual in the era of digitalization, which can be called a digital personality, which is transformed into a new format for a modern audience. It is very important for researchers and practitioners of the media sphere to have an idea of what mechanisms contributed to this and how the mass audience will develop and change in the future.

To a large extent, this affects new formats of media genres and the media market itself, which is rapidly changing under the influence of data arrays. Such areas as data journalism, augmented reality and artificial intelligence do not simply interact with a modern audience, but demonstrate a double-edged interest, when the personality changes under the influence of new media structures and at the same time has a strong influence on them. Research in recent years, presented in highly rated journals and raising issues of interaction between the audience and new media, touch on a wide variety of interests and demonstrate the transition from studying the basic categories of media activity [3] in the context of the digital environment to the formation of principles of audience consciousness depending on the specifics of genres, channels and technologies used by the media. Research on the specifics of professional journalists' work on online platforms has begun to move into a narrower specialization, for example, Chang Sup Park and Barbara K. Kaye [4] brought the issue of consuming content presented in social networks to the forefront and asked the question of the impact of large amounts of data on this audience, and in 2024, researchers have already addressed the issues of conceptualizing the significance of the media when audience interest in various channels of information transmission decreases or increases [5].

The diversity of issues regarding the formation of the media field and the place of digital technologies in it determined the choice of the abstraction method as a theoretical level of knowledge. Moving away from insignificant connections and relationships that do not directly appeal to the topic of the study, the authors turned to recording the most important elements in the creation and dissemination of information as an object of interest for the mass audience of the digital era, focusing on such a format as data journalism. By studying the specifics of creating content that is in demand today, the authors drew parallels between the theoretical basis for the successful development of the media landscape and the existing information realities that have changed both the audience and the structure of working with information.

There is already increasing evidence that digital technologies have a significant impact on brain function, mind, and human behavior at a deep neurobiological level[6]. In the process of digitalization, there were various, complex ways and levels of interaction between various changes in the world and human changes from the industrial era to the era of digitalization and ways to study and adapt them. Changing people adapt, survive and transform in an ever-changing world. At a more fundamental level, in time and space, it will be the direct and indirect interaction of various elements of digital technologies with the world, society and culture, as well as with pre-existing and changing human concepts, behavior, ideologies and identities of individual subjects.

A special place in such a study should be occupied by methods of a psychosocial and cultural-technical approach, which have the opportunity to consider changes from the industrial era to the era of digitalization in the context of potential personal and interpersonal transformation, personality concepts, fundamental experience, mental abilities, interaction and communication, ways of thinking and social interaction, behavior, mental health, as well as in consciousness and brain activity caused by the process of digitalization, at different levels through various mechanisms. Mechanisms may include self-identification and self-education, cultural assimilation, adaptation, formation, learning, information processing, social theories such as social learning and the effect of deindividuation, online phenomena such as the effect of online disinhibition, behavioral theories such as conditioning, psychological processes such as internationalization and design, various defense mechanisms as well as various neurobiological and neuropsychological processes.

Foreign researchers such as Martin Heidegger, Thorstein Veblen in the last century, starting in the twenties of the XX century, called for paying attention primarily to the essence of the concept of "technology" and the moments of perception of this concept by society, noting that there are two examples of the impact of new technologies on humanity, including "technophobia", Instrumentalism

is also a negative factor, as is the ability of harmonious coexistence and joint development. Neil Postman, already in the twenty-first century, developing this idea for the media sphere, called the existence of these two positions a "Faust deal", when it is impossible to say what exactly becomes more important for the audience and its transformation [7]. But in any case, whatever the consequences, positive or negative, important changes occur not only in the usual information field, but also in cognitive processes and information perception processes. And in this regard, there are changes in communication processes, giving rise to new formats of journalism, closely related to new technological factors.

Thus, in order to systematize the empirical material together with theoretical information, the classification method was also used as a form of controlled knowledge. It allowed to identify narrower concepts and pay attention to the necessary conceptual apparatus of scientific theoretical discourse concerning the topic of modification of the digital personality of the audience, and to study a separate format of data journalism, as well as to touch upon such issues as the use of visualization elements and artificial intelligence in it. The analysis method allowed to consider the subject of study in the totality of its components, which helped researchers to avoid limitations in matters of media activities only as an information resource and to structure the components of successful content in conjunction with the latest information technologies.

2. Research. How new technologies and new forms of communication affect the digitalization of journalism

In the process of digitalization of the information space, there were various, complex ways and levels of interaction between changes in the world and society; from the industrial era to the era of digitalization; and, accordingly, the ways of studying and adapting them varied. The study of social and empirical phenomena that are important for the transformation of the modern media field, in which the perception of video information occupies one of the dominant positions, has determined such methods of cognition as empirical research, in particular comparison, measurement and observation; the method of analysis and synthesis; a substantial method of research and forecasting. These methods allowed us to study the main trends and the impact of new communication formats, expressed in changes in the preparation of relevant content, on the structure of the modern media market.

Digitalization has a progressive and extensive impact on all aspects of daily life, including information processing, communications, infrastructure, logistics, finance and commerce, industry, economics, education, healthcare and entertainment. Along with the previously existing physical world, various types of social networks, virtual communities and realities are emerging and spreading, the world is rapidly filling the digital space for algorithms for the interaction of virtual objects with each other and with the external real environment.

The modern audience, immersed in the ultra-fast exchange of information, does not have the opportunity to adapt to it, to find a rational explanation for the change of formats, especially if this audience, due to any peculiarities, age, social, does not have the psychological and communicative skills to select and adequately perceive the information flow. The Internet is a kind of information base that accumulates, stores and reproduces an unlimited amount of information and provides quick access to a significant audience [8]. It is also necessary to note the important role in the development of civil society, expanding the possibility of social control of power through network technologies.

The information space has the opportunity to explain to its audience the social order and its phenomena, as well as to mystify, reduce the level of rationalization and reality of the surrounding world. On the one hand, while people constantly and closely interact with the world, culture and society in the era of digitalization through various processes, elements of our personality are influenced and shaped directly and indirectly, consciously and unconsciously by various components of digital technologies in the digital world. On the other hand, many were forced to develop their "digital identity" in the era of digitalization. In the beginning, the formation of a "digital persona" remained a conscious, voluntary, deliberate and purposeful process by turning on and off online

access, transactions and communications. Soon, as our awareness, volunteerism, vigilance, control and protection began to erode as a result of increasingly inevitable, intense activity in the digital world through ubiquitous 24/7 portals such as smartphones, our "digital identity" can increasingly suppress our natural, real identity in the digital world. Gradually, our original essence is increasingly extracted, expressed and presented in the digital world in conscious and unconscious ways due to the rapid growth of our "digital personality" in the digital world. Ultimately, this can lead to an increase in the image of a part of our true identity in a "digital image".

This leads us to say that the use of technical means in public life has led to the creation of a specific type of mass communication that arises between the subjects of society and the information space and thus to the creation of a new type of mass audience. If during the 20th century the main means of mass communication were traditional communication channels: printed publications and electronic means of communication, then the beginning of the new century was marked by the active development of Internet technologies that qualitatively changed the nature of politics, giving it virtual characteristics. Thanks to new opportunities, mass communication has increased the speed of information dissemination, reaching the consumer in the shortest possible time. The large amount of information transmitted, the brevity and memorable imagery of the messages increased the intensity of the impact on the audience. The blogosphere, electronic media, social networks, portals, forums and websites, and online communications are constantly being updated and have an increasingly significant impact on public opinion, shaping human consciousness and influencing personality through information in completely different forms and for different purposes, which can lead to different consequences.

The process of digitalization acts both as a mediator of pre-existing changes and as a causal factor of emerging changes in the world and humanity [9]. The object of research in this case is the interaction between the subjects of society that form the audience and the flexible area of communicative relations entering the digital space. The presence of audiovisual information in this area, previously provided by traditional media, changes the principles of journalism and at the same time puts it in front of the fact of loss of exclusivity and quality, which is unacceptable in the professional media sphere. An analysis of the possibilities of professional journalism and its presence on digital platforms, taking into account their capabilities and impact on a mass audience, is able to identify a beneficial collaboration of new media approaches to creating in-demand content.

The use of new technologies is indeed of great importance for changing both the present and the future of modern journalism. Many experts expressed their concerns that modern man-made processes affecting communication processes, which in turn generate a new type of audience, will have a negative impact on the media space. This is due, firstly, to the fact that the expansion of technical capabilities changes the structure of working with information, speeding it up and allowing technology to outpace the human brain. Secondly, since information is today the most consumed and demanded product of mankind, it forms the types of modern mass audience. Thanks to the development of information and communication technologies, the rapid improvement of technical means, the huge arena of network media, online forums and social networks has become an ideal platform for the work of a wide variety of media with a billion Internet users. The platform for creating and promoting media risks is obvious, since it is not only information technology that contributes to the development of this phenomenon. Developed psychological and psycholinguistic means occupy a special place in the technologies of working with the audience, in their collaboration with technical means, with the peculiarities of the functioning of the Internet, methods of presenting information are increasingly expanding and becoming more sophisticated, and the audience itself, increasingly immersed in high-speed information noise, becomes more sensitive and receptive to the proposed media products.

Of course, the collaboration of the achievements of the technological revolution and traditional media can be quite successful. Former reporter Jeff Kofman has developed a special algorithm that can decode recorded speech – Trint, a useful invention that greatly facilitates the work of a journalist. The DeepL machine translation service, based on the use of neural networks, has become one of the ten best examples of AI work in journalism[10]. Artificial intelligence working at the Washington

Post published 850 articles in 2016 alone using its Heliograph system, the publication informed the Associated Press that they saved 20% of time with its help and increased the volume of materials 10 times[11]. The transition to robotic reporters allows you to do quite well the work that is required for financial or sports reporting, thanks to which robotic journalism is already demonstrating its readiness to work with data journalism, which is associated with big data.

However, despite the fact that robot journalists can write quickly and in large quantities, they usually do not have the opportunity to conduct a deep and systematic analysis of the topic. Their activity only replaces a certain part of the journalist's activity, thereby helping to achieve the greatest efficiency in their work and shows an example of positive interaction in specialized genre frameworks. It is assumed that journalists who are released from routine work will have time for more significant and detailed study of topics that are of the greatest interest to the audience.

There are also theoretical and methodological problems with the media space and the use of AI in it. The media has its own algorithms for how to collect and how to use data. It is obvious that it is impossible to collect and publish data without taking into account the legal foundations of journalism, therefore, in any case, a robot journalist will need programmers and employees who will actively monitor and program his work[12]. At the same time, there is a risk of fixing various operations in the creator's brain. A few years ago, rumors were discussed that the racist idea of a South American inventor was stuck in the memory of his robot. This prompted the correspondents to reflect on their responsibility to society. AI is still not able to fully distinguish truth from false information and fight misinformation. Although OpenAI was able to create artificial intelligence in order for it to detect fake news, it turned out that the algorithms coped with their task, but worse than humans. And from a creative point of view, robots lack creativity. Smart machines can exclusively imitate the style of people, but not create something unique. The media has long used AI to record sports news and crime stories, but jokes created by robots and the novels they write still do not arouse active interest.

Nevertheless, experts are confident that in the future, using the full range of possibilities of new technologies, these difficulties will be resolved, since, according to scientists and experts, all this facilitates the work of journalists. The emerging problems and questions, in turn, are designed to emphasize the need to solve them for more successful implementation of these trends in journalistic activity in the context of competent and legitimate interaction. Augmented reality has also become an important structure in the work of artificial intelligence. We are already seeing AR technology being used in news releases around the world. Just look at the examples of Croatian television, where mockups with statistics of election campaigns were built on one of the channels. AR is also used to demonstrate various natural phenomena in weather forecasting. In one of the issues, the presenter was able to demonstrate how dangerous different water levels can be during a flood, which allowed us to create a clear picture of what is happening.

This allows us to conclude that the development of journalism working with digital formats has excellent prospects, which are already being used now and at the same time has some drawbacks that should be closely paid attention to in order to be able to use this trend more effectively, which has become a reality of the modern digital world. These examples highlight the power and possibilities of using modern information technologies, and given that many media around the world are already using both augmented reality and artificial intelligence, it can be assumed that in the near future these trends will be able to create the most effective and sustainable journalism of the digital age, becoming a real and successful assistant to the journalistic craft and opening up new horizons of the profession.

At the same time, a new and very important question arises. How do modern information technologies influence new audience formats? Since, using the conclusion drawn and taking into account the improving techniques of collecting and distributing information, it is impossible to ignore the fact that the audience, which has already become accustomed to robots, artificial intelligence, massive data formats, and virtual reality, has undoubtedly changed, both from a technical and psychological point of view. The audience has become both an object and a

subject of ongoing information and communication relations. Let's take a closer look at how the formation of a digital personality takes place and which areas of digital journalism will be closer to it.

As the new digital world penetrates our daily lives, mainly through increasingly dominant digital relationships, as well as the Internet of things and social media, fundamental changes are already taking place in our personal and interpersonal spheres. As the development and spread of the metaverse occurs at a rapid pace, in addition to the pre-existing physical world, various types of virtual communities are additionally formed, leading to various virtual realities, mixed reality and augmented reality. The development of a continuous environment of digital twins leads to the transition from the digital generation to the final coexistence of physical and virtual reality, called surreality. Some researchers have already predicted the evolution towards a digital triplet, which is assumed to be a form of intellectual world activity involving an active level of human participation. The progressive and unprecedented formation of various digital media, as well as virtual communities and realities with various inherent systems, rules, values and causal relationships, potentially have a profound impact on the actors involved, especially as boundaries become increasingly blurred and displacement becomes more frequent and intense.

Since digitalization is a continuous global process in time and space, it will take time to see its potentially unnoticeable impact on people in the process of digitalization. In particular, it can be expected that the impact of digitalization on young people, up to the levels of self, mind and brain, will be more inevitable and profound, given its more primary, dominant and direct impact on them as digital derivatives[13]. To explore such a complex, subtle and at the same time profound topic, several disciplines and traditional fields of study should be used. The study of the "digitized archetype" should focus on fixing, differentiating and analyzing changes in personality self-awareness profiles over time with probable subsequent prevention and intervention in order to mitigate any adverse effects on the psyche and consciousness of the individual. Different types of research developments and research methodologies can vary from qualitative research to quantitative research, from cross-sectional research to long-term research, as well as from target concepts to deep-object ones.

Among various aspects of human cognition, the study showed that the mere presence of devices with smartphones can already reduce available cognitive capabilities. Despite the expansion of research, the conclusions about the impact of smartphone technology on cognitive functions are still contradictory and inconclusive. The digital world can cause both acute and sustained changes in attention, memory, and social cognition, possibly manifested in brain changes. While research on digitalization and human cognition is still ongoing for a more definitive conclusion, neuroscientists argue that multitasking switches costs in the brain and psyche, since the evolution of the human brain was originally designed to perform separate tasks. Reading skills and task accuracy have been shown to decrease as people work in multitasking mode with instant messaging and various computer programs. Given the potential long-term impact, the possible adverse effects of digitalization on human cognition will require timely verification.

On an interpersonal level, our personal space has become a new privatized zone in public whenever we hold a smartphone in our hands, creating a paradoxical form of absence of a present subject, subject to individualistic manipulations of "closeness" and "openness" in daily online communication. Unlike the classical proxemic theory of interpersonal distance, new forms of interpersonal distance in the era of digitalization have been theorized in different directions and the new way of social interaction actually allows you to maintain your optimal interpersonal distance and even become hyperpersonal, since computer media can be paradoxically inaccurate for various reasons. In the age of digitalization, we are able to manipulate human interaction like never before, potentially becoming more self-centered, with a greater sense of control and possibly less authentic.

From the ordinary physical world to the digital world, originally conceived as a utopia for sharing, we can access almost any kind of information at our fingertips instantly without careful verification and deliberation. At the same time, such a huge number of digital viruses and Internet memes appear every second all over the world that it becomes difficult to adapt and distinguish between them. With further information overload in order to get spontaneous feedback, we tend to believe information

indiscriminately and quickly draw conclusions without additional fact checking, hastily taking a position and instinctively reacting on the Internet, especially when it is possible due to the status of anonymity, to undergo self-categorization and social grouping, or even go through the process of polarization in extreme situations. At the same time, the phenomenon of "fear of missing out", the prevailing fear that others will gain useful experiences in their absence, as well as the increasingly popular time-limited functional excuses on some social networks such as Instagram and Snapchat, can further perpetuate these new forms of thinking, behavior, forming opinions, attitudes and grouping.

Gradually, the process of digitalization affects our ways of thinking, information processing, social cognition and behavior. This, of course, is fraught with media risks, since the construction of reality is based on certain events in the context of which an information campaign is unfolding. By constructing the facts of social life and building a certain system of meanings, communication processes begin to influence the process of reconstructing and perceiving certain events, taking into account those cognitive capabilities that are closer to the digital audience[14]. By appealing to various sensual, irrational, emotional-volitional components of consciousness, of course, one can get support for moral, spiritual, and social foundations, but in the same way one can mobilize specific approaches in establishing the views of society on the world around them and moral values, which can be modified in accordance with the semantic load carried by virtualized and It is often a disinforming information space.

By creating an atmosphere of free, lively communication, the modern digital space at the same time creates new opportunities for manipulating public opinion, as it has already changed under the flow of new technologies and the development of alternative types of thinking, such as clip art. In addition, in extreme cases, when Internet use becomes excessive and extensive, it can lead to a new form of behavioral problems in people, namely Internet addiction, which is already becoming a modern, global and ambiguous phenomenon. It is also associated with a number of socio-demographic and psychosocial factors and associated symptoms, mental health disorders in adolescents and adults, including problems with personal qualities, mood, behavior and cognitive abilities. Internet addiction can be considered as a form of heterogeneous and complex behavioral problems associated with various activities such as gaming, gambling. This is due to various etiologies and consequences, respectively. According to existing data, there is a link between excessive Internet use and mental health problems such as mood, sleep and behavioral problems, although the trajectories of causation and the underlying mechanisms require further study.

Comprehension, presumptive mechanisms and consequences of a digitized personality. Within the framework of the duality of the processes of increasing digitalization of the modern personality and the representation of its inner essence in a digital hypostasis, when interacting with the digital world, a digitized personality is formed, which is interpreted as a new part of our "I", located between our "I" and our "digital personality" at the junction of the physical and digital worlds. Our digitized personality can be considered as a combination of a part of our personality interacting with the digital world and a part of our "Digital Persona" revealing our "I" in the process of digitization, manifested both in the digital world and in the physical world by analogy with the Freudian personality model consisting of three structures – "Ego", "Superego" and "Id", also moving through our "Conscious", "Subconscious" and "Unconscious".

The main and final question is – how much of our original and natural archetype will remain in us and who will people become when there is an increasingly continuously increasing monopoly of the omnipotent and omnipresent, extremely chaotic digitized archetype and what will humanity, personality and self-awareness of people become?

The hypothesis of the existence of a digitized archetype is motivated by the desire to provide further timely and in-depth study, definition and research to eliminate the unique and profound impact of digitalization on humanity, on individual personality qualities [15]. The relevance of studying the impact of the digitalization process requires a joint initiative to redefine human nature in order to preserve the basic elements of humanity for our future generations in the era of

digitalization and at the same time identify the most advantageous positions for expanding the possibilities of modern journalism.

Since information is an object of interest to the mass audience of the digital age, it is necessary to consider the features of creating content that is in demand today from a digital personality and using the achievements of the digital age. New media, interpreting multigenerationalism in their own way, have already made certain adjustments to the structure of information consumption, changing the appearance of a modern audience. In the era of globalization, this is becoming an important tool, because due to hypertextuality, convergence and interactivity, the format of communication is also changing and new functions are emerging, and these processes are already stable and often spontaneous. One of the results of such processes was the emergence and development of Data journalism.

3. The impact of data science on the development of data journalism

Digitalization trends, which have the opportunity to influence new ways of perception and processing of information by the audience, which in turn has changed with the advent of new technologies, have also influenced the specifics of the activities of modern journalists, who quickly mastered the territory of dynamic and progressive interchange and dissemination of information using new means of its creation and transmission. Having firmly established itself in the functional activity of the media segment, the Internet with its technical capabilities entered the structure of the QMS and now represents the largest field for the media, giving rise to the virtual world of information spaces. This interaction has a very plastic form and aims, through fundamental approaches to its study, to reveal the boundaries of the participation of the local media market in the global information space and information trends that are emerging in new conditions and have undoubted opportunities to influence economic growth and the social sphere.

One of these trends is working with databases. In a technical sense, this is the ability to store and manage information, and directly access it. Today, databases play an important role in all areas of the economy. Of course, with the change of modern media, the possibilities of working with data have also changed. In itself, obtaining information and working with it, its interpretation, is the basis of the activities of journalists, but it is the new ways that have come into their activities with changes in the digital environment and the emergence of a new digital personality that have allowed us to expand the methods and functions of presenting information so that it can interest precisely the modern audience formed in the man-made age.

The concept of structured systems, which are the basis for databases, is interpreted in the media as the basis of information, protected from manipulation of facts. Databases are stored as a characteristic of a particular topic and are used by journalists to prove or disprove a hypothesis. It should be noted here that information manipulation as a way of working with the audience and public opinion has also undergone significant changes, while remaining one of the most complex tools that easily adapts to modern realities. If earlier it was based only on context, now neural networks, the capabilities of the Internet and the psychology of social networks easily come into play, which have become a very convenient place for the dissemination of fake news and disinformation. Large volumes of data, the flow of information noise and the speed of information absorption create active media risks that can be divided into two categories.

1. Quantitative: big data offers a huge amount of information stock that cannot be fully verified;
2. Qualitative, or technological: new technologies have led to an increase in the means of collecting and storing data, as well as the number of news producers and channels for their transmission.

Since in the context of the information society, the Internet has become the most important and accessible source of knowledge, no special effort is required to access this huge repository. Comparing the data of various studies, we can safely draw conclusions about the audience's

commitment to digital information. <https://k-research.kz/research/internet> collect data on the presence of users of the Kazakhstan segment in networks, the UN provides information on changes in Internet use data in the world <https://news.un.org/ru/story/2023/11/1447172>, the TAdviser portal publishes annual studies of the global Internet audience and trends in its development, WEB-canape collects statistics on the Internet and social networks https://www.web-canape.ru/business/statistika-interneta-i-socsetej-na-2022-god-cifry-i-trendy-v-mire-i-v-rossii/?utm_referrer=https%3a%2f%2fwww.google.com%2f, Global Digital annually publishes a global report on the main figures of the Internet and social networks <https://www.sostav.ru/publication/we-are-social-i-hootsuite-52472.html>. Global studies show that more than 90% of users consider the Internet as a source of information. Social networks have become the main source of news for 47% of people. As a result, the Internet, especially social networks, have become a convenient and safe place for disseminating false information and, as a result, manipulating facts and opinions. Access to the information market and content creation through social networks has been significantly facilitated recently thanks to various applications, new technologies and programs such as Adobe Premiere Pro, Adobe Photoshop, Canva.com and others. This is all the more important because for the modern type of audience, visualization is of paramount importance.

However, we can say that journalism has more opportunities to combine technological and creative approaches in combating manipulation risks. Trends in the development of software and artificial intelligence for detecting counterfeits are complemented by professional journalistic vision.

There are three positions that are important for working with information:

1. Detecting digital integrity. Each camera has its own characteristic signature. For example, let's say you take a picture with a Nikon D780. First, this information can be found in the metadata. Then we add several photos taken with a Canon EOS 5D Mark III. Artificial intelligence can determine from the photo (pixel features, compression artifacts, etc.) that these two parts were taken at different times and combined using a graphics program such as Photoshop. This is evidence of manipulation.
2. Determining physical integrity. When manipulating an image, it may seem that the laws of physics or the laws of compositional construction have been violated when working with a frame. For example, nearby objects may cast shadows under different light sources, or skin tones may look unnatural.
3. Determining semantic integrity. The actual situation with the image must correspond to common sense and verifiable facts.

However, the future of fighting fakes looks twofold. The technology for its use will turn into a game of cat and mouse as software learns to use it more reliably. There is no direct ban on the use of programs on the Internet, it is simply technology that can be used for different purposes, although Facebook, Twitter, TikTok and others often come up with similar initiatives.

The importance of the interests of the modern digital audience in visual information is so great that not only media structures, but also government corporations related to security are paying attention to the problem of fakes. For example, the Defense Advanced Research Projects Agency (DARPA) is seriously working on developing a solution to detect and remove fake photos and videos. To understand the seriousness of this problem, we note that the agency is also working on protection against cyber and missile attacks.

DARPA program manager Matt Turek told participants of the DARPA Colloquium on Artificial Intelligence in Washington in March that in terms of what is needed to create a deepfake, "a high-end gaming computer is sufficient. Depending on the amount of data and the processor performance, training can take several hours or days" [16]. In communications with leading media platforms, DARPA sets an intermediate goal - to achieve such a level of detection and automatic filtering of fakes that such a "toy" is no longer available to people with mediocre computers without professional skills in machine learning and creating neural networks. Creating such an entry threshold will update the capabilities of professional journalism, preserving its reputation as a high-quality social institution, and not an entertainment layer. The need for a high level of awareness of important and

pressing social issues that are needed for democracy to function in society is combined with the need to disseminate technical knowledge in the journalistic environment. The data journalism format fully meets these qualities.

Information collected in the form of statistics and confirmed by a reliable source, such as an official government structure or verified statistical data, inspires confidence in the reader and allows you to raise the authority of the publication. It is important for a journalist to be able to work with such data and interpret it correctly. This, in turn, is already becoming the basis for high-quality analytical journalism and the growth of reputational capital. Such world-famous publications as Berliner Morgenpost, New York Times, Washington Post, Guardian, Financial Times have introduced the specifics of data journalism into their activities and regularly develop data projects on the relevant topic.

Modern data analysis takes place using software that was created to do all the hard work for us. We cannot work with data until we convert it into a format that is understandable to a computer. The data needs to be organized, given the appearance of rows, columns and cells. Most often, ordinary people do not use the data, on purpose or by accident, precisely because it is provided in a format that cannot be immediately used on a computer. Analysis, narration, and visualization all depend on whether a computer program can read our data. Most often, data formats are used in journalism when computer software recognizes explicit structures. Most often, this data is provided in the form of columns and rows that organize and structure individual data elements. First of all, Excel and CSV formats are used:

- Excel files (XLS): Data is saved in Microsoft Excel spreadsheets;
- Comma-separated values (CSV): a text file where each new entry is separated by commas.

These formats are usually better suited for analysis in journalism, and they can be easily worked with, for example, in Excel. When searching for data, if a journalist managed to find data in Excel or CSV format, this means that you do not have to spend a lot of time cleaning and formatting them. The tools for working with these formats are the Google spreadsheet and Microsoft Excel, which are familiar to almost everyone. The data can be in PDF, in scanned images, for example, reports on the websites of various organizations that are freely available, and in unstructured formats that require special computer programs. Recently, most countries have opened up access to statistical data. Special data portals have been created. They are an invaluable source of information and it is important to understand how to use the various interfaces of these portals to get the data you need. International, government, civil societies, University databases are all huge data sources. However, they all have their own interfaces, which are slightly different from each other. You need to explore them to understand how to navigate them. Some data, especially in large databases, is stored in packages that can be converted into websites. Most often, they require conversion to CSV or Excel for further use with spreadsheet software. But in any case, the specifics of creating a data project imply working with confirmed information, which the journalist analyzes in accordance with the topic and hypothesis. And he works in a team where a programmer and a designer will definitely be present. The very choice of data will depend on the journalist.

It is important to choose which database you want to find. On the World Bank data portal, you can select health-only data, education-only data, or development indicators to search for. On geographical issues. You can compare how much your geographical space differs from others. You can compare neighboring cities, states, countries, regions with similar levels of economic development or population. The data allows you to select indicators. Databases often give you the opportunity to tick boxes to determine which metrics need to be compared. It is best to choose a wide range, look for interesting trends, and later narrow down the search range. You can choose a time period. The longer the time interval, the more likely it is to find enough data to identify trends. In many cases, data was collected in different countries and in different years, so it's better to start with a wide range and then narrow down the time period. By this time, you will know exactly what you

are interested in. Databases often consist of tables, maps, or visualizations. These can be useful tools, as journalists themselves are interested in uploading data in CSV or Excel format.

Processing such a flow of information leads to really impressive results. The new possibilities of the digital space have become a kind of challenge in the struggle for the consumer and have set new conditions for the media – content should attract an audience with diversity, exclusivity and quality. Foreign media actively use and disseminate this experience.

Foreign empirical experience.

It is difficult to imagine that the audience could independently find out, for example, which of the US congressmen was a slave owner, as the Washington Post did [17], analyzing a huge layer of documents and creating a historical database of slave-owning politicians from the end of the XVIII to the beginning of the XX century: <https://www.washingtonpost.com/history/interactive/2022/congress-slaveowners-names-list/>. It is unlikely that a modern audience, which is used to absorbing information in separate bright pieces, would have engaged in such an analysis, journalists did it for her, creating their own project based on a database and arousing the keenest interest from their subscribers.

The use of virtual reality and 3D technologies allows teams of data journalists to visualize information that is better perceived by a modern digital audience. When Berliner Morgenpost journalists created their project on climate change, they used a 3D model of the globe, where they demonstrated which areas of the earth could eventually be affected by this phenomenon: <https://awards.journalists.org/entries/mapping-where-the-earth-will-become-uninhabitable/>. This project brought the newspaper the prestigious Online Journalism Awards for visualization and innovation in online journalism [18].

It is significant that technologies can allow data journalism not only to raise acute social issues, but also to fight for the audience's attention using only visual techniques, excluding text, which would seem unacceptable for the media, but can be accepted as a new online genre. This was demonstrated by journalists from The New York Times, creating the Front Row to Fashion Week project <https://www.nytimes.com/newsgraphics/2014/02/14/fashion-week-editors-picks/index.html> [19].

The project, consisting of multimedia visualization with a brief information component, demonstrated the versatility of the data genre, where information interprets data that has become available to journalists not in the social issue, but in the entertainment industry. The very fact that data reports can have a flexible structure brings to the forefront the importance of the ability not only to find data that is interesting to the audience, but also to process it in accordance with its requests.

In fact, the origins of data journalism go back to the requirements of analytics as high-quality content. The use of graphical data was used as early as the 19th century, when a statistical database on demography, economics, medicine and education began to be formed. The famous maps of Jon Snow and Florence Nightingale, which made it possible to identify the source of danger during the cholera epidemic in the first case and prove the danger of unsanitary conditions during the Crimean War in the second, saved the lives of thousands of people. Data journalism itself as a phenomenon begins its countdown in 1821, when the Guardian published the first study that not only used infographics, but based on database analysis, which revealed discrepancies with official statistics on the number of children enrolled in schools [20].

This has really led to the development of data journalism, combining analytics and numbers. However, it was only by the middle of the 20th century that the concept of computer-assisted reporting (CAR) appeared, when the potential of computers capable of processing statistical data began to be used. The emergence of massive information that allowed the creation of open databases, the personalization of computers, the expansion of journalistic tools and the formation of a new type of digital audience eventually led to the fact that the specifics of CAR gave way to modern data journalism, which has become practically the flagship of digital journalism.

The Guardian newspaper introduced a specialized blog dedicated to the topic of working with the data that it used in its activities, and the equally well-known Los Angeles Times created a long-term data analysis project Data Desk[21].

In addition, this enables the media to consolidate with independent agencies and specialists, thanks to whom journalists have the opportunity to use already processed data sets and carry out truly large-scale projects. For example, Sam Leon and Louis Goddard, who, in their words, “draw on more than 15 years of experience in campaigning and journalism” and arm the audience with “stories that change the course of public opinion and compelling statistics that convince politicians of the need for change”[22].

In addition to the fact that data genres are multifunctional and easily adaptable to different audience demands, as we could see earlier in the examples given, their flexibility and plasticity are also emphasized by the variety of methods of data processing necessary to prove or disprove a put forward hypothesis capable of influencing public opinion or public interests.

Let's compare two data reports:

1. Poverty Does Not = Crime, <https://www.indiaspend.com/poverty-does-not-crime-94890>;
2. Made in France, <https://made-in-france.disclose.ngo/en/chapter/yemen-papers/>.

In the first article, the problem of crime growth is analyzed through crime statistics and direct comparative analysis of social stratification. This approach makes it possible to prove, using open data analysis, that social inequality between poor and rich areas does not affect criminal inclinations or impact people's moral values. Statistics have revealed that the crime rate in poor areas is practically no different in percentage terms from its level in rich areas. This is direct data processing using analytical approaches inherent in Big Data.

Characteristics of the first material:

Velocity – the speed of accumulation of data required for such analysis was insignificant, but their processing required the use of Excel

Variety – the diversity of data in this case is structured, so the visualization is maintained within the framework of infographics and diagrams. The unstructured data flow in this case was not of great importance.

Veracity – the reliability of the data set and its analysis was very high, since all information was based on statistics.

Value – the significance of the information was high, while perception and analysis were adapted for easy understanding.

The second material shows a completely different approach. The second article takes a completely different approach. The political investigation and the hidden accusations against the authorities supplying weapons to unstable countries, while official sources spoke of neutrality and unwillingness to participate in armed conflicts, required data journalists to perform deep analysis and painstaking work with both open and closed data, and to process large amounts of it.

Characteristics of the second material:

Velocity – the speed of accumulation of data required for such analysis was very high, their processing required the use of special programs for visualization and the participation of web designers.

Variety – the diversity of data in this case is partially structured, and streaming information, such as text material supported by photos and videos, is considered unstructured, which determined the genre of the project as a longread. An unstructured data stream in this case becomes important for a more complete delivery of information to the audience.

Veracity – the reliability of the data set and its analysis, as in the first example, is very high, in combination with the first two characteristics, this demonstrates the multitasking of the material and at the same time its high relevance for a modern type of audience.

Value – the significance of the information is also high, and the perception and analysis are adapted for a deeper understanding and interest.

The analysis carried out clearly reflects the flexibility and plasticity of the data journalism format and its undoubted connection with such a direction as working with big data. These cases were offered for comparison to first- and fourth-year students majoring in journalism. The first-year

students were more interested in the first material, and the fourth-year students were more interested in the second.

This can serve as confirmation that the combination of work with data from the point of view of interaction of creative and technogenic approaches influences the formation of value attitudes of the modern audience and allows finding a balance between possible digital and media risks, between the interests of the digital personality and stable social processes. The interaction of information technologies and media development is more susceptible to positive progressive movement with competent development of prospects for the use of large data formats.

The leaders of modern data journalism are well aware that materials that can cause a public outcry are guided by the ability of a digital audience to absorb information very quickly and distinguish it from others due to the so-called rapid exclusivity, but they also understand that in order to carry their reputation high, the source of information must be official open data or a set of data from various resources that a modern journalist needs to be able to work with.

Kazakhstani journalism cannot boast of such global examples presented in world practice. Kazakhstan's publications mainly present analytical materials based on statistical data and use infographics. High-quality infographics are very important, but at the same time, most of the information should be directly taken up by an analysis of the topic, structured from a common data set. And then, based on the identified interests, these data are interpreted into forecasts that may be of interest to the reader. More statistics and infographics can be found in online publications that combine visual content with text, but most often it is found in the headings "Analytics" and "Statistics".

Most often, these sections are used by such publications as Zakon.kz, Kazinform (inform.kz), Informburo.kz, Tengri news (tengrinews.kz). Analytics based on data is found in publications of Forbes, Delovoy Kazakhstan (dknews.kz). There are no specialized subheadings devoted exclusively to data projects using the full range of information technologies in the Kazakh media. Undoubtedly, interest in numbers as information is a phenomenon peculiar to a digital audience, because numbers make up news that can attract attention. This explains the appearance and active use of infographics and maps visualized on the pages of online publications, which we can often see in publications. However, data journalism poses other challenges. Having received expanded opportunities to work with the audience thanks to the latest technologies, it should nevertheless remain a way to draw the attention of society and the state to the ongoing social, political, and economic phenomena. In any case, being in close relationship with the subjects of digitalization, data journalism continues to be a way of communicating with the audience in a language that is understandable to them, where the specifics of an interesting story should remain in the first place. Another thing is that you need to understand which language is close to the modern digital audience, the signs of which we discussed above.

4. Conclusion

During a short-term experiment with first- and fourth-year journalism students, a difference in information perception was revealed in terms of generating interest in modern audiences for materials using large amounts of data. Junior students primarily paid attention to the accessibility and clarity of visual information, which they preferred in a structured form, as graphics and precise figures that could be used to draw independent conclusions, and only then looked for an explanation in the text. Senior students preferred to get acquainted with the expected conclusions, which should be confirmed by unstructured information and analytically constructed text. This allows us to conclude that it is data journalism, having become part of the work with big data, that is capable of satisfying the needs of such a complex formation as a modern digital audience. As for the characteristics of the digital audience itself, we can conclude that the combination of methods of modern journalism and information technology not only allows us to expand the principles of working with public opinion, but also to guide the development of a digital personality in a positive progressive movement.

Since today humanity and the whole world are undergoing changes due to the rapid development of information technologies, artificial intelligence, growth of big data, global telecommunications, it seems relevant to expand the capabilities of specialized disciplines affecting the capabilities of data journalism in conjunction with related IT courses, which will expand the horizons of knowledge when setting tasks of relevant impact on a new type of audience.

The close relationship of communication processes with the development of the digital world, and as a result, the emergence of a new type of digital audience, has brought to life the urgent problem of forming a pool of professional journalists who are able to work with technical tools and be able to competently visualize content.

Virtual media, neural networks, blogs, interpreting multigenerality in their own way, have already made certain adjustments to the structure of information consumption, changing the appearance of a modern audience. In the era of globalization, this is becoming an important tool, because due to hypertextuality, convergence and interactivity, the format of communication is also changing and new functions are emerging, and these processes are already stable and often spontaneous. However, the specifics of developing and promoting modern content based on large amounts of data in the Internet space undoubtedly need to be studied in depth, since the new user habits formed serve as a source of new communications, for which the basic characteristics of the principles of high-quality journalism may be important, one of the main formats of which is careful work with volumes of data.

It seems relevant to expand the capabilities of specialized disciplines that touch upon the capabilities of data journalism in conjunction with related IT courses, which will allow us to expand our knowledge horizons when setting tasks for relevant impact on a new type of audience.

An important role in this was played by the fundamental elements of audience development under the influence of constant immersion in the digital environment, digitalization of consciousness, changing the possibilities of the real world, feelings of individuality, experience and self-perception. Our sense of experience, sense of boundary, sense of continuity, and accumulated memories will also shift and alternate between our online and offline status. When researching the modern media audience, it is necessary to monitor any changes between generations and any cultural differences in the original archetype in the process of digitization in time and space, paying special attention to any changes in the fundamental human nature and universal qualities for further necessary analysis. A joint program of various stakeholders, including multidisciplinary specialists and experts, scientists and researchers, experts in the field of information technology and communications, would be a way to explore this unique and global issue related to the existence of a digital society.

The main idea of the above-described causes of changes in our original consciousness in the pre-existing physical world, as well as in the new and gradually dominant digital world, may first include migration and manifestation from the physical world to the digital world, mainly through selective self-image and self-perception based on experiences and existence in the physical world, gradually leading to the formation of a new form of self-awareness in the digital world. In turn, the digital world also influences the primordial consciousness in the physical world. In addition to the indicated changes in people's minds, an absolutely new form of personality of the digital world is additionally being formed and growing in accordance with the new rules and conditions in the digital world for the younger generation, which can already be called the digital generation [23]. Thus, the unification of the digitized form of personality self-awareness from the physical world and the new form of personality self-awareness of the digital world as a result will lead to the formation of this new form of the digitized archetype of personality consciousness in the era of digitalization.

A relevant step in this direction can be considered the introduction of such a discipline as "Trendwatching and Trendsetting Technologies" into the educational process in addition to the course "Data Journalism", since this course allows, based on data analysis and analysis of audience interests, to predict long-term trends in various areas of social life, including economic and socio-political interests, while attracting IT technologies and expanding the boundaries of related disciplines. For modern journalism, it seems very relevant to expand the professional staff not only by journalists who are freely oriented in the digital world, but also to create special project teams that

include programmers, designers and IT specialists, which will allow editorial offices to quickly find a common language with the audience and attract it to their side.

Given the information noise, the rapid change of interests, the switching of attention among the continuous flow of Internet content, the ability to work with Big Data and the development of software has become not just a new round of development of data journalism, whose roots go back to the last centuries, but a real trend in digital media formats. In addition, these skills go beyond the media space, allowing professionals to find themselves in any field where the ability to work with data is required, and generally change the content of data journalism, which is currently developing at the intersection of technical and humanitarian knowledge.

Acknowledgements

This work, conducted without external funding, is presented as analytical material. We extend our gratitude to the conference organizers and university management: Chairman of the Board and Rector Khikmetov A.K., Vice-Rector for Global Partnership and Additional Education and Board Member Daineko E.A., Director of the Department for Research Activities Ipalakova M.T., and Vice-Rector for Academic and Educational Activities and Board Member Mustafina A.K.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] V.N. Bogatyreva, Scientific report in the journal as a modern journal, 2021, URL: <url> <https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-v-zhurnalistike-kak-sovremennyy-mediatrend>.
- [2] Digital 2023: Global Overview Report, URL: <https://datareportal.com/reports/digital-2023-global-overview-report>.
- [3] Yiping Xia, Sue Robinson, Megan Zahay & Deen Freelon. The Evolving Journalistic Roles on Social Media: Exploring “Engagement” as Relationship-Building between Journalists and Citizens. // Journalism Practice. – 2020. – Vol. 14, Issue 5. – pp. 556-573.
- [4] Park Chang Sup and Barbara K. Kay, what is it? Casual exposure to social media news, news perception, news effectiveness, and news consumption in 2020. – Mass communication and society. Volume 23, issue 2. – pp. 157-180.
- [5] Monica Jerf-Pierre, Adam Shehata and Bengt Johansson, Changes in the importance of the media and public perception of reality: how fluctuations in media attention affect the strength of citizens' sociotropic beliefs // Mass communication and society. The article is published on the Internet. <https://doi.org/10.1080/15205436.2023.2299209>.
- [6] O.V. Demidkina, K.O. Vishnevsky, Digital technologies and society: impact on human well-being and quality of life - Scientific Digest No. 7 (12), – HSE, 2022.
- [7] A.N. Pronkina, Transformation of memory in conditions of information saturation // Philosophy of Science and technology – 2020. – Vol. 25, No. 1. – pp. 110-125.
- [8] S. Barlybayeva, .Modern media technologies: monograph / S. Barlybayeva – Almaty: Kazakh University, 2019. – 260 p.
- [9] M. Castels, The power of communication: textbook – 3rd ed/ Castels M. – M.: Publishing House of the Higher School of Economics, 2020. – 591 p.
- [10] S. Botarev, Scientific report in the journal: graduate or final? 2022, URL: <url> <https://supernova.is/editorial/iskusstvennyj-intellekt-v-zhurnalistike>
- [11] The rating appeared on the pages of the magazine, 2018, URL: <https://www.kommersant.ru/doc/3414753>.
- [12] A. Ankyzy, scientific journal, 2022 URL: <url> <https://massaget.kz/blogs/26447>.

- [13] Michael Buosis, Internet Freedom, free culture and free information: Aaron Schwartz and the neoliberal turn of Cyberlibertarianism. The history of the Internet. – Digital technologies, Culture and Society, 2023. – Volume 7. Issue 3. – pp. 203-219.
- [14] Gene Burgess and Edward Harcomb, Digital journalism as a symptom, reaction and factor of change in the platform media environment. Digital Journalism, 2019. – Volume 7, Issue 3: Definition of Digital Journalism (Research). – pp. 359-367.
- [15] D.A. Belyaev D.A. Perspective anthropological models of the posthuman: transformation of human nature and superhuman attributes. The Global Future 2045: An anthropological crisis. Convergent technologies. Transhumanistic Projects, 2014. – Moscow: "Kan+" And "Republic". – pp. 43-52.
- [16] Automated tool: Pentagon is the creator of the search engine for “online punk geopolitical exchanges”, 2021, URL: <https://xn---8sbeb2annejbd6aej.xn--p1acf/v-mire/item/2801-avtomatizirovannyj-instrument-pentagon-sozdast-programmu-obnaruzheniya-onlajn-kampanij-geopoliticheskogo-vliyaniya>.
- [17] More than 1,800 congressmen once enslaved blacks. Here's who they were and how they shaped the nation, 2022, URL: <https://www.washingtonpost.com/history/interactive/2022/congress-slaveowners-names-list/>.
- [18] Mapping places where the earth will become uninhabitable, 2022, URL: <https://awards.journalists.org/entries/mapping-where-the-earth-will-become-uninhabitable/>.
- [19] Mike Bostock, Shan Carter, Erik Hinton and Ruth La FerlaFront. Row to Fashion Week, 2014, URL: <https://www.nytimes.com/newsgraphics/2014/02/14/fashion-week-editors-picks/index.html>.
- [20] Statistics of the day, 2021, URL: <https://stories.media/workshops/2021/11/05/zhurnalistika-dannikh-chast-1-istoriya>.
- [21] Los Angeles Times Data and Graphics Department, URL: <https://github.com/datadesk>
- [22] Refined Data, URL: <https://datadesk.eco/>.
- [23] Saparova D., Kanagatova A., Digital generation through the prism of culture, 2018 – Bulletin. Philosophy Series. Cultural Studies Series. Political Science Series. – No. 1 (63). – pp. 187-197.