LUBELSKI ROCZNIK PEDAGOGICZNY
T. XLIV, z. 3 – 2025

DOI: 10.17951/lrp.2025.44.3.61-76

Katarzyna Borawska-Kalbarczyk
University of Białystok
ORCID – 0000-0003-4315-2226

CRITICAL INFORMATION ANALYSIS IN RELATION TO THE THREAT OF DISINFORMATION IN DIGITAL MEDIA AND AI HALLUCINATIONS – AN EDUCATIONAL CONTEXT*

Introduction: The rapid development of the Internet has led to numerous complications in the use of digital information. It has enabled the manipulation of messages, the deliberate creation of false content, and new forms of information distortion driven by advances in artificial intelligence (AI) tools.
Research Aim: This article aims to analyze the nature and causes of selected disruptions in digital information (fake news, disinformation, and AI hallucinations). The author emphasizes the importance of critically analyzing media content in the face of false information, a topic explored in the final section of the text.
Evidence-based Facts: In recent years, the number of publications on disinformation in digital media has increased significantly. Experts consider it one of the most serious global threats in the digital space. Additionally, there is a growing concern about experiencing AI-generated hallucinations, such as those produced by ChatGPT. Studies available in the literature focus on the essence of disinformation, the reasons for its spread, and its consequences, particularly in limiting critical thinking.
Summary: Given the threats posed by false information, special attention should be directed toward children and adolescents, the most active group of Internet users. Awareness of the causes of false information, the dangers of its reception, and ways to counteract the harmful effects of disinformation highlight the need to enhance the digital skills of young people. These competencies enable students to effectively protect themselves from information manipulation and build resilience against propaganda and false narratives.

Keywords: fake news, disinformation, information hallucinations, AI literacy, critical thinking

* Suggested citation: Borawska-Kalbarczyk, K. (2025). Critical Information Analysis in Relation to the Threat of Disinformation in Digital Media and AI Hallucinations – an Educational Context. *Lubelski Rocznik Pedagogiczny*, *44*(3), 61–76. http://dx.doi.org/10.17951/lrp.2025.44.3.61-76

INTRODUCTION

The current information culture, along with all its blessings as well as challenges and threats, brought about a major revaluation of human functioning up to this day. On the one hand, it enabled access (on a scale never before experienced) to powerful information resources, but on the other, it created a space for digital abuses and threats. Along with the dynamic development of digital technologies, there arose some disconcerting possibilities of message manipulation, deliberate creation of fallacious content, as well as new forms of information distortion, in relation to the evolution of artificial intelligence (AI) tools. Digital mass media, and especially social media, cause all information, including false, to be circulated, modified, and distorted at a lightning speed and on a global scale. What is more, the research points to the fallacious content spreading faster than the truth (Vosoughi et al., n.d., p. 2), whereas social media is a perfect catalyst for disinformation.

The dominant characteristics of an information society, which is the access to information as well as its importance as an impetus for creative growth and innovation, not to mention, as a tool in the education process – understood as increasing human knowledge – quite quickly transformed from informational wealth to informational collapse. The availability of information that is currently experienced, the extent, the plurality of content, and speed of circulation are difficult to compare with anything else in the history of humankind thus far.

RESEARCH AIM AND QUESTION

Taking into account the spread of different kinds of disinformation in the digital space, which only deepen the infamous pathology of Internet culture, particular attention should be paid to the educational conditions of the development of digital competences among children and the youth, as to the group of the most involved Internet users. The purpose of this article is to analyse the nature and causes of digital information distortion. Featured among them are various types of fallacious information (including disinformation), as well as content of a hallucinatory nature derived from AI tools. The analysis undertaken by the author in the article aims to answer the question about the nature, manifestations, and causes of the spread of various types of informational disorder, including AI-generated hallucinations. This text presents a thesis that one of the major conditions of counteracting the adverse effects of informational chaos is the development of critical thinking among the students, which is put forth in the end section of this article.

The reflections contained in this article are theoretical and based on literature studies (c.f. Paré et al., 2015). They are to be defined within the framework of an critical discussion concerning the issue of threats to one's cognitive functioning

posed by disinformation in digital media. The literature review also forms the context for indicating the importance of the requirement for development of critical thinking among students as well as for the articulation of the original conclusions and practical recommendations, which are directly related to the students' critical thinking abilities forming process in the course of their media education, considered as part of lifelong learning.

EVIDENCE-BASED REVIEW

Disinformation in digital media

Making the information consumer status equal with that of the creator – a development brought about by the dizzying rise of social media – engendered perfect conditions for the creation and circulation of an immeasurable volume of information, including that of dubious quality, unreliable, and implausible. Authors and researchers who engage in discourse concerning the presence of information in the digital space distinguish certain types of it, which bear hallmarks of manipulation. It seems that the most widespread classification category of disinformation is fake news.

Extrapolating from many available definitions, it should be assumed that "fake news refers to news articles that are intentionally and verifiably false and could mislead readers" (Allcott & Gentzkow, 2017, p. 213). In the opinion of Palczewski (2019),

> a masterfully prepared piece of fake news is essentially indistinguishable from a true message. Informational value of fake news is often identical to that of an actual news piece (e.g. topicality, proximity, intensity, scale, importance, surprise, luridness, prominence of persons and countries, personalisation, or negativity). (p. 24)

The above-mentioned characteristics are the reason why many of us so easily and quickly succumb to the illusion of truthfulness generated by fake news.

In the literature concerned with the analysis of the nature of fake news and their influence, is it typical to single out the following categories: satire or parody, misleading content, imposter content, fabricated content, false connection, false context, manipulated content (Wardle, 2017). Quite often, the phenomenon of disinformation is assumed synonymous with fake news. According to the definition devised for the European Commission by the High-Level Group of Experts,

> disinformation is understood as verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm. (…) Disinformation does not include reporting errors, satire and parody, or clearly identified partisan news and commentary. (European Commission…, 2018, pp. 3–4)

Taking into account the degree of complexity and intricacy of this term, it shall be only noted that disinformation is classified as part of the so-called informational chaos or disruption of the information environment – just next to the so-called mis-information or mal-information (Wardle & Derakhshan, 2017). Due to this, disinformation should be classified not as a term, but as a process, as among the deciding factors for fake news to become characterised as disinformation features, in particular, the intent behind their creation. The phenomenon of intentional, purposeful circulation of fake news in media space is defined precisely as disinformation, whereas if the origin of their creation was connected to a low level, or even lack of knowledge or insufficient awareness on the part of the creator, then we are dealing with malinformation. What is more, the content classified as malinformation is truthful, but circulated with malicious intent. The third of the listed categories of informational chaos is misinformation – it means that the content is verifiably false, but it is not circulated with malicious intent. The assumption is that the creator has no intention of deceiving the recipient of the message, only that they are negligent towards their own extent of knowledge or thoughtless in the process of creating the message (Lelonek, 2020).

Regardless of the underlying intent behind the creation of fallacious information, they contribute to the ever-growing deluge of junk news. These involve low-value content, created to cause a stir and geared towards attracting attention of the widest possible audience, with a sensationalistic title, often containing one-dimensional information bearing the hallmarks of conspiracy theories. Their essence lies in the action of spreading, but not fakeness (Venturini, 2019). The following fake news definition also makes a reference to the junk metaphor: "news content published on the internet that aesthetically resembles actual legitimate mainstream news content, but that is fabricated or extremely inaccurate. Also referred to as false, junk, or fabricated news" (Pennycook & Rand, 2021, p. 389).

Selected causes of informational chaos

While attempting to find an answer to the phenomenon of fallacious content seducing the minds of the public, we should refer to the emotions that are activated when such information is received. Indicated here is an adversely proportional relation – heightening of the emotions triggered by a text or image containing the message yields lower chance of rational analysis of how the content is legitimised and its modality validated (Kajfosz, 2013). Another way of multiplying the disinformative content is human affirmation and the pleasure caused by the perception of novelty. Fallacious messages very often bear the hallmarks of novelty, at the same time being "more novel – and people share new information more frequently. On social media, the people who share new information are regarded as better informed" (Baron-Polańczyk, 2019, pp. 223–224). To explain the phenomenon of circulation of the fallacious content, Pareto principle (from the name of the Italian economist)

can be applied – it is also known as 80/20 rule, and used in theory of economy and management. It describes an imparity of phenomena occurring, indicating that usually 20% of causes generates up to 80% of effects (Jemielniak, 2024). Transferring this relation to the sphere of content production on social media, it turns out that disinformation originating from politicians, celebrities, and other widely known personalities amounts to about 20% of it all, but is responsible for 69% of total engagement on social media (Bagherpour & Nouri, 2020). Social media profiles and information channels of the so-called opinion leaders, or persons of prominent opinion-making status, can influence the views of their followers in a significant way, while being carriers of more or less intentional disinformation.

Algorithms play a key role in accelerating the spread of mis/disinformation. Even children are aware of the dynamics of algorithms that maximize attention. Seeking popularity and validation, children post on social media hoping that something will "go viral" and they will gain popularity and visibility, which requires high engagement. It is likely these motivations among online users that lead to misleading or fabricated content often surpassing credible information (Howard et al., 2021).

In an overview of the general characteristics of content placement for the purposes of disinformation, some formula-content-function features can be indicated, such as: domination of anonymity; viral spreading; presence of "friend of a friend" formula as a reinforcement of validity; citing difficult to verify scientific data as hard facts; presenting existentially important or interpretively complex matters; describing the world according to the "we – them: ours – alien" model (Chudzik, 2021, pp. 150–151).

Informational disruptions of generative AI

In the light of the latest report of the World Economic Forum, the experts deemed disinformation and misinformation to be the most severe global risk anticipated for the next two years, and listed them as fifth in their long-term, 10-year perspective (World Economic Forum, 2024). As the authors of the report indicate, nowadays, we experience an escalation of volume, reach, and negative efficiency of fallacious information – the primary cause of which is the technological progress, connected to generative AI, among others. The tools of generative AI are currently revolutionising the way of searching for information and utilising the results created by their algorithms. The article, in the following part, focuses on the phenomenon of hallucination in the context of generative AI, with particular attention to language models. Generative AI represents a subset of broadly understood AI and refers to systems that not only analyse data but also generate new content – texts, images, sounds – based on patterns learned from large datasets. In contrast to classical AI systems, which primarily perform predictive, classificatory, or decision-making tasks based on clearly defined input data, generative AI relies

on probabilistic modelling and can produce statements that appear coherent but are detached from factual reality. It is within this class of systems that the phenomenon of hallucination – that is, the generation of false yet plausible information – takes on particular significance, both in practical and epistemological terms.

One of the latest phenomena among AI tools is ChatGPT – a chatbot equipped with a conversation-based, AI interface, developed by OpenAI. Made available in 2023, ChatGPT now has around 100 million users. It was able to reach this result in mere two months, which makes it world's fastest-growing application of this type (Walsh, 2024). Neff presents a disturbing vision of digital reality, related to the threat of AI-created content. She writes on the Wired website:

> in 2024, we will face a grim digital dark age, as social media platforms transition away from the logic of Web 2.0 and toward one dictated by AI-generated content. Companies have rushed to incorporate large language models (LLMs) into online services, complete with hallucinations, inaccurate, unjustified responses, and mistakes, which have further fractured our trust in online information. (Neff, 2024)

Additionally, the research indicates that "potential concerns of chatbots haven't been investigated as much. (…) It is therefore important to investigate the concerns of using this technology, ChatGPT, in education to ensure safe use" (Tlili et al., 2023, p. 2).

One of the more serious threats is the possibility of inducing hallucinations from ChatGPT, that is generating content that seems to be authentic, but is, in fact, fallacious. It seems valid to ask whether ChatGPT is a "bullshit spewer" (see: Rudolph et al., 2023), however, this question is complex and multi-faceted. On the one hand, ChatGPT and similar AI tools can generate text that is not always precise nor reliable, which raises concerns regarding the authenticity and quality of the content produced. Some researchers claim that excessive reliance on such tools may result in students presenting works that are devoid of originality, critical thinking, and actual understanding of the topic (c.f. Farhi et al., 2023).

The results of "2024 AI & ML Report: Evolution of Models & Solutions" by Aporia, a leader in the AI control platform sector, offer a cause for concern regarding the growing trend for increasingly common hallucinations and prejudice in the field of generative AI. AI hallucinations refer to the cases in which AI models may generate results that are "incorrect, nonsensical, or disconnected from reality. These hallucinations can range from minor inaccuracies to significant errors, including the generation of biased or potentially harmful content" (Tardif, 2024). What is more, biased content may solidify stereotypes or unfair practices, whereas in sensitive applications (health, finances, law), such errors may entail serious consequences, altering human decisions (Tardif, 2024). In light of the research conducted by an interdisciplinary consortium of experts

appointed to analyse nine cases from different areas of security (e.g. using mobile phones while driving, supervising children near bodies of water, principles of managing a crowd, operating heavy machinery, etc.) concerns were formulated, connected to the potential misuse of ChatGPT, especially in terms of providing inappropriate or harmful information regarding safety in the analysed situations (Oviedo-Trespalacios et al., 2023). The text features an outright appeal by the researchers:

> given the risks associated with following insufficient safety advice, we caution any users interacting with ChatGPT to source safety-related information and advice. Given the popularity of ChatGPT and its potential, safeguards are urgently needed to help prevent misuse for safety advice. (Oviedo-Trespalacios et al., 2023, p. 18)

Therefore, it appears that the currently ongoing rapid evolution of technology employed in many chatbots, including ChatGPT, has little influence over the fact that they may exhibit harmful behaviours, classified as information manipulation and disinformation. As a result, utilising such content may have particularly negative meaning for the recipients with low levels of digital competency (Tlili et al., 2023). Moreover, as the researchers of this phenomenon indicate, information that is easily generated by ChatGPT may have an adverse influence on critical thinking and problem solving skills. The main reason for this is identified as the simplification of obtaining answers and information, as it may even reinforce cognitive laziness and counteract the motivation of learners to seek out their own intellectual pursuits, conclusions, or solutions (Kasneci et al., 2023). Similar conclusions arise from other scientific reports, indicating that, in actuality, ChatGPT has proven that it is priceless as an educational tool, however, an overreliance on it may potentially impair the development of critical thinking and problem-solving skills – both key competencies for students. The risk is that they may have too much faith in ChatGPT's answers, which reduces their own abilities to explore and analyse (Hasanein & Sobaih, 2023).

Students' critical thinking development in the process of education

Despite the fact that children are considered to be "born digital", we cannot equate the familiarity with technology with critical thinking and media consciousness. To move carefully across the complex online landscape, a certain level of reading and writing abilities is needed, as well as maturity, which children may not attain solely by way of online activity (Howard et al., 2021). Confronting the risk of disinformation experienced in relation with the content originating from social media, along with the confabulation and hallucinations generated by AI tools, is inductive toward constant refinement of children and youth's digital education. In 2018, the Commission of the High Level Expert Group on Fake News and Online Disinformation highlighted the importance of media literacy as "a response to dis-

information because it can empower individual users as suggested above and mass empowerment of users will lead to greater social resilience against disinformation and perhaps other disorders of the information age" (European Commission: Directorate…, 2018, p. 25).

In the course of the current information revolution caused by the presence of generative AI tools, advanced media education becomes the more significant in relation to the development of students' digital competencies.

> To equip both students and teachers to become apt in the use of AI for their academic purposes, a new "culture of AI" seems in order. An AI-culture should permeate academic life, creating an environment where AI is not feared but readily used, understood and – most importantly – critically evaluated. (Walter, 2024, p. 8)

Without a properly planned and carried out media education, there is a significant risk of the students not fully grasping the importance of a responsible use of AI tools and not recognising their limitations. As a result, a lack of critical assessment abilities could make the students put too much faith in AI-generated content, which places the quality and integrity of their independent work in jeopardy. On the other hand, using the AI tools correctly may offer a chance to form critical thinking skills (Walter, 2024).

A developed critical thinking ability allows to defy media manipulation and use digital content in a creative way – which was indicated by Kellner and Share (2005). The process of developing critical thinking abilities – in relation to the constant evolution of digital and media environments, which resonate with existing concerns connected to the AI dynamism – shows the hallmarks of a continuous education. It is understood as a need to educate individuals who exhibit creative and dynamic attitude towards life and culture, have the capacity for perpetual self-improvement, as well as for increasing the quality of their own existence. In relation to the people with developed digital competencies, including that of critical thinking, it is much more probable that they will be able to assess digital content, identify sources that can be trusted, and make conscious decisions.

While invoking the notion of critical thinking, one should refer to Freire's interpretation of the term, who in 1970 noted that "critical literacy views readers as active participants in the reading process and invites them to move beyond passively accepting the text's message to question, examine or dispute the power relations that exist between readers and authors" (National Literacy Trust, 2018, p. 16). While considering how critical literacy may be incorporated into media education, it might be worth to contemplate Luke and Freebody's "four resources model" updated for the digital era in 2013 by Hinrichsen and Coombs. According to this model, students attain the abilities of code breaking, meaning making, text using, and text analysis, becoming: Code breakers – Meaning makers – Text users – Text analysts (National Literacy Trust, 2018, p. 17). The addition of the

fifth element, called Persona, meant sensitisation to the matters of reputation, identity, and membership in different digital contexts, as well as the development of a sense of belonging and a determined role of a digital space user (Hinrichsen & Coombs, 2013).

In the face of the dynamic evolution of digital technology tools based on AI, it may be assumed that the process of critical thinking development is worth including in the scope of the so-called AI literacy. In an effort to define what AI literacy is, Ng et al. (2021) distinguished four definition categories:

> [1] Know & understand AI: Know the basic functions of AI and how to use AI applications in everyday life ethically; [2] Apply AI: Applying AI knowledge, concepts and applications in different scenarios; [3] Evaluate & create AI: Higher-order thinking skills (e.g., evaluate, appraise, predict, design) with AI applications; [4] AI ethics: Human-centered considerations (e.g., fairness, accountability, transparency, ethics). (p. 505)

Other researchers note that "the components of AI literacy are ill-defined and it is unclear to what extent middle school students can engage in learning about AI as a sociotechnical system with socio-political implications" (Zhang et al., 2023, p. 290). Nevertheless, they highlight that students should have an

> understanding of three domains in order to become prepared for participating in AI-infused fields and industries of the future: (1) an age-appropriate technical knowledge and skills in AI, (2) an understanding of AI's ethical and societal implications, and (3) a knowledge of AI's impact on jobs (AI career futures) and how to adapt to the future of work (career adaptability). (Zhang et al., 2023, p. 291)

Conducting classes on AI literacy may be an excellent opportunity for developing critical thinking skills related to using AI tools: result analysis, assessment of their reliability and usefulness, identification of biases using AI algorithms, or understanding the limitations as well as the assessment of ethical implications of the decisions made by AI. Here, case studies would be of value – chosen so that they encompass a wide variety of topics, which would make it possible to analyse situations that are close to reality as well as inspire reflection on ethical, moral, and social consequences of using AI tools (Walter, 2024). In the context of developing critical thinking, it is essential to cultivate the ability to evaluate information generated by AI tools – an integral aspect of what is commonly referred to as information literacy. "Information literacy is the ability to think critically and make balanced judgements about any information we find and use" (CILIP, 2018, p. 3). Therefore, the development of information literacy constitutes a fundamental prerequisite for fostering critical thinking.

SUMMARY

Carefully chosen goals and content related to media education are undoubtedly not only required for adapting students to technological changes, but also a vital element of shaping them into future members of society who are able to consciously analyse and make decisions in a world that is increasingly overrun by technology. In the age of ubiquitous technology and rapidly increasing volume of information available on the Internet, critical media content analysis is becoming a remarkably vital ability.

Moreover, media education should not be restricted only to learning how to recognise fallacious information, or to cybersecurity. A critical reflection on the influence of media on society, culture, and values should also be among its goals. This would give the students the ability to analyse the social, ethical, and moral implications that digital technologies have on human functioning in the digital world. As a result, they may have a chance to understand the consequences of their actions in the hybrid online-offline world. When faced with a need to select information and assess their reliability in the world inundated by fallacious news originating from digital media, the development of critical thinking as part of media education becomes very much vital.

CONCLUSION

In today's information environment, characterized by data overload, rapid dissemination, and the presence of disinformation, the development of critical thinking has become a key task for effective media education. The ability to critically engage with digital content allows not only for the assessment of source credibility, but also for the identification of manipulative communication techniques, the distinction between facts and opinions, and the analysis of the quality of arguments. These competencies enable students to navigate the media landscape consciously, reducing their susceptibility to fake news, clickbait, or AI-generated content without proper quality control. Particularly important in this context is cognitive reflexivity – the ability to recognize one's own cognitive biases, emotions, and mechanisms that influence content interpretation. Equally crucial is the ability to compare different sources of information and recognize diverse perspectives, which promotes a multifaceted understanding of complex social and cultural phenomena. The cultivation of such attitudes not only enhances resilience to information chaos but also develops civic competencies, responsibility for content shared online, and a culture of dialogue. Thus, critical thinking should not be regarded as one of the goals of media education, but as its foundation and a condition for its effectiveness.

Developed cognitive abilities, based on critical thinking skill, are a decisive factor in the students' future quality of social life, as they facilitate the evaluation

of information and profound reflection while processing them. The development of critical thinking enables young people to protect themselves from information manipulation and builds immunity to propaganda and false narrations. A student who thinks critically does not stop at the surface-level, simple absorption of media messages, but rather gets involved in the process of deeper inquiry – comparing sources, verifying information, and understanding the context of the content creation. As a result, having developed competencies of information analysis and interpretation, they have a chance to become an active participant in social life, capable of conscious engagement in public discussions and making rational decisions.

The length restricted of this text does not allow for presenting precise and detailed solutions regarding the execution of media education objectives. However, it should be stressed that the creation of a worthwhile learning environment is conditional on the teacher applying methods and strategies known in the field of didactics as problem-based. Among their hallmarks is the prevalence of learning over teaching as well as multilateral activation of the students. These methods require the students to perform analysis, synthesis, and assessment of the information in use, which foster practice of various ways of reaching the truth. They emphasise active creation, self-reliance in acquiring knowledge, ability to search, select, and verify information.

However, there is no doubt that implementing this is connected to systemic changes throughout all aspects of education. It would involve a departure from expository teaching, dominated by facts, definitions, and rules, towards an education based on independent thinking and uncovering meaning. Certain methods of education are of considerable importance here, among them: discussion, project-based lesson format (including WebQuest method), didactic games (e.g. brainstorming or case-based methods), games involving simulations, source material analysis, argument mapping, and others. Their common feature – typical of problem-based methods to which they belong – is the aim of encouraging students to work independently and to engage them in a multifaceted way, with a strong emphasis on creative thinking and innovation. Through various forms of stimulating students' creative activity and offering diverse opportunities to develop cognitive skills such as comparing, classifying, inducing, deducing, identifying errors, constructing arguments, abstracting, and analyzing viewpoints, these methods are more effective in teaching than traditional approaches. As problem-based methods, they require students to independently develop new strategies for solving problems – strategies that are previously unfamiliar to them.

Student group work organisation is also key, as they share their own knowledge and experience in the process. It is even more beneficial, considering that the correctly utilised digital space may become a valuable environment for collective learning. It is conducive to building community and cooperation among students,

which also constitutes added value of the modern media education – remembering that the primary message of social media was participation, not disintegration and polarisation.

Moreover, it is important to encourage the students to reflect on their own thought processes, ask questions, and seek out other approaches to solving problems, which is also conducive to strengthening critical thinking abilities. It is also vital that the teachers themselves exhibit critical thinking patterns through their teaching methods, including by active listening, receptivity to questions and doubts being raised by and appreciating the diversity of the students' views.

In conclusion, I note that educating on the sustainable use of the benefits of digital media should not be limited solely to children and adolescents. The differentiation between digital natives and digital immigrants introduced by Prensky (2001) seems less sound nowadays, considering the ongoing digital transformation (whose direction is often difficult to predict). Adopting such a dichotomy glorifies the younger generation (so-called digital natives), somewhat automatically granting them high media competence. And conversely, labelling the older generations as digital immigrants, equates them to people who encounter numerous difficulties when trying to function in a virtual space, whereas both can display varying levels of digital competence. Essentially, consistent access to digital space (which constitutes a typical attribute of the younger generation) does not determine the quality of their competence regarding the way of processing information as well as critical content analysis. Therefore, both the so-called digital immigrant and the digital native may become a digital outcast – not as a consequence of the lack of access to digital media, but, above all, as a result of a dearth of ability to manage digital information as well as interpret them. Thus, every generation faces the challenge of adjusting to new circumstances formed by the state of the digital world. It needs to be acknowledged that the key condition of active participation in digital media is lifelong learning. The concept of lifelong learning, therefore, appears to be an activity determined by the ever more complex and unpredictable development of digital technologies as well as a process enmeshed in qualitatively new and changing relations between an individual and the virtual world. The lifelong learning paradigm triggers the need for change in media education, in the course of which students should acquire complex digital competences – considered not in a technical, but cultural context. It entails figuring out crucial mechanisms of information generation, how it is received and assessed, creating and presenting, especially in the current environment of informational chaos.

The theoretical reflections above do not exhaust the issue of threats arising from digital content disinformation and the need for shaping critical thinking among students. From a theoretical perspective, the problem addressed in this article – especially the part concerned with the analysis of the influence of AI-generated content – indicates the need for further study aimed at changing the philosophy of

teaching. That means refining the abilities of teachers regarding not only the use of AI tools in education, but also the recognition of their limitations and dangers. It is assumed in this context that education in the fields of both digital competencies of current media users and critical thinking among students constitute issues requiring constant reflection and actualisation with regard to the dynamic evolution of digital media spaces, shaping the lives of us all.

REFERENCES

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Bagherpour, A., & Nouri A. (2020). *COVID misinformation is killing people. This "infodemic" has to stop*. https://www.scientificamerican.com/article/covid-misinformation-is-killing-people1

Baron-Polańczyk, E. (2019). Boty, trolle i fake news – uważaj, kto cię uczy! *Edukacja – Technika – Informatyka*, *2*(28), 218–226.

Chudzik, A. (2021). Pandemiczne mity i teorie spiskowe w memach internetowych. In W. Świerczyńska-Głownia, A. Hess, & M. Nowina-Konopka (Eds.), *Komunikowanie w procesie zmian* (pp. 145–164). Instytut Dziennikarstwa, Mediów i Komunikacji Społecznej Uniwersytetu Jagiellońskiego; Wydawnictwo ToC.

CILIP. (2018). *CILIP definition of information literacy 2018*. https://cdn.ymaws.com/www.cilip.org.uk/resource/resmgr/cilip/information_professional_and_news/press_releases/2018_03_information_lit_definition/cilip_definition_doc_final_f.pdf

European Commission. (2018). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. *Tackling Online Disinformation: A European Approach*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0236

European Commission: Directorate-General for Communications Networks, Content and Technology. (2018). *A multi-dimensional approach to disinformation.* Publications Office. https://data.europa.eu/doi/10.2759/739290

Farhi, F., Jeljeli, R., Aburezeq, I., Dweikat, F.F., Al-shami, S.A., & Slamene, R. (2023). Analyzing the students' views, concerns, and perceived ethics about chat GPT usage. *Computers and Education: Artificial Intelligence*, *5*, 1–8. https://doi.org/10.1016/j.caeai.2023.100180

Hasanein, A.M., & Sobaih, A.E.E. (2023). Drivers and consequences of ChatGPT use in higher education: Key stakeholder perspectives. *European Journal of Investigation in Health, Psychology and Education*, *13*, 2599–2614. https://doi.org/10.3390/ejihpe13110181

Hinrichsen, J., & Coombs, A. (2013). The five resources of critical digital literacy: A framework for curriculum integration. *Research in Learning Technology*, *21*, 1–16. http://dx.doi.org/10.3402/rlt.v21.21334

Howard, P.N., Neudert, L.M., & Prakash, N. (2021). *Digital misinformation/disinformation and children. UNICEF Office of Global Insight and Policy*. https://www.unicef.org/globalinsight/media/2096/file/UNICEF-Global-Insight-Digital-Mis-Disinformation-and-Children-2021.pdf

Jemielniak, D. (2024). Polski internet jest pełen potężnych teorii altermedycznych. *Gazeta Wyborcza,* January 31.

Kajfosz, J. (2013). Faktoid i mistyfikacja w nowych mediach, czyli o strategiach czarowania umysłu. In P. Grochowski (Ed.), *Netlor: wiedza cyfrowych tubylców* (pp. 65–90). Wyd. Nauk. UMK.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of Large Language Models for education. *Learning and Individual Differences*, *103*. https://doi.org/10.35542/osf.io/5er8f; https://osf.io/preprints/edarxiv/5er8f

Kellner, D., & Share, J. (2005). Toward critical media literacy: core concepts, debates, organizations, and policy. *Discourse: Studies in the Cultural Politics of Education*, *26*(3), 369–386. https://doi.org/10.1080/01596300500200169; http://web.mit.edu/schock/www/docs/towardscritmedlit.pdf

Lelonek, A. (2020). *Leksykon bezpieczeństwa informacyjnego, cz. 2*. https://nci.org.pl/leksykon-bezpieczenstwa-informacyjnego-cz-2/

National Literacy Trust. (2018). *Fake news and critical literacy. The final report of the Commission on Fake News and the Teaching of Critical Literacy in Schools*. https://nlt.cdn.ngo/media/documents/Fake_news_and_critical_literacy_-_final_report.pdf

Neff, G. (2024). *The New Digital Dark Age*. Wired. https://www.wired.com/story/the-new-digital-dark-age/

Ng, D.T.K., Leung, J.K.L., Chu, K.W.S., & Qiao, M.S. (2021). AI literacy: definition, teaching, evaluation and ethical issues. In *Proceedings of the Association for Information Science and Technology*, *58*(1), 504–509. https://doi.org/10.1002/pra2.487

Oviedo-Trespalacios O., Peden, A.E., Cole-Hunter, T., Costantini, A., Haghani, M., Rod, J.E., Kelly, S., Torkamaan, H., Tariq, A., Newton, J.D.A., Gallagher, T., Steinert, S., Filtness, A.J., & Reniers, G. (2023). The risks of using ChatGPT to obtain common safety-related information and advice. *Safety Science*, *167*, 1–22. https://doi.org/10.1016/j.ssci.2023.106244

Palczewski, M. (2019). Dyskurs fake newsa. *Annales Universitatis Paedagogicae Cracoviensis, Studia de Cultura*, *11*(1), 15–31.

Paré, G., Trudel, M.C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), 183–199. https://doi.org/10.1016/j.im.2014.08.008

Pennycook, G., & Rand, D.G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Prensky, M. (2001). Digital natives, digital immigrants, part 1. *On the Horizon*, *9*(5). https://www.marcprensky.com/writing/Prensky%20%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf

Rudolph, J., Tan, Sa., & Tan, Sh. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, *6*(1), 342–263. https://doi.org/10.37074/jalt.2023.6.1.9

Tardif, A. (2024). *Rosnące obawy związane z halucynacjami i uprzedzeniami sztucznej inteligencji: raport Aporii za rok 2024 podkreśla pilną potrzebę wprowadzenia standardów branżowych*. https://www.unite.ai/pl/rosn%C4%85ce-obawy--zwi%C4%85zane-z-halucynacjami-sztucznej-inteligencji-i-aporiami-uprzedze%C5%84-Raport-z-2024-r.-podkre%C5%9Bla-piln%C4%85-potrzeb% C4%-99-wprowadzenia-standard%C3%B3w-bran%C5%BCowych/

Tlili, A., Shehata, B., Adarkwah, M.A., Bozkurt, A., Hickey, D.T., Huang, R., &Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environment*, *10*(15), 1–24. https://doi.org/10.1186/s40561-023-00237-x

Venturini, T. (2019). From fake to junk news, the data politics of online virality. In D. Bigo, E. Isin, & E. Ruppert (Eds.), *Data Politics: Worlds, Subjects, Rights* (pp. 1–15). Routledge.

Vosoughi, S., Roy, D., & Aral, S. (n.d.). *The spread of true and false news online. MIT initiative on the digital economy research brief*. https://ide.mit.edu/wp-content/uploads/2018/12/2017-IDE-Research-Brief-False-News.pdf

Walsh, M. (2024). *ChatGPT statistics (2024) – the key facts and figures*. https://www.stylefactoryproductions.com/blog/chatgpt-statistics

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, *21*(15), 1–29. https://doi.org/10.1186/s41239-024-00448-3

Wardle, C. (2017). *Fake News. It's Complicated. First Draft News*. https://firstdraftnews.com/fake-news-complicated/

Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. Report to the Council of Europe. Strasbourg Cedex: Council of Europe. https://shorensteincenter.org/information-disorder-framework-for-research-and-policymaking

World Economic Forum. (2024). *The Global Risk Report* (19th ed.). https://www.zurich.com/knowledge/topics/global-risks/the-global-risks-report-2024

Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y., & Breazeal, C. (2023). Integrating ethics and career futures with technical learning to promote AI Literacy for middle school students: An exploratory study. *International Journal of Artificial Intelligence in Education*, *33*, 290–324 https://doi.org/10.1007/s40593-022-00293-3

KRYTYCZNA ANALIZA INFORMACJI W OBLICZU ZAGROŻEŃ WYNIKAJĄCYCH Z DEZINFORMACJI W MEDIACH CYFROWYCH I HALUCYNACJI AI – KONTEKST EDUKACYJNY

Wprowadzenie: Dynamiczny rozwój Internetu przyczynił się do licznych komplikacji w zakresie korzystania z cyfrowych informacji. Powstały zarówno możliwości manipulowania przekazem, celowego tworzenia treści fałszywych, jak i nowe formy zniekształcania informacji wynikające z rozwoju narzędzi sztucznej inteligencji.
Cel badań: Celem artykułu jest analiza istoty i przyczyn wybranych zakłóceń informacji cyfrowych (*fake news*, dezinformacja, halucynacje AI). Autor podkreśla znaczenie krytycznej analizy treści medialnych w obliczu fałszywych informacji, co omawia w końcowej części tekstu.
Stan wiedzy: W ostatnich latach znacząco wzrosła liczba publikacji dotyczących dezinformacji w mediach cyfrowych. Eksperci uznają ją za jedno z najpoważniejszych globalnych zagrożeń w przestrzeni cyfrowej. Dodatkowo pojawia się problem doświadczania halucynacji generowanych m.in. przez ChatGPT. Dostępne w literaturze badania koncentrują się na istocie dezinformacji, przyczynach jej rozprzestrzeniania się oraz skutkach ograniczenia myślenia krytycznego.
Podsumowanie: Wobec zagrożeń fałszywymi informacjami szczególną uwagę należy poświęcić dzieciom i młodzieży jako grupie najbardziej aktywnych użytkowników Internetu. Świadomość przyczyn powstawania informacji fałszywych, niebezpieczeństw wynikających z ich odbioru, i sposobów przeciwdziałania szkodliwym wpływom dezinformacji skłania do doskonalenia kompetencji cyfrowych dzieci i młodzieży. Umożliwiają one uczniom skuteczną obronę przed manipulacją informacyjną oraz budują odporność na propagandę czy fałszywe narracje.

Słowa kluczowe: *fake news*, dezinformacja, halucynacje informacyjne, AI *literacy*, krytyczne myślenie