Információs Társadalom

[Information Society]

A SOCIAL SCIENCE JOURNAL

Founded in 2001

English Issue Vol. XXIV, No. 2 (2024)

Published by INFONIA Foundation

Principal sponsor: Budapest University of Technology and Economics, Faculty of Economic and Social Sciences

Technical partner: SZTAKI

ISSN 1587-8694 (Print)
ISSN 2063-4552 (Online)
ISSN-L 1587-8694

Produced by Server Line Print & Design, Budapest

The journal Information Society (In Hungarian: Információs Társadalom, abbreviated as InfTars) aims to provide a platform for research and discussion of the theories and applications of Information Society Studies. Currently every second issue is published in English, the rest are in Hungarian.

The journal is fully Open Access and freely available at https://inftars.infonia.hu/

InfTars is indexed in both the *Web of Science Social Sciences Citation Index and Scopus*, and all titles are automatically submitted to *Crossref*.

Since Vol. VIII, No. 1 (2008) the journal has been included in the Thomson Reuters index (Social Sciences Citation Index, Social SciSearch, Journal Citation Reports/Social/Sciences Edition)

E-mail: inftars-lapman@ponens.org

TARTALOM

LECTORI SALUTEM

The editorial board welcomes the readers of the No. 2 issue of 2024!

This issue is partially a special one, edited in cooperation with the Science and Society Research Group, with several papers related to the topics of a conference organized by The Science and Society Research Group. The Research Group operates under the umbrella of the Faculty of Public Governance and International Studies of the Ludovika University of Public Service. They are convinced that the public understanding of science and the importance of relations between science and society is particularly relevant in today's information society: effective communication of science with policymakers, industry, market, civil society and between academic disciplines, shaping the role and trust of science in society; reducing the alienation from science in society.

Given the increasing relevance of disinformation, misinformation and information manipulation in the changing information, communication and media environment, this is a highly critical topic. Thus, the studies and practices on communication and relations between science and society require a complex approach. Founding members of the Research Group focus on the aspects of this complexity (such as science communication, information disturbance, academic capital and knowledge production, networking, media- and info-communication, digital culture, and socio-cultural impact of digital and artificial intelligence technologies), on the interconnection of these topics following the current international academic and professional trends, and aim to create a common platform for this network.

The Research Group's central theme in 2023 was the complex relationship between science, society and disinformation. On 13 October 2023, the Research Group organized the conference "Science and Disinformation: How can Science support Society against Disinformation?" which offered a wide range of topics, including the legal aspects of dealing with disinformation, the impact of AI on science communication, the impact of digitalization on experiential education; the relationship between disinformation and cyber security; conspiracy theories; and the relationship between evidence-based decision-making and science communication. The panel discussion at the conference focused on trust and distrust in science.

In the first paper, Bányász et al. aim to empirically analyze the cyberattacks that occurred in the context of the Russian-Ukrainian conflict between 2022 and 2023, specifically focusing on the impact of these attacks on civilian infrastructure and institutions.

Cholyshkina et al. analyze the use of artificial intelligence in optimizing education management processes. The study examines AI in education management and highlights innovative approaches introduced by AI that allow us to move from standard teaching and education management methods to individual and effective strategies.

Dodé and Falyuna explore how language elements of specialized and political communication are represented in parliamentary discourse. The results of their in-

vestigation can be relevant for studies on populist communication, the relationship between technocracy and democracy, and the methodological aspects of computer-assisted discourse analysis.

Falyuna et al. frame the insights of the panel discussion at the conference, which focuses on the issue of social trust and distrust in science, the logic of (science-related) conspiracy theories, the impact of digital culture and AI on science communication, disinformation, trust and credibility, and relations between science and business.

Paksi deals with the question of artificial intelligence as a materialist anthropomorphism. He investigates the problems and contradictions stemming from the materialist interpretation and presents a coherent emergentist concept of machines, in which machines are clearly unique between simple material things and living beings.

Singh's exciting essay concerns the technophilosophical underpinnings of the HBO TV show Westworld (2016–2022). The paper elicits several philosophical arguments, dilemmas, and questions concerning technology and artificial intelligence (AI). It discusses the dispute between memory and consciousness in the context of an artificially intelligent robot, delves into the issues of reality and morality for humans and AI, and finally, speculates on the potential of social interaction between sentient AI and humans.

Török, in his study, explains that the challenge of disinformation lies in the fact that digital technologies, platforms and social media have significantly subverted the previous operating mechanisms of the public sphere. However, clear principles and criteria for regulating public discourse have been developed in recent decades. This article summarizes the most representative European principles of freedom of speech that are highly relevant in the legal answer to disinformation. Fehér, in her article, summarises an experimental PhD. seminar that focused on the academic trends in the AI-driven synthetic worlds through risk assessments and ethical concerns and highlights the importance of comprehensive education in technology with a critical and ethical approach.

Lu's research concerns the Chinese perspective on artificial intelligence-generated content and copyright. Can pieces of artificial intelligence-generated content (AIGC) be considered works? Can AI be the author of AIGC? The paper seeks to provide a comprehensive and systematic analysis of the literature of Chinese scholars to sort out their different perspectives on the relevant issues.

Finally, Yao and Zhu tackle the willingness to adopt smart-ageing services, based on evidence from Anhui Province, China, taking a questionnaire survey of 1600 households. The results show that subjective norms, perceived usefulness and perceived ease of use are the key factors influencing the adoption intention of smart senior care services. Further, age has an impact on perceived usefulness and adoption intention, education has an impact on perceived ease of use and adoption intention, and monthly household income has an impact on perceived usefulness and adoption intention.

We wish you a pleasant reading.

Empirical analysis of the cyberattacks of the Russian–Ukrainian war

This research study aims to empirically analyze the cyberattacks that occurred in the context of the Russian–Ukrainian conflict between 2022 and 2023, with a specific focus on the impact of these attacks on civilian infrastructure and institutions. The data collection for this study is based on publicly available sources from the Cyber-Peace Institute, taking into account various types of incidents such as malware, distributed denial of service (DDoS) attacks, spam, information operations, and website defacements. The study employs a network theory approach to examine the structure and dynamics of incidents and campaigns, while additional statistical methods and trend analysis are used to assess sector-specific and geographic patterns, as well as changes in attack frequency and severity. The research aims to contribute to the existing literature on cyber warfare and to provide valuable insights into the cyber threats faced by civilian infrastructure and institutions during times of conflict.

Keywords: *Russian–Ukranian war, cyberattacks, network analysis, sector analysis*

Acknowledgments

This research has been carried out as part of project no. TKP2021-NKTA-51, implemented with support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NKTA funding scheme.

Author Information

Péter Bányász, University of Public Service
https://orcid.org/0000-0002-7308-9304
Adrienn Kiss, University of Public Service
https://orcid.org/0000-0003-4953-5367
Sándor Magyar, University of Public Service
https://orcid.org/0000-0002-6085-0598
Dávid Kiss, University of Public Service
https://orcid.org/0009-0004-2034-2734

How to cite this article:

Bányász, Péter, Adrienn Kiss, Sándor Magyar, Dávid Kiss. “Empirical analysis of the cyberattacks of the Russian–Ukrainian war”.
Információs Társadalom XXIV, no. 2 (2024): 9–32.
https://dx.doi.org/10.22503/inftars.XXIV.2024.2.1

1. Introduction

The impact of information technology (IT) has been growing not only in the lives of individuals but also in the operations of organizations. As far back as the 2000s, some research was already looking into the future of the information society. According to studies published in recent years on the information society, human knowledge has become the most important factor of development in the information society, and among capital goods, human resources have become the most prominent (Majó 2006). There was a time when the drivers of technological development were general education and technological innovation, but these drivers have disappeared over time (Karvalics 2007). With an increasing dependence on IT services in the digital ecosystem, the ease and convenience they provide are accompanied by a growing threat of service outages and cyberattacks. Information is power, and its importance is indisputable. The complexity of information is judged on subjective grounds, but information is a force that is nowadays used to launch various attacks (Krzanowski 2023). Given the well-defined target that IT systems and their users present, such attacks pose a significant challenge. During the Russia–Ukraine war, it became clear that these attacks could affect both local and global digital ecosystems, emphasizing the importance of digital services for the continuity of operations in the region (Aviv and Ferri 2023). While emerging and disruptive technologies such as artificial intelligence, autonomous aerial vehicles, and hypersonic aircraft offer significant benefits for civilian and military domains, their potential for malicious use also poses a significant threat. As a result, cyber operations are taking on an increasingly important role alongside conventional warfare, which itself is evolving with new technologies and tactics.

As IT becomes increasingly essential for modern society, the potential impact of cyberattacks on a nation's security also increases. This is why cyberspace is now recognized as an important operational domain in contemporary warfare, including in the Russian–Ukrainian conflict. Hybrid warfare strategies further underscore the importance of these domains, as they provide a viable battlefield for all parties to achieve their strategic objectives, both in peacetime and in times of conflict. The choice of targets for cyberattacks can support military objectives, even at the strategic level, making it necessary for both the military and civilian sectors to develop appropriate protection against such attacks. The use of cyber operations in the Russian–Ukrainian conflict also revealed the capacity of cyberspace to conceal attackers' identities. The ability to collaborate with non-state actors can be an advantage for cyber attackers. On the Russian side, intelligence services have been primarily responsible for these activities, with support from non-state actors, including "patriotic" hackers and private companies (Miron and Thornton 2024).

In cyberspace, hidden attackers from third countries can be recruited or voluntarily join. Due to the difficulty of attribution in cyberspace, as a result, electronic information systems, including those in government agencies, public services, financial institutions, and critical infrastructure, have become established targets for state actors. These actors may cause significant damage through the use of hacker

groups funded by state actors or through services available on the dark web, such as "cybercrime as a service."

The wide range of cyberattack tools available to attackers provides them with significant opportunities. Electronic information systems' vulnerabilities can be exploited through various tactics, techniques, and procedures that are easily deployable. The availability of hacking tools is increasing, enabling a range of outcomes beyond attacks on the target country's territory, such as accessing adversary data, rendering systems inaccessible, and conducting reputation attacks. In addition, psychological operations against soldiers and the population accessible via cyberspace, particularly in the area of influence (disinformation, misinformation, malinformation), can also be highly effective and yield easier successes. As a result, many countries are dedicating significant attention to the development of offensive cyber-operational capabilities. The Russian–Ukrainian conflict has demonstrated that the cyber domain is evolving alongside changes in conventional warfare, and can deliver appropriate results on the battlefield, in the hinterland, and even more broadly in the international environment. In certain cases, it can even serve as a complement to or substitute for military operations.

In addition to the challenge of effectively attributing cyberattacks in the legal realm, such attacks may pose risks even in the absence of open armed conflict (Fiala and Worrall 2024). The Russian invasion of Ukraine serves as a prime example of a gray-zone contest in which nation-states' and other actors' main narratives – particularly their changing intensity and tone – affect diplomatic, economic, and military domains (Hoffman and Hofmann 2018). The political and military gains made in the gray-zone area are significant and contribute to the realization of conflict objectives without relying on arms. The scope of the impact of activities in cyberspace raises important questions about the development of international law for managing cyber conflicts and highlights cases that may require military responses.

The present study employs a comprehensive methodology to analyze documented cyberattacks during the Russian–Ukrainian conflict between 2022 and 2023. Since 2013, Russia has carried out various cyber operations targeting Ukraine on multiple occasions (Lunn 2023). Antecedent to the invasion, scholars had foreseen a widespread cyber conflict; however, the magnitude of the recorded cyberattacks has been limited, plausibly because of Ukraine's enhanced cyber defenses established through collaboration with Western allies (Lonergan et al. 2023; Kostyuk and Brantly 2022).

The research objectives were twofold:

- to conduct an empirical analysis of the characteristics and dynamics of cyberattacks that were documented during the Russian–Ukrainian conflict between 2022 and 2023;
- to identify the sector-specific attacks, as well as assessing the changes in the frequency of said attacks.

Consequently, the research study has identified the following research questions to be addressed:

- What were the most commonly occurring types of cyberattacks during the Russian–Ukrainian conflict?
- What is the distribution of cyberattack types between different sectors?
- How did the frequency of cyberattacks change during the course of the conflict?

In conclusion, the research study proposes the following hypotheses:

H1: Cyberattacks reach their peak during the winter months.

H2: The energy sector has been the most affected by cyberattacks due to its critical role in civil infrastructure.

H3: In terms of frequency, distributed denial of service (DDoS) attacks are the most common type of cyberattacks.

H4: The frequency of cyberattacks conducted by state-sponsored attack groups is significantly higher when compared to those perpetrated by smaller attack groups.

2. Methods

The data under analysis was gathered through the Cyber Conflicts project of the CyberPeace Institute (CyberPeace Institute 2023). The Institute was established with the aim of mitigating the negative impacts of cyberattacks, aiding vulnerable communities, and promoting responsible conduct in cyberspace. Cyber Conflicts concentrates primarily on cyberattacks that occurred during the Russian–Ukrainian war. Any incident that falls within the ambit of cyberattacks and operations defined by the CyberPeace Institute, especially those carried out by a threat actor with the intention of disrupting, disabling, destroying, manipulating, surveilling, controlling, or extracting computing environments/infrastructure and/or data using a computer network or system, is covered. These incidents include but are not limited to hack and leak, where the attacker aims to hack into the target's data and then steal and use critical information from the victim (Traficom 2023); DDoS, in which an attacker floods a server with internet traffic to prevent users from accessing related online services (Fortinet 2024); and defacement, which is an attack on a website that alters its informational content or visual appearance (Kaspersky Lab 2024).

The Institute refers to incidents as campaigns if they satisfy all of the following conditions:

- The incident is linked to the same threat actor and happened within an eight-hour period targeting more than two entities at the same time within the same country or one entity more than twice.
- The incident targeted the same entity over two consecutive days.
- The incident that targeted more than two entities in more than two countries is linked to the same threat actor using the same modus operandi.

Where a threat actor targeted entities in various sectors during a campaign, the Institute creates an incident record for each sector, not for each targeted entity. The primary focus for data collection is on cyber incidents that affect institutions and facilities in the sectors listed in the United Nations International Standard Industrial

Classification of All Economic Activities. Data collection pertains to cyber incidents in the context of the Russian–Ukrainian war, including incidents in Ukraine, the Russian Federation, and other countries. It is worth noting that confirming incidents, especially in the Russian Federation and Belarus, presents particular challenges.

Although it is not always possible to determine if a specific cyberattack or operation was carried out with political, military, activist, and/or strategic motives related to the conflict, this forms the basis of the data collection scope. For instance, incidents are documented relating to the leak of data from Russian organizations committed in the name of pro-Ukrainian activism, the disruption of services after a country took a public political or economic position on the conflict or provided military aid, and collateral damage in a third country that spills over from an incident originally targeting an entity in either the Russian Federation or Ukraine.

For the timeline's purpose, the Institute collects information on cyberattacks that is available publicly (open source) by monitoring news/media outlets, government, cybersecurity companies, computer emergency response teams (CERTs), and civil society organizations' reports, advisories, blogs, and social media feeds, among other sources. Every identified incident and associated content is reviewed by at least two internal analysts, and wherever possible, the incident is linked to at least two distinct sources of information. The Institute continuously scans for information on previous incidents to update the timeline on societal harm and attribution, which is often reported significantly after the actual incident. As publicly available data is relied upon, documented cyberattacks have been assigned a classification of certainty based on the reliability of the information source. The classification levels are as follows:

- Confirmed: Attacks in this category are based on official government reports/records, official press releases by the targeted organization, official letters addressed to customers by the target organization or the government, or social media communication by the targeted organization. If an incident has been self-attributed by a threat actor, and a government entity has confirmed the attack, it will be classified as confirmed.
- Probable: Attacks in this category are based on media reports of a press conference by the targeted organization, social media communication by the targeted organization, or quotes from the targeted organization's staff in media articles. If an incident has been self-attributed by a threat actor, and the attack has been corroborated by a third party through independent research or the analysis of stolen data, this is also classified as a probable incident. Incidents identified and reported on as a result of a technical/forensics investigation will also be classified as probable.
- Possible: Attacks in this category are based on media reports with no direct reference to primary source information. This can be in the form of a news article that mentions a letter sent to patients or a blog post that references a statement published by the targeted organization, but no direct record of this material is available. This category also includes data published by a threat actor online with no further corroborating information.

Based on the gathered data, 187 cases were confirmed, 117 were classified as probable, and 655 were deemed to be possible. The CyberPeace Institute does not

publicly document data related to "hearsay" incidents, which contains uncorroborated information originating from a third party, that is, as a result of media reporting of the allegation by a third party (CyberPeace Institute 2023).

Our methodology for investigating cyberattacks integrates five main techniques – time-series analysis, trend analysis, heat-map visualization, cluster analysis, and network analysis. Time-series analysis is used to examine the temporal distribution of cyberattacks, while trend analysis focuses on the evolving nature of attacks in different sectors and the tactics used by threat actors. Heat-map visualization provides a clear and intuitive picture of the concentration and distribution of cyberattacks, while cluster analysis groups similar types of attacks based on various attributes. Finally, network analysis examines the relationships and interactions between different entities involved in cyberattacks. Together, these methods offer insights into temporal trends, sectoral vulnerabilities, geographic hotspots, common attack patterns, and the complex web of relationships between different stakeholders in the cyber domain, serving as the foundation for developing informed cybersecurity strategies and policies.

3. Results

As an initial step in the time analysis, we examined the monthly frequency of attacks between 13 January 2022 and 31 December 2023. The cyberattacks that occurred during the initial period of the conflict are as follows:

- January 2022: 6 attacks
- February 2022: 33 attacks
- March 2022: 61 attacks
- April 2022: 47 attacks
- May 2022: 19 attacks.

The aforementioned data indicates a significant increase in the number of attacks in February, with a peak in March, followed by a decline in April and a further decrease in May.

The frequency of cyberattacks exhibits a periodicity that can be delineated into approximately six-month cycles. Specifically, the second cycle spans from May 2022 to the conclusion of November 2022, whereas the third cycle encompasses December 2022 to April 2023. Subsequently, a diminishing trend is observed during the fourth cycle from April 2023 to January 2023. In each cycle, a discernible pattern emerges, characterized by a gradual increase in the frequency of attacks, followed by a peak and a subsequent decline. This pattern can be attributed to the constraints imposed by the attackers' capacity and resource limitations, which allow for approximately six months of preparation and activation. When examining the trend in the number of attacks averaged over time, a minimal upward trend is observed, indicating that the attackers are either unable or unwilling to execute large-scale attacks even after a considerable length of time has elapsed.

Subsequently, we employed visualization techniques to better understand the dynamics and possible patterns of monthly attacks over the entire period (see Figure 1).

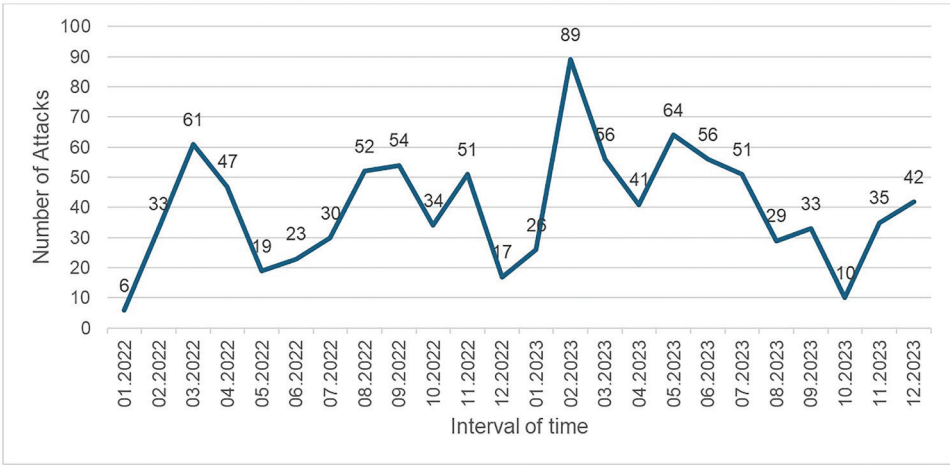

Figure 1. Distribution of cyberattacks during the period of analysis (own edition based on CyberPeace Institute database)

In total, 959 cyberattacks were identified during the period of analysis. In February 2022, as the Russian military advanced toward the Ukrainian border, experts in cybersecurity began to envision the potential employment of cyberattacks by the Russian government to undermine the Ukrainian defenses. Various governmental agencies and private sector entities predicted that the Russian forces would unleash a rapid and devastating series of electronic assaults aimed at disrupting the country's critical infrastructure, including power plants and air traffic control networks, thereby causing widespread destruction. However, while cyberattacks from Russian sources have indeed been a factor in the conflict, their impact thus far has not been as substantial as anticipated by some analysts (Givens et al. 2023).

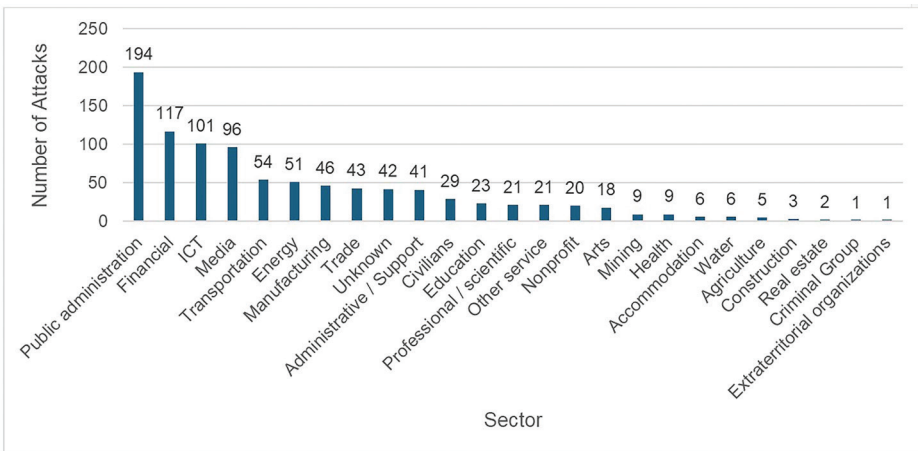

Figure 2. Distribution of attacks by sector (own edition based on CyberPeace Institute database)

We subsequently examined the sector-specific distribution of attacks to ascertain which sectors were most impacted during this period. The findings are depicted in Figure 2.

The aforementioned distribution indicates that the public administration, finance, information and communication technology (ICT), and media sectors were the most impacted by attacks during the period under analysis. These sectors offer critical infrastructure and services, which may elucidate why they were coveted targets for attackers.

The next part of our analysis focused on the types of attacks. First, we examined the frequency of attack types in the general context (see Figure 3), and second, we looked at the distribution of these attack types across sectors, which, in view of the attacks that have occurred, we limited to the 10 sectors that have suffered the most attacks (see Figure 4).

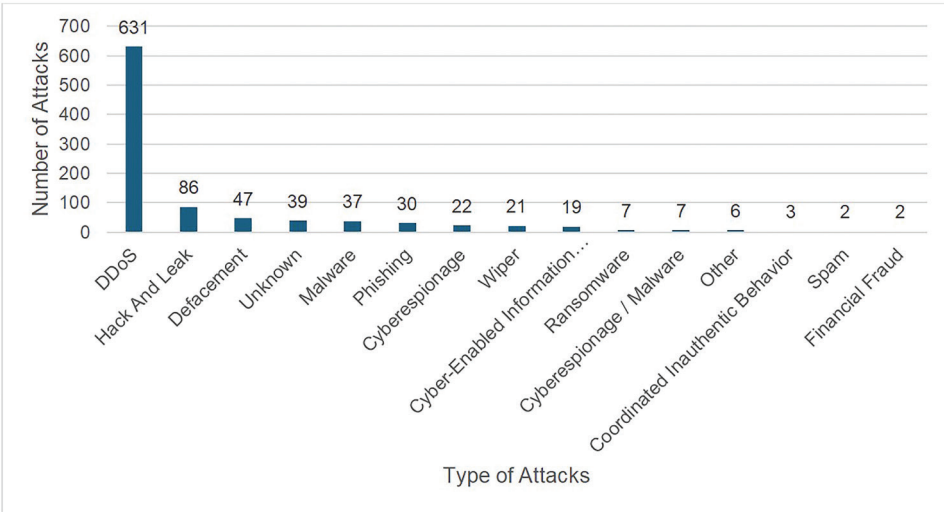

Figure 3. Distribution of attack types
(own edition based on CyberPeace Institute database)

The aforementioned distribution demonstrates that DDoS attacks (631) were the most prevalent during this period, substantially outnumbering other types of attacks. Hack and leak (86) and defacement (47) attacks also constituted a considerable number of incidents, while the other attack types occurred relatively less frequently. This analysis can aid in comprehending the attack techniques that attackers favor and emphasize the priorities for cybersecurity defenses.

Figure 4 reveals that the public administration (194) sector experienced the highest number of security breaches, succeeded by the financial (117) and ICT sectors (101). The DDoS attacks emerged as the most prevalent form of security intrusion across all the three sectors (public administration 118, financial 95, ICT 65). The term "sectors" encompasses not only traditional sectors but also criminal groups. This is

due to the fact that multiple hacktivist groups, which support either Ukraine or Russia, have launched attacks against each other, resulting in the exposure of sensitive information about the groups' operations or members, and in some cases leading to arrests. As evidence, three indictments in distinct federal jurisdictions have been unveiled, accusing multiple Russian cybercrime actors associated with the Trickbot malware and Conti ransomware stratagems (2023).

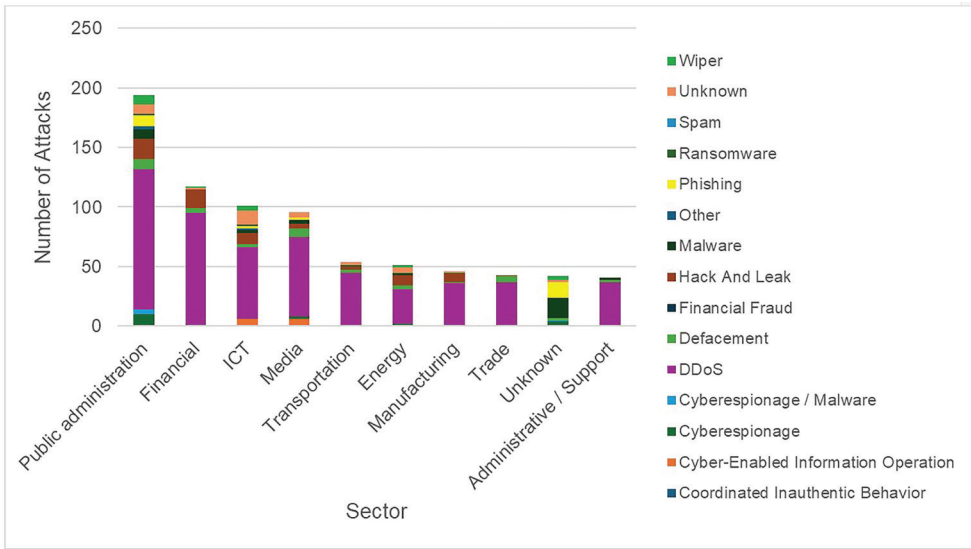

Figure 4. Distribution of the 10 most attacked sectors by attack types (own edition based on CyberPeace Institute database)

The graphical representation in Figure 5 illustrates the top 10 attack groups along with their frequently employed attack types. The data reveals that certain attackers prefer a diverse range of attack methods, whereas others concentrate on specific techniques. The majority of the attacks can be linked to the People's Cyber Army (249), known to have ties with Russia, which surpasses the number of attacks conducted by the Ukrainian IT Army (85), the second most active group, by almost three times. It is noteworthy that the IT Army occupies the third position in the chart, as the perpetrator of 107 attacks could not be ascertained. Despite being a group linked to Russia, the People's Cyber Army has surprisingly launched attacks on four occasions against Russian targets. Furthermore, six identified groups have attacked both Russian and Ukrainian targets, namely Anonymous Russia, Phoenix, Mirai, KillNet, and GURMO. With the exception of GURMO, the majority of these groups have primarily targeted Ukraine and are commonly associated with Russia. As GURMO is the Ukraine's Military Intelligence Service, consequently the one attack executed a cyber-enabled information operation against Russian television channels broadcasting in Crimea, which targeted Ukraine. Crimea has been occupied by Russia since 2014 (Lunn 2023).

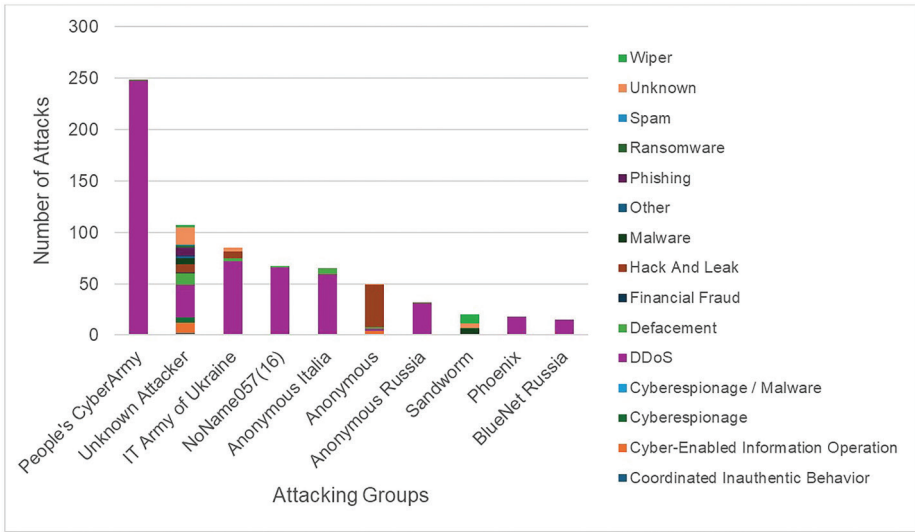

Figure 5. Distribution of the 10 most active attacking groups by attack types (own edition based on CyberPeace Institute database)

The conflict between Ukraine and Russia has spurred the involvement of young hackers in hacktivist groups, which have carried out cyberattacks against Russia (Security Alliance Limited 2022). The cyber facet of the conflict has observed a considerable quantity and heterogeneity of cyber occurrences; nonetheless, the participation of non-state actors and their readiness to carry out cyberattacks outside the domain of military operations had not been predicted (Chukhua 2023). The initial majority of pro-Ukraine groups has shifted over time, with pro-Russia groups taking the lead. This shift can be attributed to various factors, including the establishment of the IT Army by the Ukrainian government. Furthermore, the outrage of pro-Ukraine groups regarding other geopolitical events has diverted their attention to other political targets. Additionally, frequent rivalries among pro-Ukraine groups have led to divisions and the suspension of their operations. Notably, some groups supporting Russia have been identified as Russian cybercriminal groups. It is widely believed that there exists an unspoken agreement between Russian national security services and Russian cybercriminal groups, whereby cybercriminals are permitted to operate, provided that they do not target Russian interests (Miron and Thornton 2024). Alternatively, if the interests of the Russian state require it, these groups may carry out their activities in accordance with Russian interests.

As per the Ukrainian government, the IT Army boasts over 200,000 active members. However, this number is believed to be overstated. Nonetheless, a significant number of cyber volunteers are involved on both sides of the conflict (Willett 2022), with most of them aged between 13 and 25 years. The duration and outcome of the war remain uncertain. However, one of the primary concerns is the future actions of the hundreds of thousands of cyber volunteers who have participated or are participating in the attacks. These young individuals have acquired skills in penetrating

secure systems and concealing their identities. While there is a great need for ethical hackers to ensure a safer cyberspace, there is apprehension that most of these volunteers will opt for the easier route of pursuing illicit activities for financial gain (Feledy and Virág 2022).

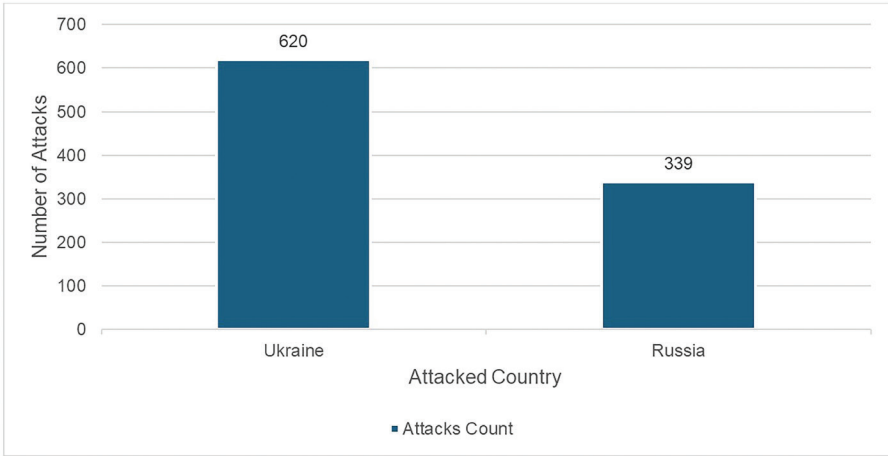

Figure 6. Distribution of attacks by country
(own edition based on CyberPeace Institute database)

Figure 6 depicts the distribution of attacks targeting both Russia and Ukraine, while Figure 7 provides a visual representation of the distribution of the attacks over time.

During the period under review, Ukraine suffered almost twice as many cyberattacks as Russia.

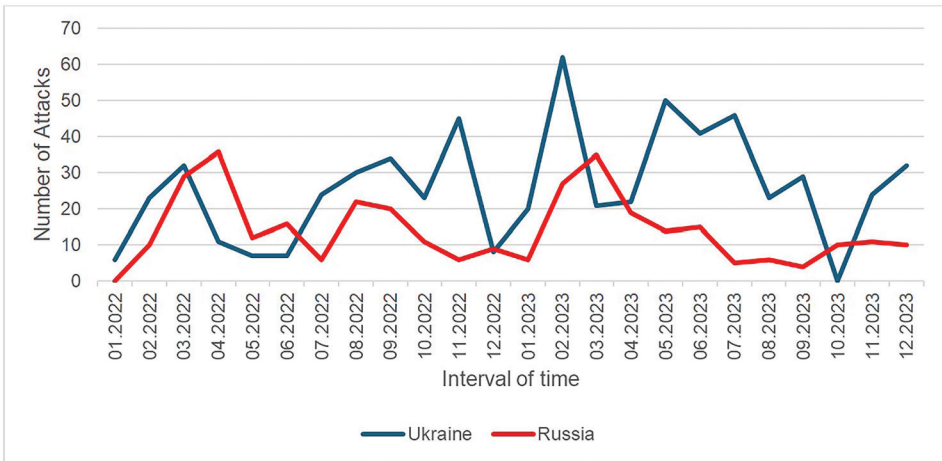

Figure 7. Distribution of attacks by country over time
(own edition based on CyberPeace Institute database)

As the graphical representation shows, the month of February 2023 witnessed the highest number of cyberattacks in Ukraine among all the months, while October 2023 was the only month devoid of any such incidents.

Examination of the dispersion of affected industry sectors by nation accentuates the extensive array of economic and societal sectors that have been impacted by cyber intrusions in both Ukraine and Russia (see Figure 8). The findings indicate that:

In the instance of Ukraine, a majority of the attacks were directed toward the public administration, energy, trade, and transport sectors, denoting the strategic significance of these domains in the conflict.

In Russia, the administration, energy, and transport sectors were the most prominently impacted; however, it is of significant note that the agricultural and mining sectors were also targeted.

These results can help us better understand the dynamics of cyber conflict and how attackers choose their targets in each country.

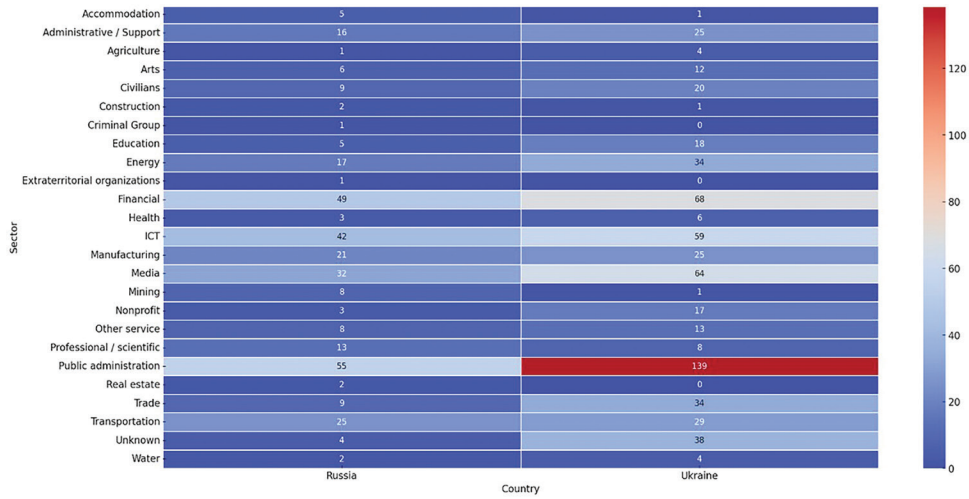

Figure 8. Distribution of sectors affected by the attack by country (own editing in Python, based on CyberPeace Institute database)

The presented heat map provides a detailed analysis of the distribution of cyberattacks across various sectors in Russia and Ukraine. The color saturation within each individual cell represents the frequency of attacks, providing a clear and concise evaluation of the affected sectors in both nations. The findings of the analysis confirm the notable targeting of specific sectors, including public administration, energy, and transportation in both countries, which was previously observed. The heat map also allows for the identification of additional intricacies, such as the relative susceptibility of different sectors and variations between the two countries in terms of the most targeted sectors. The data highlights the need for increased cybersecurity measures in these specific sectors, particularly in public administration, energy, and transportation, to mitigate the risks of cyberattacks.

Furthermore, we have exhibited the relationships among attackers and the affected industries (Figure 9), as well as between attackers and their frequently used attack techniques (Figure 10), through the use of heat maps. Such a graphical representation can potentially facilitate comprehension of the interconnections between perpetrators and their intended victims, as well as their modus operandi.

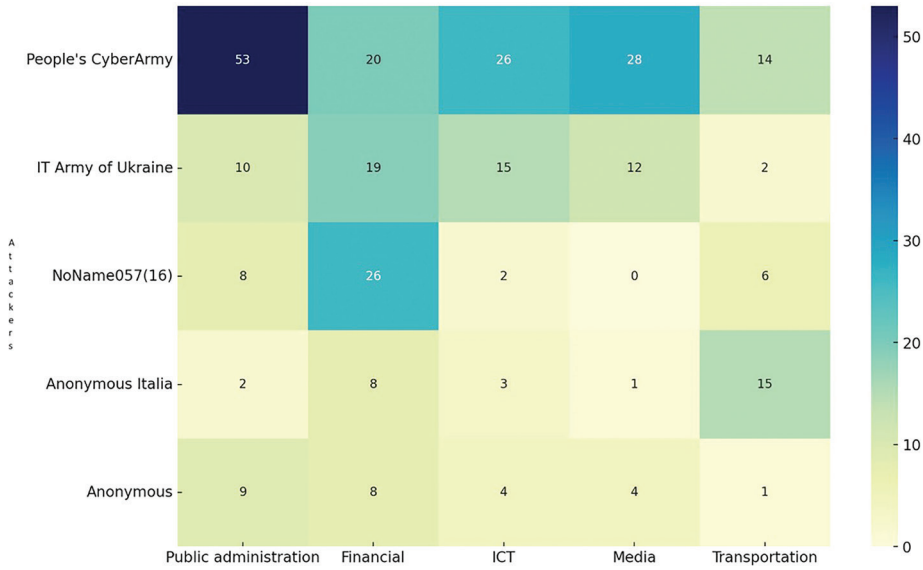

Figure 9. Distribution of sectors most frequently targeted by attackers (own editing in Python, based on CyberPeace Institute database)

Figure 9 presents a heat map that provides a visual representation of the frequency of cyberattacks in different sectors. The chromatic intensity within each discrete unit signifies the rate of occurrences of assaults, with darker shades indicating a higher frequency of attacks. The results of the heat-map analysis reveal that public administration is the most vulnerable sector to cyberattacks, followed by the financial, ICT, and media sectors. The correlation between the previous analyses and the heat-map data highlights the need for increased cybersecurity measures in these sectors to protect against the growing threat of cyberattacks. Figure 10 corroborates the previous analyses, indicating that DDoS attacks are the most prevalent among all types of cyberattacks, and they are predominantly carried out by a specific group of threat actors. The figure provides a clear visual representation of the distribution of attack types and the groups responsible for them. The data reveals that DDoS attacks stand out significantly, with a considerably higher frequency than other forms of cyberattacks. Furthermore, the figure illustrates that a select group of attackers is responsible for the majority of these attacks, which highlights the need for targeted measures to combat this specific threat. It is clear that DDoS attacks pose a significant risk to organizations, and identifying the responsible groups should be a priority in developing effective strategies to mitigate the risk of such attacks.

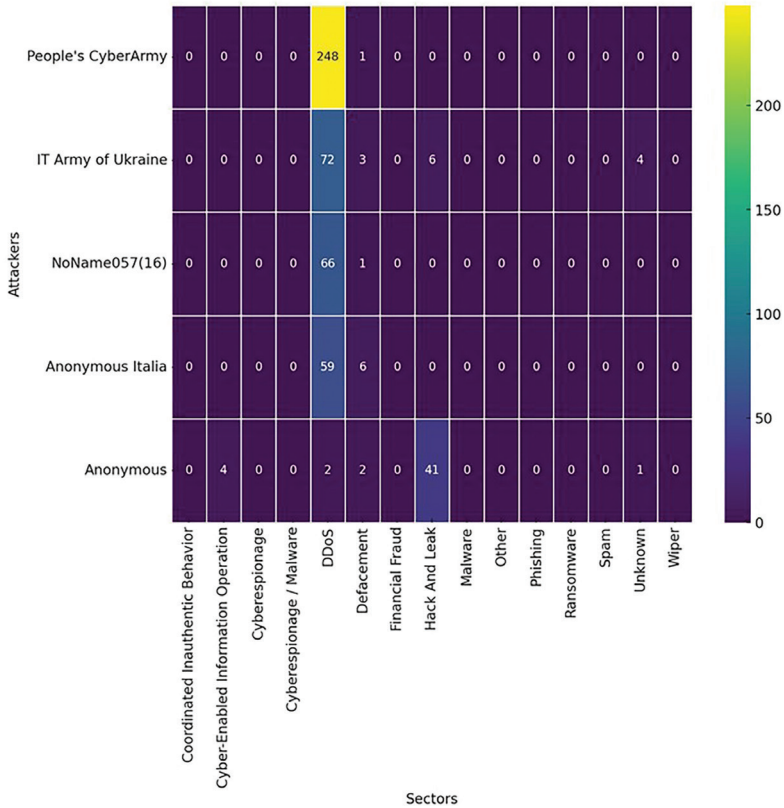

Figure 10. Distribution of the most common types of attacks used by attackers (own editing in Python, based on CyberPeace Institute database)

In order to attain a more comprehensive understanding of the attacks' nature, it is advisable to employ supplementary methodological approaches in analyzing the accessible data. Network analysis offers the possibility to visually represent the connection between the attackers and the attacked sectors, therefore enabling us to draw further inferences. To establish the foundation of our network, we must first define its structural principles. Within this network, we interpret all attackers and attacked sectors as nodes. Specifically, there are 25 sectors and 95 attackers represented as nodes in this case. It is crucial to note that we do not segregate the Russian and Ukrainian sectors from each other, but rather consider the direction of the attack, i.e., the sector itself, as a node in the network, following a general principle. To elaborate, the network's disassociation from the targeted country of attacks is crucial, as it enables us to focus on the attacks' essence. Extracting data from the database, we treated all attacks directed from an unidentified attacker or toward an unidentified target as a single point, labeled as Unknown Sector and Unknown Attacker. This is a critical component of our analysis and warrants emphasis. The

next fundamental element of the network is its connection scheme, which entails the determination of links or edges. A link between two points is established when an attacker has targeted that specific sector at least once during the analysis period.

A representation of the fundamental network based on these principles is depicted in Figure 11. All the network diagrams were generated employing the Gephi data visualization software.

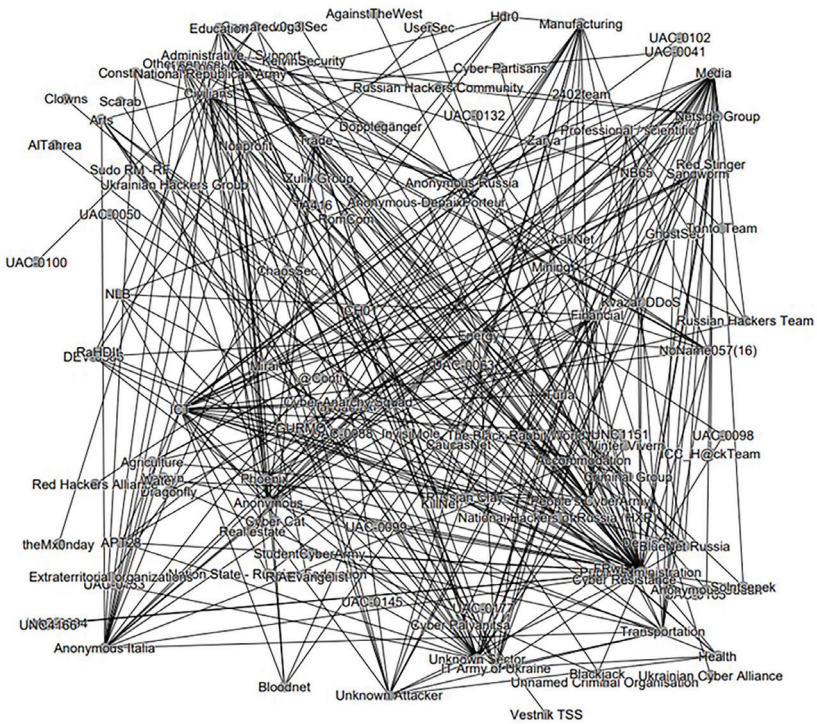

Figure 11. Network of cyberattacks by sector and by attacker (own editing in Gephi, based on CyberPeace Institute database)

The network structure, in its present state, remains incomplete. In addition to defining connections between nodes, we must also determine whether or not the edges in our network are directed. As attacks toward a sector have an evident directionality (given that the sector cannot retaliate), we have designated our network as directed, with links always originating from the attacker and terminating at the sector. Both analysis and interpretation can be further enhanced through the color-coding and weighting of edges and nodes. While many criteria can be used to weigh the relationship between nodes, such as the extent of damage or the resources utilized to execute an attack, this study utilized a coherent system where the weights represent the number of attacks. In other words, each additional attack toward a given sector resulted in a one-unit increase in the attacker's value toward that sector, regardless

of whether the attack targeted a Russian or a Ukrainian sector. For nodes, a similar weighting system was employed based on the number of degrees, i.e., the number of edges entering or leaving a node. The size of a sector node in the network corresponds to the number of attacks it sustained, while an attacker node's size reflects the number of attacks it executed. Building on the edge weighting system, we also factored in edge weight when determining node size rather than simply considering the number of edges. Consequently, the size of each node was determined based on its respective weighted degree.

The introduction of these criteria has resulted in a significant transformation of our network, as demonstrated in Figure 12.

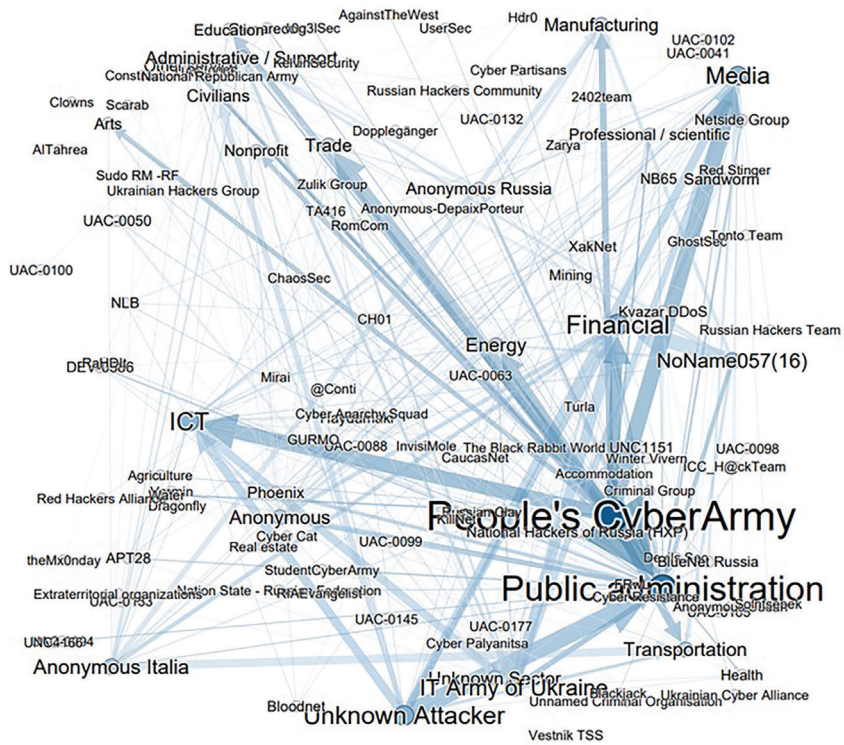

Figure 12. Distribution of the weighted network of cyberattacks by sector and attacker (own editing in Gephi, based on CyberPeace Institute database)

Figure 12 displays the network nodes in the same position as in Figure 11; however, the application of color-coding and weighting has emphasized the network's crucial nodes, namely, the attackers with the most attacks and the sectors that sustained the most damage.

Following the weighting process, it is crucial to identify the network's primary features. The network is classified as a bipartite graph, meaning that its nodes can

be categorized into two distinct sets (attackers and sectors), where nodes within each set are never connected to each other, but solely to nodes in the other set. To illustrate this grouping of nodes, an additional criterion can be incorporated, which, when used in conjunction with the visualization tools and layouts provided by Gephi, can yield a representative depiction of the network's nodes in terms of both color-coding and positioning (as evidenced in Figure 13).

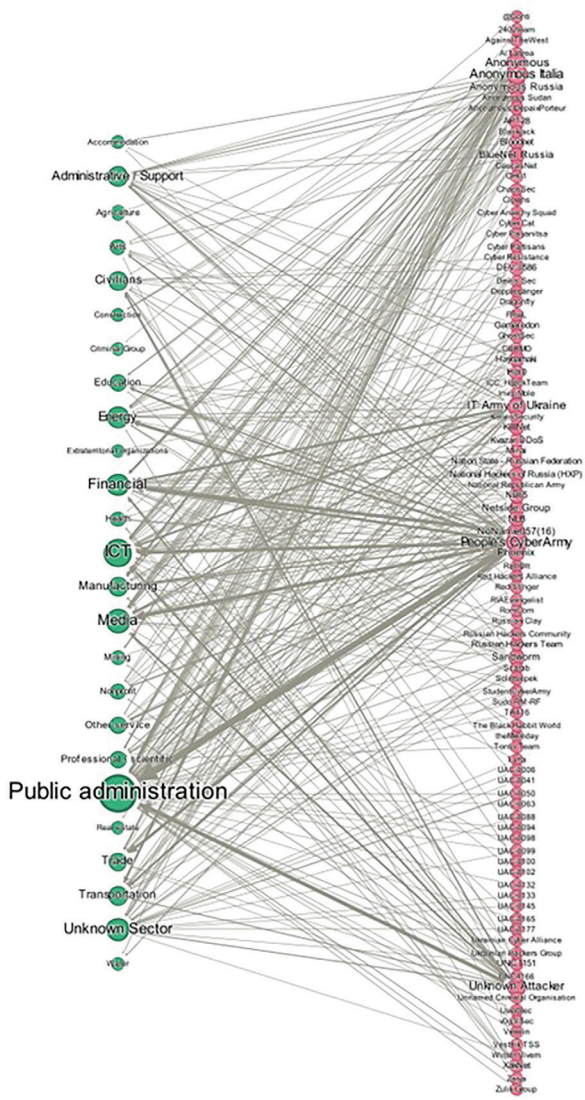

Figure 13. Network of cyberattacks, categorized by both sector and attacker, with weights assigned to each node and edge
(own editing in Gephi, based on CyberPeace Institute database)

It is noteworthy that the network comprises two distinct components, implying that points within the network are not entirely traversable; in mathematical terms, it is not fully walkable. One of these components pertains to the attack orchestrated by the @Conti group, which was specifically targeted at a criminal group. This attack was unrepeatable and did not extend to other sectors during the observation period, nor did it involve any future assaults on the criminal group, which has been categorized as a sector. Therefore, these two points can be regarded as entirely independent in the network.

Given that the diameter of the bipartite graph, which represents the distance between the two farthest points in the network, is 1, and the average path length, which denotes the average distance between the points, is also 1, this statistical measure does not offer any additional insights into the current network. However, the distribution of the weighted degree number indicates a scale-independent pattern (Jeong et al. 2000), signifying that a majority of the nodes have only one connection, while the number of nodes in the network decreases as the number of connections increases (Figure 14). Specifically, 43.16% of the attackers are found to have only one connection, 71.58% have fewer than six connections, and only 13.68% of the attackers have more than 10 connections.

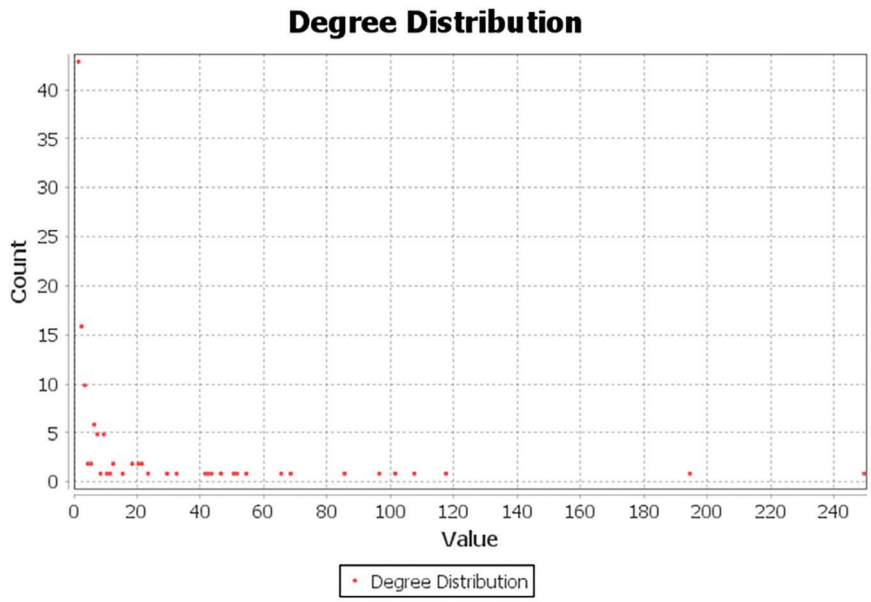

Figure 14. Distribution of the weighted degree number of points in the network (own editing in Gephi, based on CyberPeace Institute database)

In the context of the network's content, the aforementioned observations suggest that the cyberattacks examined are often sporadic and occasional, orchestrated by a single organization, and that just a few organizations have undertaken systematic and organized attacks targeting one or more sectors during the given period. From a

numerical standpoint, this implies that the top 10 organizations responsible for the most attacks constitute over 70% of the total attacks—73.9% to be exact.

In order to gain a more comprehensive understanding of the key components within the network, an alternative visual tool in Gephi is employed. The Force Atlas layout organizes the points in the network based on their weighted degree number, thereby highlighting the strategically significant elements of the network (as depicted in Figure 15).

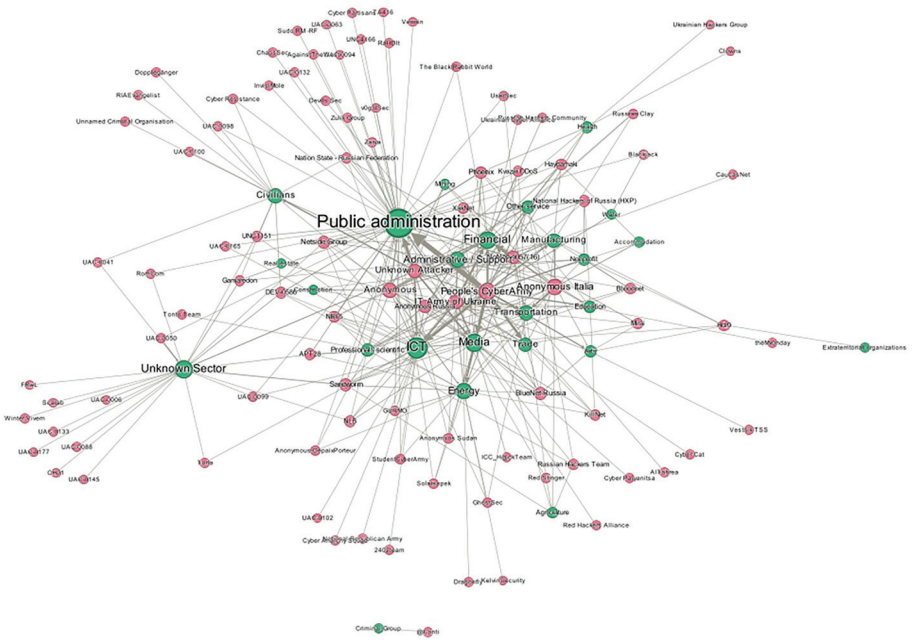

Figure 15. Weighted network of cyberattacks by sector and by attacker, focusing on the key elements of the network
(own editing in Gephi, based on CyberPeace Institute database)

Although the centrality, which quantifies the role and significance of each node in the network, can be computed for bipartite graphs, it does not provide relevant information such as distance and diameter. However, the distribution of the weighted degree number can be leveraged to sort the network based on the nodes with the highest weighted degree number. In this regard, Gephi utilizes this criterion to center the network on the sectors that have been consistently targeted by a greater number of cyberattacks (in descending order of weighted degree: public administration—194, financial—117, ICT—101, media—96, transportation—54) as well as the attackers that have carried out the most attacks during the observation period (in descending order of weighted degree: People's Cyber Army—249, Unknown Attacker—107, IT Army of Ukraine—85, NoName057(16)—67, Anonymous Italia—65). This information facilitates identification of the most targeted sectors and the most active attackers. Notably, Gephi's ordered graph layout illustrates the network's most

significant points from left to right, based on their weighted degree numbers (refer to Figure 16).

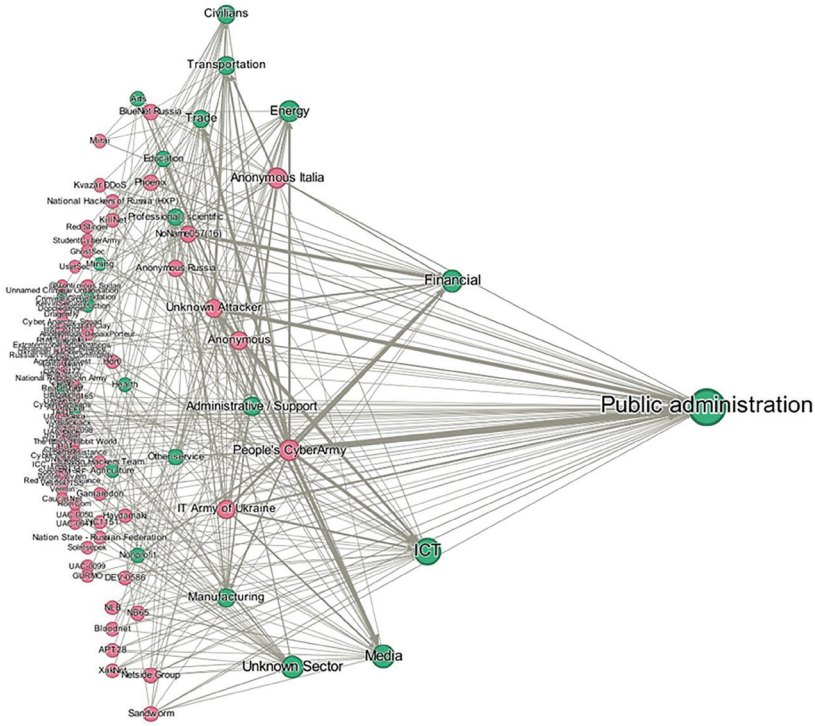

Figure 16. Weighted network of cyberattacks grouped by sector and by attacker, with the network's priority points ranked from left to right (own editing in Gephi, based on CyberPeace Institute database)

However, in order to investigate the characteristics and patterns of the attacks, it is necessary to cluster the points by computing their clustering indices. Gephi facilitates this computation and the resulting network map that visualizes the clusters with ease (as shown in Figure 17).

The program allocates a "branch" to each cluster within the network, resulting in a total of six separate clusters. One of these comprises the two nodes mentioned earlier, representing a distinct component of the network. In the remaining five clusters, it is evident that each cluster terminates in a sector with a high degree number (i.e., the farthest point on the branch from the center of the network), which is also the highest degree number point within the cluster. Subsequently, the points gradually exhibit fewer degree numbers as they approach the center of the network. The primary principle underlying clustering is to group points that are most strongly connected to each other. Thus, in the present network, all the most connected sectors and attackers are grouped together in a cluster. To elaborate, the most frequent

attacker targeting a sector is placed in the same cluster. If an attacker targets multiple sectors and is the most dominant among them, the corresponding sectors are also clustered together.

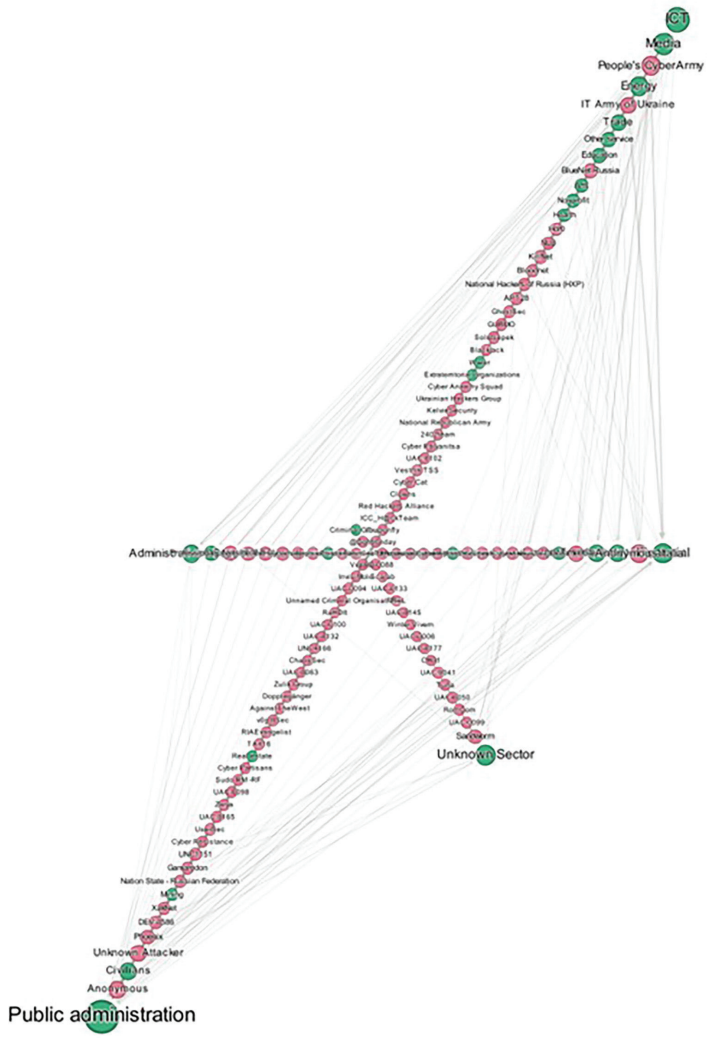

Figure 17. Weighted network of cyberattacks by clusters (own editing in Gephi, based on CyberPeace Institute database)

Clusters offer insights into the nature of attackers and their attacks by highlighting the most vulnerable sectors and those that are collectively susceptible to the same group of attackers. This, in turn, reveals the characteristics and patterns of the attacks and the attackers perpetrating them.

4. Discussion

In contemporary society, the crucial role of IT services in maintaining the operation of critical infrastructures has been widely acknowledged. As a result, cyberattacks targeting IT services have the potential to cause significant disruption to critical infrastructure operations. The Russia–Ukraine conflict has provided a stark illustration of the impact of cyberattacks on critical infrastructures and the security measures available to safeguard these systems. Against this backdrop, the present study aims to analyze the documented cyberattacks that occurred during the Russia–Ukraine conflict between 2022 and 2023, identify sector-specific patterns of cyberattacks, and assess any changes in the frequency and severity of these attacks. To achieve this research objective and test the formulated hypotheses regarding the sectors most vulnerable to cyberattacks and the potential frequency and severity of these attacks, a range of scientific research methods was employed.

Drawing on the conducted studies, our results are as follows:

R1: The data analysis indicates a semiannual periodicity in the frequency of attacks during the study period. It can be inferred that the winter season of 2023 witnessed the highest number of attacks, distributed monthly, compared to the other periods under consideration.

R2: As the sector-wise analysis shows, the majority of the cyberattacks were directed toward the public administration sector.

R3: The analysis conducted during the reviewed period predominantly identified DDoS attacks.

R4: The results of our research indicate that state-sponsored groups have perpetrated cyberattacks with greater frequency and diversity.

This paper aims to enhance our understanding of attackers' preferences for different methods of cyberattacks and to identify the most critical sectors that require protection. Additionally, it provides valuable insights into the attack patterns and networks of attackers, while also offering a forecast of the attack trends likely to emerge during the years 2022 and 2023 of the Russia–Ukraine conflict.

References

Aviv, Itzhak, and Uri Ferri. "Russian-Ukraine Armed Conflict: Lessons Learned on the Digital Ecosystem." *International Journal of Critical Infrastructure Protection* 43 (2023): 1–31. https://doi.org/10.1016/j.ijcip.2023.100637

Chukhua, Ilona. "Russian Aggressive Cyber-Policy during Russia-Ukraine War." In *Cyber Security Policies and Strategies of the World's Leading States*, edited by Nika Chitadze, 224–238. Hershey, PA: IGI Global, 2023.

CyberPeace Institute. 2023. "Cyber Attacks in Times of Conflict." Accessed March 10, 2024. https://cyberconflicts.cyberpeaceinstitute.org/

CyberPeace Institute. 2023. "Data & Methodology." Accessed March 10, 2024. https://cyberconflicts.cyberpeaceinstitute.org/

Feledy, Botond, and Csaba Virág. "An Assessment of Cyber Volunteer Groups in Interstate Conflicts and Their Impact on Public Policies." *Scientia et Securitas* 3, no. 1 (2022): 12–18. https://doi.org/10.1556/112.2022.00091

Fiala, Otto C., and Jim Worrall. "Unconventional Warfare in the Information Environment." In *Great Power Cyber Competition: Competing and Winning in the Information Environment*, edited by David V. Gioe and Margaret W. Smith, 157–172. Abingdon, UK: Routledge, 2024.

Fortinet. 2024. "What Is DDOS Attack?" Accessed March 10, 2024. www.fortinet.com/resources/cyberglossary/ddos-attack

Givens, Austen, Max Gorbachevsky, and Anita Biernat. "How Putin's Cyberwar Failed in Ukraine." *Journal of Strategic Security* 16, no. 2 (2023): 96–121. https://doi.org/10.5038/1944-0472.16.2.2099

Hoffman, Mark, and Martin O. Hofmann. "Challenges and Opportunities in Gray Zone 'Combat'." In *Advances in Cross-Cultural Decision-Making: Proceedings of the AHFE 2017 International Conference on Cross-Cultural Decision Making, July 17–21, The Westin Bonaventure Hotel, Los Angeles, California, USA*, edited by Mark, Hoffman, part of the series Advances in Intelligent Systems and Computing (AISC, vol. 610), 156–166. Cham: Springer, 2018.

Jeong, Hawoong, Bálint Tombor, Réka Albert, Zoltán Oltvai, and Albert Barabási. "The Large-Scale Organization of Metabolic Networks." *Nature* 407, no. 6804 (2000): 651–654. https://doi.org/10.1038/35036627

Karvalics, László. "Az információs társadalom gondolatának európai szálláscsinálója." *Információs Társadalom* VII, no. 1 (2007): 124–136. https://dx.doi.org/10.22503/inftars.VII.2007.1.13

Kaspersky Lab. 2024. "Website Defacement." Accessed March 10, 2024. https://encyclopedia.kaspersky.com/glossary/deface/

Kostyuk, Nadiya, and Aaron Brantly. "War in the Borderland through Cyberspace: Limits of Defending Ukraine through Interstate Cooperation." *Contemporary Security Policy* 43, no. 3 (2022): 498–515. https://doi.org/10.1080/13523260.2022.2093587

Krzanowski, Roman. "Information: Modern Concepts." *Információs Társadalom* XXIII, no. 4 (2023): 73–91. https://dx.doi.org/10.22503/inftars.XXIII.2023.4.5

Lonergan, Erica, Margaret Smith, and Grace Mueller. "Evaluating Assumptions about the Role of Cyberspace in Warfighting: Evidence from Ukraine." In *15th International Conference on Cyber Conflict: Meeting Reality (CyCon)*, 85–102. Tallinn, Estonia: NATO Cooperative Cyber Defence Centre of Excellence, 2023. https://doi.org/10.23919/CyCon58705.2023.10182101

Lunn, Stephen. "Human Security and the Digital Threat: Russia and Ukraine." In *Perspectives on Justice, Indigeneity, Gender, and Security in Human Rights Research*, edited by Laura E. Reimer and Katerina Standish, 263–283. Singapore: Springer Nature, 2023.

Majó, Zoltán. "Úton az információs társadalom felé: tudjuk-e, hová tartunk?" *Információs Társadalom* VI, no. 1 (2006): 30–39. https://dx.doi.org/10.22503/inftars.VI.2006.1.4

Miron, Marina, and Rod Thornton. "Russian Cyberspace Operations Against Ukraine in the 2022 War: How Effective Have They Been and What Lessons for NATO Can Be Drawn?" In

Great Power Cyber Competition: Competing and Winning in the Information Environment, edited by David V. Gioe and Margaret W. Smith, 57–72. Abingdon, UK: Routledge, 2024.

Traficom (Finnish Transport and Communications Agency National Cyber Security Centre). 2023. "Cyber and Information Influence Activities Come Together in the Hack and Leak Phenomenon." Accessed March 10, 2024. www.kyberturvallisuuskeskus.fi/en/ajankohtaista/hack-and-leak

Office of Public Affairs, United States Department of Justice. 2023. "Multiple Foreign Nationals Charged in Connection with Trickbot Malware and Conti Ransomware Conspiracies." Accessed March 10, 2024. www.justice.gov/opa/pr/multiple-foreign-nationals-charged-connection-trickbot-malware-and-conti-ransomware?fbclid=IwAR0hcMYb-92uYkmevnCXcScdeQXHO2iFTFJhVA6h-1U6YdipSms87J7I-uk

Security Alliance Limited. 2022. "The Changing Landscape of Hacktivism." Accessed March 10, 2024. www.secalliance.com/blog/the-changing-landscape-of-hacktivism?fbclid=IwAR2S5nByA9uV_rP5AuBwHR7m2lDyZcDEl7lCkGz3FzWunWSPrLcRQcuUtNA

Willett, Marcus. "The Cyber Dimension of the Russia–Ukraine War." *Survival* 64, no. 5 (2022): 7–26. https://doi.org/10.1080/00396338.2022.2126193

OLGA CHOLYSHKINA, ANDRII ONYSHCHENKO,
VOLODYMYR KUDIN, MYROSLAVA GLADKA,
SERHII OLEKSIIENKO

The use of artificial intelligence in optimising education management processes

Innovation and transformation of educational systems

As a strategic technology, artificial intelligence (AI) contributes to the transformation of the economy and symbolises a new stage not only in the history of digital technologies but also in the global development of modern civilisation. It also plays an important role in improving the quality and accessibility of education. The use of AI allows us to move from standard methods of teaching and education management to individual and effective strategies. This article analyses the use of AI in the field of education management and highlights innovative approaches introduced by AI. The potential disadvantages and ethical issues arising from the integration of these technologies into the field of education are considered. Prospects and directions of AI use in education are outlined. Conclusions are drawn about the importance of AI for current and future education.

Keywords: *information technology, artificial intelligence, educational process, learning technologies, education management*

Author Information

Olga Cholyshkina, Interregional Academy of Personnel Management
https://orcid.org/0000-0002-0681-0413

Andrii Onyshchenko, Taras Shevchenko National University of Kyiv
https://orcid.org/0000-0003-3975-8946

Volodymyr Kudin, Taras Shevchenko National University of Kyiv
https://orcid.org/0009-0005-9071-1213

Myroslava Gladka, Taras Shevchenko National University of Kyiv
https://orcid.org/0000-0001-5233-2021

Serhii Oleksiienko, Donetsk State University of Internal Affairs
https://orcid.org/0009-0006-9150-8329

How to cite this article:

Cholyshkina, Olga, Andrii Onyshchenko, Volodymyr Kudin, Myroslava Gladka, Serhii Oleksiienko. “The use of artificial intelligence in optimising education management processes”. *Információs Társadalom* XXIV, no. 2 (2024): 33–47.
https://dx.doi.org/10.22503/inftars.XXIV.2024.2.2

1. Introduction

In recent years, artificial intelligence (AI) technologies have rapidly entered the educational and scientific spaces, transforming the economy and society, and symbolising a new stage not only in the history of digital technologies but also in the global development of modern civilisation as a whole. AI has been defined as a strategic technology that offers many benefits to citizens and society as a whole, provided it is human-centric, ethical, sustainable and respects fundamental rights and values. At the end of 2022, OpenAI introduced the world to a number of its digital services that use the principles of human brain organisation and functioning, commonly referred to as AI. ChatGPT and DALL-e allowed the company to break all possible world records in terms of user growth. According to *The Guardian* (Milmo 2023), two months after the launch, the number of users exceeded 100 million, i.e. in January 2023, more than 590 million visits to the website from 100 million unique users were recorded. This has triggered a real AI revolution in the world. At the same time, the complexity of the technology itself, the ease of its use (it is accessible even to users who know only how to use messengers, as the ChatGPT interface is implemented similarly) and, most importantly, the quality of the text generation has sparked an active discussion in the global educational environment, including the university environment.

Undoubtedly, humanity will benefit from the new capabilities of AI, but this also entails solving many strategic tasks. First of all, the main one is the reliable use of AI systems' capabilities for the development of societies and human life, as the significant impact of these technologies is radically changing human existence. Second, being aware of the significant advantages and achievements of AI systems, it is necessary to predict unexpected challenges and risks that may arise as a result of such a continuous development of these digital systems, and accordingly to outline ways to prevent and eliminate them.

The use of AI in the educational process is not a fantastic idea for the future but a reality today. In January, a survey was conducted among Stanford University students. According to the results, almost 20% of students have already used ChatGPT to help with homework, exams and academic projects (Cu and Hochman 2023). A recent survey of teachers in the United States showed that two-thirds of educators have experienced a situation where students used ChatGPT to prepare homework without their permission. This figure is undoubtedly growing every month. It is obvious that in Ukraine, a significant number of pupils and students are already using ChatGPT, and this figure will soon become more and more significant (Intelligent 2023).

We must understand that in real life and professional activities, people will increasingly use such applications, and it is likely to become the norm in the coming years. Accordingly, the ability to use them effectively and ethically is an important skill, as is the ability to use Wikipedia responsibly. The relevance of this study lies in the fact that the use of AI and neural networks in student learning is one of the most promising areas of development in modern education (Semenets-Orlova, Rodchenko et al. 2022). AI technologies can have a positive impact on all aspects of the

educational process, help improve the quality of learning and ensure more-efficient knowledge transfer.

The purpose of the study is to analyse the use of AI in education management and its impact on process optimisation and transformation of educational programmes, highlighting various innovative approaches implemented with the help of AI. Exploring the role of AI in educational data analytics allows adapting educational processes to the individual needs of educational entities. Determining how AI technologies contribute to the effectiveness of management decisions in educational institutions allows improving planning, coordination and evaluation of educational processes and uncovering the potential challenges and ethical issues that arise in the process of integrating these technologies. The study will also outline prospects and directions for the development of AI in education and the importance of AI for current and future education.

2. Materials and Methods

The methodology for studying AI in the field of education management is based on the integrated application of various methods of analysis and synthesis, including analogy, abstraction, induction, deduction, dialectics and analytics. The analysis method is used to study the impact of AI on process optimisation and the transformation of educational programmes. This includes a thorough analysis of the links between AI and educational processes. The synthesis method explores innovative approaches introduced through AI, creating a holistic view of strategies for managing resources and improving education.

The study of the role of AI in education and training uses an inductive method to determine how technology can adapt educational processes to meet the individual needs of learners. The deductive method is used to formulate general conclusions and principles of the study. The dialectical method allows us to consider different perspectives and the interaction of technological innovations in the context of learning activities.

The method of analogy is used to identify potential challenges and problems that arise in the process of integrating information technology into education. The method of abstraction helps to discuss the prospects and directions of AI development in education. The generalisation method allows us to draw general conclusions about how AI contributes to the effective planning, coordination and functioning of educational processes. This methodology provides a systematic and in-depth study of the use of AI in education, taking into account various aspects and ensuring a comprehensive understanding of its impact and prospects.

3. Results

Current changes in the development of postmodern society are happening somewhat abruptly due to the creation and implementation of AI technologies. AI is a

product of human activity that, thanks to its ability to perform certain human functions, has significantly integrated into various industries and areas of social interaction, including the education sector, offering new solutions to improve learning and teaching and thus determining the future of society. Educational institutions should adapt their development strategy and focus on promoting responsible, high-quality, ethical and transparent use of AI tools by their employees and students in line with the current legislation, with the subsequent mitigation of the negative effects of its impact on society, culture and the economy. In general, the evolution of information technology-assisted learning dates back to the early 1980s and the publication of the first *International Journal of Artificial Intelligence in Education* in 1989, while the International Society for Artificial Intelligence in Education was founded only in 1993 (Tolochko and Godunova 2023).

There are great opportunities and prospects for using AI in education to transform the educational process into a more innovative, inclusive, efficient and effective one by introducing new high-quality learning methods that are fast, personalised and student-centred (Viznyuk et al. 2021). The real goal of AI in education systems should be to maximise the individualisation of education by offering students personalised learning paths according to their strengths and weaknesses and didactic material adapted to their characteristics while maintaining the quality of education and the integrative principle of education systems (Goksel and Bozkurt 2019; Mironova et al. 2022).

Significant changes in the use of AI in education, in particular in educational and research activities, are currently taking place in the following key areas:

1. assessment (in particular, automatic assessment and evaluation of students' learning progress and attitudes to learning, individual and group assessment, etc.);
2. prediction of learning status (prediction of student dropouts, risk groups, innovative abilities, career decisions), productivity or satisfaction, improvement of educational experience;
3. assisting (supporting students in their education, for example through anthropomorphic presence, which includes virtual agents and persuasive intervention through digital programs);
4. tutoring (tailoring individual strategies and approaches to the characteristics and needs of students);
5. learning management (learning analytics, sequencing of curricula and programmes, instructional design and student placement) (Peven, Khmil and Makogonchuk 2023).

One of the key arguments in favour of using neural networks is optimisation of the learning process. Delegating routine tasks to AI (creating drafts of lectures, assignments, assessment automation systems; expressing answers to frequently asked questions, or general answers to written questions, for example for a course project) can speed up and relieve teachers' work. AI technologies can support the digitisation of content, making it easier to find and use the right information. In addition, AI can provide learners with access to more diverse and relevant sources of information, enabling them to obtain comprehensive and useful information for their learning (Nagao 2019).

AI has the potential to stimulate students' motivation to learn. For example, the system is ready to offer gaming elements in the learning process to help students be more interested and engaged in their studies. Another advantage of using AI in the learning process is the automatic assessment of students' knowledge and skills. AI systems allow assessing not only closed-ended answers in a test format but also descriptive answers (Drach et al. 2023).

AI systems provide a more objective assessment of knowledge and skills as they use standardised algorithms and evaluation criteria. This helps to avoid subjectivity in student assessment, which can be present in manual assessment. To combat the problem of cheating in online exams, many institutions use automated security software. These tools monitor students through their webcams during tests. Using AI, the software can point out suspicious movements or actions, such as when a student frequently looks away from the screen or someone else enters the room. AI systems provide high reliability of assessment because they are not affected by emotions and fatigue. AI also identifies mistakes in students' work that may be missed during manual assessment.

AI helps to create curricula that meet the personal needs and interests of all students. Each student has individual needs and interests that can be met by AI systems. For example, if a student has certain interests, the system can recommend learning materials that relate to this particular information area (Terepyshchy 2023). Thus, individualised learning helps students who have significant differences in prior knowledge and skills.

Later, each student can have an assistant, such as Siri or Alexa. This system has enough data and experience to enable the neural network to learn itself. The main functions must be laid out in such a way that timely questioning is a Socratic principle. And then the dream of ancient philosophers may come true – that every citizen will be educated as a philosopher (Sharov 2023).

The use of AI and neural networks is significantly increasing the accessibility of education. The use of online platforms and applications with built-in AI will allow students to get an education anywhere and anytime. AI can help provide access to education for people who had to move due to the war and continue their education in other countries. For example, AI-powered online learning systems can help students with internet access to get an education from anywhere. AI systems can make a significant contribution to improving the efficiency of management decisions in educational institutions through a variety of technologies and tools. The use of AI technologies in education can greatly facilitate management processes, contributing to the quality of learning and the efficiency of educational institutions. Here are some ways in which this can happen (Krupenina, Fedorchuk and Sabadosh 2023).

To begin with, AI systems can integrate and synthesise data from different sources, allowing for a more complete and in-depth understanding of a situation or problem. This is especially important in education, where individual attention to each student is key to achieving optimal results. Automated data processing allows you to create personalised learning approaches based on the unique needs and characteristics of each student. Machine learning algorithms are powerful tools for

analysing and understanding learning processes. The ability of these algorithms to identify complex patterns and trends in massive data sets allows not only immediate response to events but also planning management strategies to improve the learning process.

AI can also be used to predict and manage changes in the educational environment. Algorithms can analyse educational trends, predict possible risks, and recommend strategies to avoid or reduce the impact. This allows for more flexible and adaptive educational programmes that can meet the needs of a rapidly changing society (Yatsenko 2023). Automation through the use of specialised systems can play an important role in optimising resource decomposition and management. Scheduling automation systems not only simplify the process of scheduling events and classes but also take into account various parameters such as seat availability, availability and qualifications of teachers and other factors that affect the efficiency of the educational process (Semenets-Orlova, Shevchuk et al. 2022).

However, along with many positive opportunities, the use of AI in school education can cause certain problems. AI is already actively used by some students writing essays and papers, making it difficult for a teacher or someone who wrote the paper to understand either the AI or the teacher. Only China has developed a system for checking essays and other works using AI, but this system has not yet been implemented (Chen, Chen and Lin 2020).

The use of AI could lead to the automation of some processes in school education, such as checking assignments and assigning grades to students. This could lead to job losses for teachers and other educational institutions from a global perspective. The other problem is the restriction of freedom of expression and creativity in the learning process, as the AI system offers only a limited set of tasks and solutions. AI can lead to changes in communication between teachers and students, as well as between students. This can affect social interaction and the development of communication skills. Interaction with AI teachers is less emotionally charged, which affects students' motivation to learn. One more threat posed by AI is the lack of group interaction. The use of AI systems reduces students' ability to cooperate with other group members and develop social skills. One of the negative consequences of using AI is that students feel isolated and distant from the learning process. They may feel confused, not knowing who to turn to for help or advice, especially if they need one-to-one instruction.

The use of AI can lead to the collection and use of students' data without their consent or protection. There also can be such problems as dependence on technology and the loss of skills that could be useful in real life. AI can be programmed to stereotype and prevent discrimination based on race, gender, nationality or other categories. If AI is used to evaluate students, it raises questions about the objectivity of such an assessment and its fairness. AI and neural networks require a lot of high-quality data to be successful. If the data contains errors or inaccuracies, it can lead to incorrect results (Chen, Chen and Lin 2020).

AI and neural network systems are less flexible than teachers, which limits the ability to adapt to the specific needs of students. In addition, AI has a limited ability to perceive the context and the individual characteristics of students.

The use of AI to manage cybersecurity risks in education is rapidly becoming an integral part of the modern educational environment. As the use of technology in education increases, so does the risk of cyberattacks. AI-based cybersecurity solutions can help protect educational institutions from intruders, but implementing these solutions can pose several challenges.

The use of AI to manage cybersecurity risks in education is rapidly becoming an integral part of the modern educational environment. However, implementing these solutions can pose several challenges. Educational institutions need to ensure that their staff is properly trained to effectively use and manage AI solutions with access to the necessary resources to ensure that their AI solutions are properly maintained and updated and that they are properly integrated into their existing security infrastructure (Luan et al. 2020). By addressing these challenges, educational institutions can ensure that their AI solutions are effective in protecting their institutions from attackers.

There is an undeniable need for proper training to fully utilise the potential of AI. Before introducing new tools such as AI-powered writing and research assistants, students should learn how to formulate queries, interpret results and use the tool to complete tasks during practical classes. When using AI for grammar checking, teachers should emphasise to students that while the tool can detect mechanical errors, it does not always pick up on nuances of tone or style. Thus, teachers should encourage students to use the tool as a preliminary check but rely on their judgement for the final touch. Setting the right expectations ensures that AI is seen as a tool for improvement, not a magic wand (Alam 2021).

It is also important to remember that the use of AI in education requires significant investments that are not available to all educational institutions. AI and neural networks require a lot of high-quality data to function and learn. If the data being trained contains errors or inaccuracies, it leads to incorrect results. Another example is the use of an automated assignment grading system. If the system has been trained on data that contains errors, it will give students incorrect grades. Thus, poor data quality can lead to inaccuracies and incorrect functioning of the AI system in education. This reduces the quality of teaching and the provision of correct information to students.

The above disadvantages of using AI and neural networks in education show that these technologies cannot completely replace the human factor in education. However, they can be a useful complement to traditional teaching methods (Ouyang and Jiao 2021). The use of AI raises ethical concerns because it is capable of collecting and analysing large amounts of data, including students' data, which may contain sensitive information. This violates the privacy and security of such students. For example, an AI-assisted learning system collects and stores data on students' activity, including their progress, answers to tasks and knowledge gained. This data can be used to evaluate and address ethical issues such as discrimination and privacy violations. There are also ethical issues of accountability for the results produced by an AI learning system. If the system produces incorrect results, this has negative consequences for students, especially if their grades are used to make decisions about graduate school or employment.

The use of ChatGPT-like applications can reduce the spread of cheating in education, which involves writing assignments for money. On the other hand, new ethical challenges arise that need to be addressed now. The results of the Stanford University survey show that students actively use ChatGPT to help generate ideas for future essays and receive instant automatic feedback on their work. In this context, the use of AI tools should not be condemned but rather encouraged (Hwang et al. 2020).

AI applications can perform typical essays or labs with 'perfect marks', so this type of task should be a thing of the past. Instead, tasks that require analysis, an individual approach and critical thinking should take up an increasing share of the learning process. It is important to teach students not just to repeat existing ideas but to use them as a basis for creative solutions and new research. In this sense, ChatGPT-like applications open up additional opportunities for students and educators.

4. Discussion

AI, as a key player in the development of education, will not only bring significant changes to current approaches to learning but also provide the education sector with new opportunities to create effective and adaptive learning environments. The introduction of AI into education is a promising area that can significantly transform the current education system and prepare students for the challenges of the future.

Although AI may still have its opponents and sceptics, this technology will not disappear. Those working to improve AI are confident that this innovative technology will only become more widespread in schools in the next five to 10 years. Researchers believe that AI-enabled educational programmes will be widely used in different countries, especially as more and more educational institutions expand their hybrid curricula. In addition, virtual reality AI will enable students to communicate with classmates at home and vice versa. This can create a more optimised and homogeneous remote classroom experience during distance learning. As AI research continues to advance and the field evolves, the implications of these new technologies for the educational landscape are enormous. It's almost impossible to imagine a future where AI isn't being used somewhere in the classroom, simply because the potential benefits are too compelling to ignore.

With AI-enabled teacher assistants, integrated virtual reality classrooms and gamified lessons, the ability to develop personalised instruction for individual students is becoming a major goal for educators. Teachers are well aware of how effective differentiated learning is for students with special needs. AI is designed to create unique individualised lesson plans for students without putting an additional burden on teachers' time and resources. This change promises to open up learning to a whole generation of students who otherwise find it difficult to learn.

AI has already been widely used to help catalogue and grade student work. Right now, AI programs can already determine whether an article is plagiarised or written with the help of an AI program such as ChatGPT. These programs can also help with bulk grading, significantly reducing the amount of time an instructor has to spend

checking assignments. As these learning models grow in scale and capability, they will continue to be applied to more and more specific and complex projects.

An AI tutor can complement a student's education in several pragmatic ways. Virtual assistants can help students develop flashcards, mock exams and other practice tasks. Speech synthesis speech-to-text learning models or speech-to-text generation can help with learning new languages or be used to provide examples and learn phrases. These technologies can even be used in conjunction with traditional tutoring to create a hybrid tutoring session, giving students access to tutor guidance no matter where they are (Knox 2020).

By combining new approaches and technologies, you can shape the learning environment. Let's look at a few examples of how AI is being used in online learning. AI is changing online learning. It can be effective due to the ability to analyse student data and change the learning process itself based on the results of the analysis. The simplest and most direct use of AI is to monitor knowledge, automate assignment checking, detect and correct errors and help instructor systems determine the autonomy of assignments during grading. AI also helps to eliminate cheating by analysing camera images and user browsing activity. AI-based security systems determine the autonomy of tasks.

The positive effect of using AI for knowledge and skills, the ability to recognise text and voice with subsequent natural language analysis, which has begun to be used in foreign language learning, is significant. The chatbot allows for multiple rereads when translating words, pronunciation, etc. – i.e. it acts as a tutor – which is important in online learning. Learning outcomes and assessment, quality of learning and speed of learning allow us to adapt learning to the characteristics of each person, their needs and mental abilities. AI systems allow educational institutions to develop individual learning profiles. Based on this, educators can assess students' capabilities to improve their knowledge and academic performance.

The use of voice assistants in online learning is positive. Students can use these tools to interact with course materials and meet their information needs. Instead of traditional handouts, teachers provide students with voice messages to advise them. Hence, this practice is one of the best uses of AI, as it reduces the human workload.

At the educational institution level, AI also helps to organise the learning process, managing the teaching team, students' performance, administrative tasks, etc. AI is greatly benefiting various foreign language translation services, such as Google, which has implemented Google Neural Machine Translation (GNMT) in its translators since 2016. GNMT significantly improves the quality of translation by training the system from information contained in millions of online sources. Thanks to its extensive end-to-end structure, the system learns to produce better translations over time. Let's take a look at some of the technologies used in the learning process and their role in it (Huang, Saleh and Liu 2021).

By combining these approaches and technologies, competent methodologists can shape the learning environment. Let us consider some approaches to the use of AI in online learning. Regardless of whether we are talking about academic learning or other life skills, each person is different. Everyone has their own needs and different levels of mental capacity. Teachers believe that it is necessary to cope with this

situation. Therefore, they believe that it is better to teach on an average. So, when AI comes into the picture, there is a whole new level of courtesy towards individual needs.

AI-powered systems now allow educational institutions to develop an individual learning profile according to each student's abilities. Based on this profile, educators around the world will be able to assess the abilities of their students better than ever before. Accordingly, they will also be able to help students significantly improve their academic performance.

Voice assistants can improve the perception of information. Amazon's Alexa and Microsoft's Cortana are used in this regard as indispensable assistants for teachers. Students can use these tools to interact with course materials and meet their information needs. Instead of traditional handouts on research materials, teachers provide students with voice assistants; for example, the University of Arizona has been practising this form of education for some time now to assist undergraduate students. This approach reduces the need for internal constant support from the teacher. In addition, students receive a higher level of education. This contributes to their academic experience. Thus, such a practice is one of the best examples of the use of AI, as it reduces the level of human workload.

AI can assist teachers in solving various tasks. In education, academic teaching and the presentation of educational material are not a single problem. At the level of the educational institution, it is necessary to manage the teaching staff and student performance, as well as to perform administrative tasks that ensure the continuous operation of the academic environment. For example, such tasks include human resources and learning environment management, core document management, group materials management, as well as non-teaching duties such as assessment and student collaboration.

In 2020–2021, Finland will provide EU residents with free access to the Elements of AI online course. The programme was created by the University of Helsinki and the Finnish technology company Reaktor. It is designed to encourage people to learn the basics of AI, regardless of their age and education. Everyone needs digital skills and awareness nowadays. The initiative costs the Finnish budget €1,679,000. Already, 700 Flanders schools in Belgium have started using AI in education this academic year. The local authorities signed an agreement with the British company Century Tech, which has developed a platform that allows for personalised learning. The goal is to enable each student to learn at their own pace and, at the same time, reduce the administrative burden on teachers. The idea is to move away from the traditional model of education, when teachers try to teach children at different levels, to one where the AI platform helps to adapt the types of tasks, time for their completion and assessment for each student. This is a kind of constant diagnostics and planning of a child's progress (Cope, Kalantzis and Searsmith 2021).

In practice, it looks like this: schools upload the curriculum to the Century system, and then the platform breaks down its content into micro-lessons. Children take a quick assessment to determine who is at what level. The system then develops learning for each student, based on what they need to work on more. Unlike other adaptive technologies, learning based on clear algorithms and rules, the Century

platform uses AI that learns and adapts its knowledge of the student. During the learning process, it is constantly updated. The basic principles of neuroscience allow us to individually assess the level of concentration and perseverance of students, the pace and best time to study, and how long it takes for information to move from short- to long-term memory. The company claims that its technology saves teachers an average of six hours a week. Personal information about the child is only required to log in to the platform. AI does not use it.

China has plans to become a leader in the use of AI. The government spares no expense on large-scale projects that bring together leading IT companies, startups and schools. The number of classrooms equipped with AI cameras and brainwave trackers is increasing in China. Journalists of the American edition of the *Wall Street Journal* visited one of the primary schools in Shanghai. The lesson begins with all children wearing a special tracker on their foreheads, which looks like a hoop. A special sensor that lights up in different colours is located above the eyebrows. There are special sensitive electrodes there. They are also on both sides where the tracker is attached near the child's ears. The lesson begins with meditation. The tracker can determine the child's level of concentration at a certain moment. The information is immediately sent to a computer, and teachers and parents can see it.

The use of AI in schools in different countries is still in its early stages. However, there are already concerns about the confidentiality of personal data and the undermining of the status of teachers. Some believe that modern technologies will facilitate the rapid acquisition of information, while others argue that children are not ready for such experiments and do not perceive such innovations well. But we definitely cannot ignore or avoid all the opportunities and risks, achievements and disruptions that AI brings with it (Nguyen et al. 2023).

The issue of AI use in Ukraine has been one of the most important in the development of educational and scientific activities since the first days of the young state's independence. The Institute of Artificial Intelligence Problems of the Ministry of Education and Science of Ukraine and the National Academy of Sciences of Ukraine was established in 1991, and its scientists developed a draft National Strategy for the Development of Artificial Intelligence in Ukraine 2021–2030. The Institute's scientists proposed two areas of research based on application of the principles and mechanisms of human brain functioning, in particular its consciousness. The first area includes research for science, education, medicine, defence industry, agriculture, etc., and the second is the creation of a competitive new generation computer with a basic smart unit – artificial consciousness and a high level of AI. About the use of AI in research and education, research areas in computing systems have been identified. The government approved the Concept for the Development of Artificial Intelligence in Ukraine and an action plan for its implementation for 2021–2024 (Panukhnyk 2023).

The key problems in this area in Ukraine are insufficient funding for AI research projects in educational institutions; insufficient quality of higher education and educational programmes for training relevant specialists; and uncertainty of ethical approaches to the use of AI technologies. The functioning of real mechanisms for the protection of human rights and universal values, and not just their declarative

nature, is particularly critical, particularly in the context of the Russian Federation's armed aggression against Ukraine. That is why it is necessary for the national, institutional and individual levels to:

- facilitate attraction of grant funding for research activities in the field of AI;
- partially compensate researchers' expenses for participation in international conferences;
- support international cooperation in sharing best practices;
- promote the use of AI technologies in the fields of science, as well as interdisciplinary research at the intersection of AI and other fields of science;
- develop clearly defined rules for the use of AI based on the value principles justified by the EU and the European Association of Universities. Regulation of the use of AI should be carried out at the institutional level in line with the mission, vision, goals and values of the educational institutions based on the key ethical principles;
- assist teachers at the institutional level in mastering modern developments in the use of AI, as well as rethinking the role of the teacher in teaching and learning through AI technologies;
- since the application of AI in higher education can be carried out at all stages of the educational process, adapt approaches to teaching, learning and assessment for the effective and efficient use of AI, taking into account the socio-psychological consequences;
- inform participants in the educational process about the use of AI in educational and scientific activities as a tool or research method based on responsibility, ethics, transparency and academic integrity;
- take into account industry specifics, in particular pedagogy, which involves orientating applicants to quality standards within the framework of interaction between the applicant and the created 'rules of the game', as well as awareness of the nature of interaction with the AI system.

About compliance with ethical standards, the Concept envisages, first of all, the need to regulate social relations in the field of AI development at the legislative level and the development of an Ethical Code of Artificial Intelligence. The development and implementation of AI technologies in Ukraine is defined as one of the tasks of other regulatory acts, in particular the National Informatisation Programme for 2022–2024.

4. Conclusions

AI is playing an important role in improving the quality and accessibility of education. The use of AI allows us to move from standardised teaching methods to individualised and effective strategies. AI also enhances learning through the use of virtual reality and other immersive technologies, creating learning environments that encourage active participation and facilitate understanding of complex concepts. AI-enabled online platforms provide global access to learning resources, reducing geographical limitations and making education more accessible to a wider audience, especially in regions where access to quality education may be limited.

AI also helps automate routine tasks, such as grading, allowing teachers to focus on personalised guidance and stimulating students' creative thinking. However, the successful implementation of AI in education requires careful consideration of moral and security aspects. Maintaining a balance between technological capabilities and the role of the teacher is essential to ensure a warm and supportive educational environment where not only the intellectual but also the social and emotional aspects of students' personalities are developed. This synergistic approach maximises the potential of AI to create high-quality and accessible education.

The introduction of AI in education is a transition from reactive to strategic management, where the ability to predict trends plays an important role in the development of effective management strategies. The use of AI analytics allows us to gain a deep understanding of educational processes and respond to changes promptly, ensuring a higher level of adaptability and innovation.

AI can adapt to the different learning speeds and perception styles of students, providing a personalised approach. However, it is important to remember that educational success is not limited to advanced technologies. Teachers play an important role in nurturing, mentoring and creating an environment of emotional support for students. By striking a balance between technological capabilities and a humane approach, the educational process can become a holistic system that develops both the intellectual and the social aspects of the individual. Such a harmonious approach can prepare students for the challenges of the modern world by developing their technical and interpersonal skills for the successful functioning of society.

References

Alam, Ashraf. 'Possibilities and Apprehensions in the Landscape of Artificial Intelligence in Education'. In *2021 International Conference on Computational Intelligence and Computing Applications*, 1–8. Nagpur, India: IEEE, 2021.

Chen, Lijia, Pingping Chen and Zhijian Lin. 'Artificial Intelligence in Education: A Review'. *IEEE Access* 8 (2020): 75264–78.

Cope, Bill, Mary Kalantzis and Duane Searsmith. 'Artificial Intelligence for Education: Knowledge and Its Assessment in AI-Enabled Learning Ecologies'. *Educational Philosophy and Theory* 53, no. 12 (2021): 1229–45.

Cu, Mark Allen, and Sebastian Hochman. 'Scores of Stanford Students Used ChatGPT on Final Exams, Survey Suggests'. *Stanford Daily*, 22 January 2023. Accessed 11 November 2023. https://stanforddaily.com/2023/01/22/scores-of-stanford-students-used-chatgpt-on-final-exams-survey-suggests/

Drach, Iryna, Olha Petroye, Oleksandra Borodienko, Iryna Regheilo, Oleksandr Bazelyuk, Nataliia Bazelyuk and Olena Slobodianyuk. 'Use of Artificial Intelligence in Higher Education'. *International Scientific Journal 'Universities and Leadership'* 15 (2023): 66–82.

Goksel, Nil, and Aras Bozkurt. 'Artificial Intelligence in Education: Current Insights and Future Perspectives'. In *Handbook of Research on Learning in the Age of Transhumanism*, edited by Serap Sisman-Ugur and Gulsun Kurubacak, 224–36. New York: IGI Global, 2019.

Huang, Jiahui, Salmiza Saleh and Yufei Liu. 'A Review on Artificial Intelligence in Education'. *Academic Journal of Interdisciplinary Studies* 10, no. 3 (2021): 206–17.

Hwang, Gwo-Jen, Haoran Xie, Benjamin W. Wah and Dragan Gasevic. 'Vision, Challenges, Roles and Research Issues of Artificial Intelligence in Education'. *Computers and Education: Artificial Intelligence* 1 (2020): 100001.

Intelligent. 2023. '8 in 10 Teachers Approve of Student Use of ChatGPT, Nearly All Use It Themselves'. Accessed 11 November 2023. www.intelligent.com/8-in-10-teachers-approve-of-student-use-of-chatgpt-nearly-all-use-it-themselves/

Knox, Jeremy. 'Artificial Intelligence and Education in China'. *Learning, Media and Technology* 45, no. 3 (2020): 298–311.

Krupenina, Nataliia, Anna Fedorchuk and Vasyl Sabadosh. 'The Role of Teachers in the New Era of Distance Learning and Artificial Intelligence'. *Current Issues in Modern Science* 10, no. 16 (2023): 565–78.

Luan, Hui, Peter Geczy, Hollis Lai, Janice Gobert, Stephen J. H. Yang, Hiroaki Ogata and Chin-Chung Tsai. 'Challenges and Future Directions of Big Data and Artificial Intelligence in Education'. *Frontiers in Psychology* 11 (2020): 580820.

Milmo, Dan, and agency. 'ChatGPT Reaches 100 Million Users Two Months After Launch'. *Guardian*, 2 February 2023. Accessed 23 October 2023, www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app

Mironova, Nadya, Hanna Koptieva, Iryna Liganenko, Ayta Sakun and Daryna Chernyak. 'Modeling the Selection of Innovative Strategy for Development of Industrial Enterprises'. *WSEAS Transactions on Business and Economics* 19 (2022): 278–91.

Nagao, Katashi. 'Artificial Intelligence in Education'. *Artificial Intelligence Accelerates Human Learning: Discussion Data Analytics* (2019): 1–17.

Nguyen, Andy, Ha Ngo, Yvonne Hong, Belle Dang and Bich-Phuong Thi Nguyen. 'Ethical Principles for Artificial Intelligence in Education'. *Education and Information Technologies* 28, no. 4 (2023): 4221–41.

Ouyang, Fan, and Pengcheng Jiao. 'Artificial Intelligence in Education: The Three Paradigms'. *Computers and Education: Artificial Intelligence* 2 (2021): 100020.

Panukhnyk, Olena. 'Artificial Intelligence in the Educational Process and Scientific Research of Students of Higher Education: Responsible Limits of AI Content'. *Galician Economic Bulletin* 83, no. 4 (2023): 202–11.

Peven, Kateryna, Nataliia Khmil and Nataliia Makogonchuk. 'The Impact of Artificial Intelligence on Changing Traditional Models of Learning and Teaching: An Analysis of Technologies to Ensure the Effectiveness of Individualized Education'. *Perspectives and Innovations of Science* 11, no. 29 (2023): 306–16.

Semenets-Orlova, Inna, Larisa Rodchenko, Inna Chernenko, Oleh Druz, Mykola Rudenko and Ruslan Poliuliakh. 'Requests for Public Information in the State Administration in Situations of Military Operations'. *Anuario De La Facultad De Derecho. Universidad De Extremadura* 38 (2022): 249–70.

Semenets-Orlova, Inna, Roman Shevchuk, Bohdan Plish, Andrii Moshnin, Yaroslav Chmyr and Ruslan Poliuliakh. 'Human-Centered Approach in New Development Tendencies of Value-

Oriented Public Administration: Potential of Education'. *Economic Affairs (New Delhi)* 67, no. 5 (2022): 899–906.

Sharov, Serhii V. 'The Current State of Artificial Intelligence Development and Directions of Its Use'. *Ukrainian Studies in the European Context* 6 (2023): 136–44.

Terepyshchy, Serhii. 'Media Literacy in the Age of Artificial Intelligence: Integration of Artificial Intelligence Tools and Methods into Modern Pedagogical Approaches'. *Pedagogy: Actual Issues of Humanitarian Sciences* 60 (2023): 195–202.

Tolochko, Svitlana, and Anastasiia Godunova. 'Theoretical and Methodological Analysis of Foreign Practices of Using Artificial Intelligence in Education and Science'. *Bulletin of Science and Education* 7, no. 13 (2023): 832–48.

Viznyuk, Inessa, Natalia Buglai, Larysa Kutsak, Anna Polishchuk and Viktoriia Kylyvnyk. 'Use of Artificial Intelligence in Education'. *Modern Information Technologies and Innovation Methodologies of Education in Professional Training Methodology Theory Experience Problems* (2021): 14–22.

Yatsenko, Oleksandr. 'Artificial Intelligence Technologies: The Main Directions of Implementation in the Educational Process of a Higher Education Institution'. In *The 11th International Scientific and Practical Conference 'Scientific Research in the Modern World'*, 252–7. Toronto: Perfect Publishing, 2023.

The language and motivations of expertise in political discourse

Characteristics of specialised communication in the Hungarian ParlaMint corpus

The study aims to explore how language elements of specialised and political communication are represented in parliamentary discourse. Within the ParlaMint corpus, we examine Hungarian parliamentary speeches between 2020 and 2022 on the 'KATA' (specific tax of small taxpayer businesses). The specialised terms, terms used in standard language, sentiment and attitude values are analysed. Results show that there is a significant difference in term use and the sentence sentiments between the different discourse participants (policy actors, opposition, governing party). There are only a few emotionless speeches, and no large differences in the proportion of terms and emotions are observed. The results can be relevant for studies on populist communication, the relationship between technocracy and democracy, and the methodological aspects of computer-assisted discourse analysis. The paper also outlines future orientations for research, including extended thematic analysis and investigation of contextual shifts in the actual meaning of terms within political discourse.

Keywords: *terminology usage, sentiment analysis, parliamentary discourse, ParlaMint corpus, specialised communication*

Acknowledgement

For Nóra Falyuna the János Bolyai Research Scholarship of the Hungarian Academy of Sciences supported the paper (grant number: BO/00729/23/9). We thank Kinga Mátyus for her contribution to the research.

Author Information

Réka Dodé, Hungarian Research Centre For Linguistics
https://orcid.org/0009-0008-3880-2750, Email: dode.reka@nytud.hun-ren.hu

Nóra Falyuna, Ludovika University of Public Service
https://orcid.org/0000-0002-5636-903X, Email: falyuna.nora@uni-nke.hu

How to cite this article:

Dodé, Réka, Nóra Falyuna. "The language and motivations of expertise in political discourse". *Információs Társadalom* XXIV, no. 2 (2024): 48–67.
https://dx.doi.org/10.22503/inftars.XXIV.2024.2.3

1. Introduction

The study presents an analysis of the discourse of Hungarian parliamentary sessions. Parliamentary sessions exhibit distinctive discourse characteristics: they address policy issues, thereby incorporating elements of both specialised discourse and political discourse. Additionally, they target a diverse audience, reflecting multifaceted goals. The study corpus consisted of speeches delivered from 2020 to 2022 within the Hungarian ParlaMint subcorpus (Ligeti-Nagy et al. 2023), focusing on the topic of the KATA (acronym of the 'kisadózó vállalkozások tételes adója', the specific tax of small taxpayer businesses in Hungary). The research aimed to explore how the linguistic and language usage elements of specialised and political communication manifest within parliamentary discourse. It sought to determine to what extent these elements could be considered specialised or political communication based on linguistic, language usage and contextual features. To achieve this, we examined terminologies, sentiment and attitude values commonly used in both specialised and political discourse, which are most readily identifiable using computational tools. Beyond contributing to the discussion on the use of specialised language in political discourse, the study also contributes to the methodological aspects of computer-assisted analysis of specialised and political discourses. Importantly, the research highlights the existence and research potential of the ParlaMint corpus (Erjavec et al. 2023) and emphasises the societal importance of responsibly and critically interpreting statements made within the parliamentary environment.

2. Political discourse and specialised discourse

The utterances made during parliamentary sessions belong to political discourse, but they also occupy a specific place in it:

1. Policy issues are also discussed in parliamentary sessions, so there are also elements of specialised and political discourse.
2. Parliamentary sessions are highly regulated, institutionalised, with a set agenda (the language, types of speeches, duration, order of speakers and other rules of the session are laid down in the Rules of the Hungarian National Assembly n.d.), but there are also somewhat spontaneous speeches alongside speeches written in advance (van Dijk 1997).
3. Public accessibility is regulated by law (speeches given in the Hungarian Parliament are available in the form of transcripts and video recordings), so they necessarily become part of the public discourse.

The linguistic and language-using features of parliamentary utterances are thus influenced by multiple purposes, audiences and discourse features. Due to the policy themes, elements of specialised discourse appear alongside political and social issues (van Dijk 1997). This is also reflected in the fact that the political terminology is characterised by the mixing of specifically political terms, units of general language, terms from different fields, diplomacy, economy, culture (Kotenko et al. 2023) and the multiple target audiences, which also imply the intention to persuade and

inform. The main characteristics of specialised communication and political communication can be described on the basis of several analytical frameworks.

2.1. Specialised discourse

Specialised communication is carried out by a group of specialists embedded in their specialised cultures, based on a specialised language, which they use according to the goals and expressive needs of the community (Kurtán 2010). The specificity of terminological units (the units of specialised language), as opposed to other language units at the same structural level and with the same meaning, lies in the fact that all three of their components (cognitive, grammatical and pragmatic) satisfy restrictive conditions. Among other things, they depend on the context of the subject (cf. also Faber 2009); they have a well-defined place in the conceptual structure, which determines its explicitly recorded meaning recognised and used by the specialised community. As a lexical unit, it may coincide formally with units of standard language use (Cabré 2003).

Specialised speech communities represent specialised cultures in which common knowledge and specialised background knowledge play a significant role. With regard to the community of speakers, it is crucial whether communication takes place between specialists in the same field, between a specialist in one field and a specialist in another field, or between a specialist and a non-specialist or the public (Kurtán 2010). However, a text is specific only if it is written by a specialist in the field; for example, a text about a disease is specialised if it is written by a doctor (Cabré et al. 2014). Nevertheless, many communicative scenarios fit into this communicative framework, with one important condition: they convey specialised knowledge. For example, it covers communication between specialists, between specialists and technicians, and the sharing of scientific or technical information widely with non-specialists (Cabré 2003).

Participants in communication communicate in different ways regarding the level of abstraction, the level of specialised knowledge and preparedness, and the intentions. In terms of context, language use is manifested in a variety of situations, in typical activities and in typical texts. What distinguishes special discourse (specialised language use) from others, apart from its preference for certain text types and strictly controlled knowledge structures, is that it presents information in a systematic way; it uses linguistic units that either are exclusively used in the subject or are more widespread but have limited meaning in this context; it works with texts with a specific content and a more concise, systematic, explicit and less inferential way of expression.

2.2. Political discourse

Based on these findings, parliamentary discourse may be one of the scenarios of specialised communication, but the contextual features of the parliamentary discourse

influence the features of specialised communication in this case. The presence of different audiences at the same time, the institutionalised framework in which spontaneous linguistic utterances can occur, the different goals and intentions, the fact that the participants in the parliamentary session refer to their own politics, ideologies, evaluations, attitudes towards political issues and their political actions in the decision-making discourse – all this means that the discourse is at least partly about politics itself (van Dijk 1997). Speeches in parliamentary sessions can be of various kinds (motions for a resolution, interpellations, reports, questions, requests, draft policy statements, policy debates, bills), and are mostly characterised by formal style and an argumentative, debating character, but they can also be characterised as political discourse by insulting the interlocutor and by influencing the public (Schirm 2009; Zimányi 2008). The role of the public in political communication has been particularly relevant since the emergence of new media (see, e.g., Merkovity 2012).

The persuasion strategy (Wilson 2015) can be based on ethos (the authority/credibility of the speaker); pathos (emotional appeal to the audience); and logos (proof of real or apparent truth, supporting reasons and information, argumentation). Political discourse analysis looks more broadly at the use of the linguistic choices of a political utterance for political purposes and their function in the construction of political reality (van Dijk 1997; Wilson 2015). Influence is also a prominent element here: the relationship between language and the image of the world is one in which politics is able to manipulate language for its own ends, to create worldviews that suit its goals and to negate others (Wilson 2015). Critical political discourse analysis has already argued that political utterances are tools of persuasion and power struggles, and this character determines the linguistic and stylistic features of texts: 'language is politics, politics assigns power, and power governs how people talk and how they are understood' (Lakoff 1990, 7). The main analytical focus is on the emergence and interaction of ideologies, power techniques and manipulation in texts, for example through linguistic constructions of polarisation, speech acts, implicatures, topical focus (see, e.g., Wodak and Chilton 2005; Vadai 2016).

2.3. Using specialised language in the political discourse

The appearance, function and even the meaning of the units of specialised language must therefore be investigated in this political context: 'whereas metaphors in classroom discourse may have an educational function, metaphors in politics will function in a political context, for instance in the attack on political opponents, the presentation of policies or the legitimation of political power' (van Dijk 1997, 24).

Expertise and its linguistic representation are fundamentally related to legitimation in political discourse (Reyes 2011). Studies on public policy emphasise the role of knowledge in the political process, with different definitions of the concept of knowledge utilisation pointing, among other things, to its symbolic use as an expression of the perceived rational basis for decisions. Justifying a decision with information is one way of symbolising the decision-making process as legitimate (Radaelli 1995). Politicians use legitimisation (a purposeful process to justify, which aims to

gain the support or approval of the interlocutor and thus persuades them) to justify their political agenda, in which different linguistic choices can be made, with strategies including appeals to emotion and authority. Politicians are inherently authoritative, as is (the appearance of?) expertise. The elements that represent expertise (such as reliable sources, figures) are thus legitimisation tools (Reyes 2011). Specialised language use can be such a tool too, as 'terms carry the specific knowledge that experts have but lay people do not (necessarily), and which knowledge can put experts in a dominant position, asymmetric relations with lay people, and make it easier to influence lay people (Falyuna 2017, 96); if only because, for example, readers make judgements about the communicator's reliability based on the use of specialised language (Hendriks and Kienhues 2019 based on Thon and Jucks 2017). The use of specialised language is … about the demarcation of boundaries and, in a sense, about control and power (over knowledge)'(Falyuna 2022, 56–7). So credibility, (the appearance of) expertise and authority are the basis of persuasion, but these can also be tools of manipulation, especially when the speech is addressed to non-specialist people (cf. the publicity of parliamentary sessions):

> for meaningful political discussion to take place a basic prerequisite is some degree of comprehension of the subject matter being discussed … In the case of expert knowledge, Turner maintains 'there is very often no such comprehension [by the masses] and no corresponding ability to judge what is being said and who is saying it' … both expert knowledge and the ability to identify who is an expert and can speak on a given matter being potentially inaccessible or incomprehensible to the public. (Thomas and Buckmaster 2013 based on Turner 2003, 46)

In addition to manipulation, they are also tools of power. Márton Szabó contrasts the 'boredom' of public policy, sectoral policies with 'the colourful world of party battles', and argues that the approach to the success of public policy proposals is often to remove politics from them, 'i.e., to reduce them to a simple technical issue in a narrow field' (Szabó 2003, 51). This relates to his statement that 'public policy programmes are implemented by experts and advisors, and decision-makers and executors usually defend their positions with scientific arguments' (Szabó 2012, 1). He further explains that

> state and party bureaucracies have monopolised the management of public policy, and through the positivist-objectivist language of thematization, have removed the presuppositions and political constraints of interpretations and proposals … Even those interested in public policy were confronted with the fact that the real problem in formulating and answering questions was the uncharacteristic jargon, or language itself, which assumed the appearance of neutrality and objectivity. (Szabó 2003, 52)

On this basis, he argues that two approaches have emerged in the development and implementation of public policy programmes and recommendations: 'One is

that the implementation of public affairs is a special professional task involving trained specialists in law, economics, sociology, management and leadership' (Szabó 2012, 2). Relating to the other approach,

> [a] public policy programme is only effective if it is not set and managed solely by the specialised apparatus, but also involves the citizens concerned in some way. … This enhances the role of values in decision-making and makes it clear that specialised knowledge does not take precedence over lay knowledge in a self-evident way in public and policy issues to be implemented collectively. (Szabó 2012, 3)

Compare also in this respect the relationship between technocracy and democracy, where democracy is based on legitimate consensus, free elections and participation, and technocracy sees expertise as the sole basis of power and authority (see, e.g., Radaelli 1995; Stehr 2007). See also the problem of expertise, where knowledge becomes part of politics but enters the political process along with interests (Collins and Evans 2007; summarising this specifically in public policy, Thomas and Buckmaster 2013). Compare also with topics of knowledge policy (see, e.g., Stehr 2010). Thus, the publicity of parliamentary sessions without their being publicly understandable does not in fact give the public insight into the political process, and the lack of clarity and transparency results in it being only the discourse of politicians and experts.

2.4. Emotions and attitudes in political discourse

In the context of the study, the means of public comprehensibility and public involvement may mean reducing the number of terms and using terms that exist in the standard language and are publicly understandable instead. But comprehensibility can also refer to the importance of interpretative frameworks in the political discourse (Szabó 2012), in the construction of which emotions, ideologies and beliefs play a prominent role. Although the percentage of emotions in political discourse depends on the choice of genre, discursive events and topics, for example in the case of routine legislative activity, informational content is stronger than emotional content, while the reverse is true for 'hot' political issues (Kenzhekanova 2015). Gennaro and Ash (2022) developed a measure based on computational linguistics tools to scale the emotionality of political language. In their study, they looked at how using emotion and reasoning in the US Congress has changed over the past 150 years, by topic and congressional speakers. They mention that emotional displays spike in wartime, but since the late 1970s there has been a significant increase in emotionality, coinciding with the introduction of television coverage of congressional debates. Further, patriotism, foreign policy and social issues are the most emotionally debated, while within economic policy, issues of taxation and redistribution have seen the greatest increase in emotion in recent years (particularly among Republicans) (Gennaro and Ash 2022). They also find that emotions are prominent among disadvantaged, minority groups (emotions can help politicians cope with loss of control or frustration of expectations, Gennaro and Ash

2022 based on MacLeod 1996; Lin et al. 2006, or serve to enforce policy positions, Gennaro and Ash 2022 based on Jerit, Kuklinski and Quirk 2009).

3. Empirical study

In our research, utilising the Hungarian subcorpus of the ParlaMint corpus, containing parliamentary speeches and their metadata from 6 May 2014 to 31 July 2023, we examined speeches on an economic issue, the KATA. The study corpus encompasses speeches delivered from 2020 to 2022, considering this period as changes to the law on KATA began during this time. It was a popular form of taxation in Hungary for self-employed individuals, but was significantly restructured in 2022. It was a hot and complex topic of social, political and economic discourse in Hungary. This is mainly because the government's proposals to amend the KATA came on the agenda after social discontent during the pandemic and inflation following the Russia–Ukraine war. The ParlaMint corpus (2023) is freely accessible and provides a rich research material for researchers in humanities and social sciences.

3.1. Hypotheses

Taking into account that discourse is specialised, scientific, political and general at the same time – targeting multiple audiences – we examine terms used in scientific context, terms used in standard language, sentiment values and attitude values (emotions and ideologies), which are among the most graspable elements of specialised and political discourses using computer tools.

Sentiment analysis examines the polarity of texts and evaluates/emotional content (positivity, negativity, neutrality) expressed in texts. Sentiment analysis can be carried out in various ways within a given text. We can examine the sentiment value of a given text or text segment (e.g., sentence) (Mifrah and Benlahmar 2022), as well as evaluative expressions related to a particular entity (whether it be a proper noun or a common noun) (Laki and Yang 2023).

We formulated the following hypotheses:

H1: There is a lower proportion of terms in the speeches of opposition and governing party actors, particularly a lower proportion of specialised terminologies, than in the speeches of policy actors.

H2: The opposition will make negative statements while the governing party will express positivity.

H3: Policy actors will express themselves neutrally.

3.2. Method

The participants of the parliamentary discourse were divided into groups based on the metadata of the ParlaMint corpus:

1. Members of Parliament (LMP [‘Politics Can Be Different’], MSZP [‘Hungarian Socialist Party’], Jobbik [‘Jobbik-Conservatives’; the name comes from ‘Movement for a Better Hungary’, and Jobbik means ‘better’ in English], DK [‘Democratic Coalition’], Mi Hazánk [‘Our Homeland Movement’]) – opposition parties
2. Members of Parliament (Fidesz, KDNP [‘Christian Democratic People’s Party’]) – governing party
3. non-members of Parliament, policy actors (state secretaries, experts, guests such as the state secretary in charge of tax issues).

We compared the speeches of Members of Parliament and non-members of Parliament because political affiliation can influence the political stance and communication of a person (van Dijk 1997). The next preparatory phase involved extracting texts related to KATA using the NoSketch Engine corpus querying tool (Sketch Engine 2024a). We examined the context of the extracted sentences and gathered those text excerpts for further research that touched upon the specific tax of small taxpayers, not just instances where the term was explicitly mentioned. The search was conducted for the terms *kisadózó* (‘small taxpayer’) and *KATA*, resulting in a total of 185 occurrences. Table 1 shows the total number of text words in the text excerpts. The number of tokens in the text excerpts was important information for calculating the proportion of term candidates.

	Max.	Min.	Average	Median
No. of tokens in an excerpt	2576	72	495	287

Table 1. Total number of text words in the text excerpts (compiled by the authors)

3.2.1. Extraction and analysis of term candidates

We manually extracted term candidates from the text excerpts containing the terms *KATA* and *kisadózó* as described above. This step will also be performed automatically in the future.

In our research we refer to the extracted expressions as term candidates, since we are not professionals in the field. Also, we have not verified the expressions in any terminology databases; thus, they can only be considered as term candidates.

Examples of selected term candidates: *NAV-törvény* (law related to NAV [Nemzeti Adó- és Vámhivatal ‘National Tax and Customs Administration’]), *adóhatóság* [‘tax administration’], *munkabér* [‘wage’], *feketefoglalkoztatás* [‘illegal employment’], *jogviszony* [‘legal relationship’], *kamara* [‘chamber’], *büntetőadó* [‘penalty tax’], *bevétel* [‘income’], *juttatás* [‘allowance’], *tevékenység(i kör)* [‘scope of activities’], *adótanácsadó* [‘tax consultant’], *adó* [‘tax’], *kisvállalkozás* [‘small businesses’].

We then decided whether the expression is specialised (clear in meaning only for the specialised context) or commonly used in standard language, understandable

for non-specialised (comprehensibility). To determine this, we employed the following method.

Based on the initial idea by Fóris (2005), if a term appears in a general dictionary, it can be considered commonly used in standard language beyond the specialised community. Nowadays, the role of general dictionaries has been taken over by general corpora. One such publicly available Hungarian corpus is the Hungarian Web 2020 (huTenTen) corpus (Nemeskey 2020), which is accessible in NoSketch Engine (Sketch Engine 2024a).

We used the frequency of occurrence of term candidates to determine whether such a sequence is an expression of the standard language and understandable for the general public. We set the threshold at 10,000 occurrences. Therefore, if a term candidate appeared between 0 and 10,000 times in the corpus, we marked it as a specialised term candidate, and if it appeared more than 10,000 times, it was considered an expression of plain language.

The advantage of corpora and corpus query tools is that they allow searching for multi-word expressions in various ways.

However, a disadvantage of using a web corpus as reference is that the principles of representativeness, balance and sampling (Sinclair 1991) are not upheld because it contains texts from various sources unchecked. The threshold of 10,000 occurrences was our own decision based on observation, but there are some cases where it did not prove to be adequate. For instance, *adónem* ('types of tax') occurred 11.022 times; however, *kisadózó* with 2.413 occurrences is a specialised term, although, due to the law, it is known by quite a lot of people (but this does not necessarily mean that they also know the meaning of it).

Table 2 shows the term candidates and their proportion compared to the words of the text, as well as the special term candidates and their proportion compared to the term candidates. It shows the highest, lowest proportion, and the average and median. Median is more tolerant of outliers.

	No. of term candidates	Proportion of term candidates (%)	No. of specialised term candidates	Proportion of specialised term candidates (%)
Max.	75.0	16.8	48.0	90.0
Min.	4.0	2.5	2.0	25.0
Average	24.4	6.4	14.2	57.1
Median	17.0	5.4	9.0	56.7

Table 2. Numbers and proportion of term candidates (compiled by the authors)

3.2.2. Measurement of sentiment values and attitude values

The importance of sentiment analysis lies in its ability to capture, through automated means, when text aims to evoke emotions or convey ideology, and when it seeks to persuade. As Laki and Yang (2023) express, there are various approaches under development for sentiment analysis; initial efforts have focused on categorising documents and texts according to their overall polarity (negative, positive or neutral) (Pang, Lee and Vaithyanathan 2002). Another approach is the aspect-based method, which offers a more detailed analysis by identifying specific aspects of an object or entity that contribute to the overall sentiment (Pontiki et al. 2014). An alternative strategy is referred to as sentence-level sentiment analysis, where the focus is on examining individual sentences within a document to ascertain their level of opinionisation (Feldman 2013).

The provision of sentiment and attitude measurements for statements containing references to *kisadózó* and/or *KATA* was done automatically using tools developed at the Hungarian Research Centre for Linguistics (Laki and Yang 2023). The values can be: 1 (positive), 0 (neutral), −1 (negative). These values are compared in the following manner:

- How are they distributed relative to each other in the speeches of politicians and other participants (experts)?
- What are the proportions concerning political affiliation?

Table 3 shows the sentence-label and aspect-label values of all speeches.

	negative	neutral	positive
sentence-label	88	57	40
aspect-label	125	33	27

Table 3. Sentence-label and aspect-label values of all speeches (compiled by the authors)

4. Results

4.1. Results of usage of term candidates

The hypothesis related to usage of terms was as follows:

H1: There is a lower proportion of terms in the speeches of actors of the opposition and governing parties, particularly a lower proportion of specialised terms, than in the speeches of policy actors.

Opposition parties		Governing party		Policy actors	
Proportion of term candidates (%)	Proportion of specialised term candidates (%)	Proportion of term candidates (%)	Proportion of specialised term candidates (%)	Proportion of term candidates (%)	Proportion of specialised term candidates (%)
4.9	**51.5**	**6.8**	**60.0**	**7.9**	**61.4**

Table 4. Proportion of terms in the speeches of opposition parties, governing party and policy actors (compiled by the authors)

The results show (Table 4) that speeches of opposition members have the lowest proportion of term candidates, especially specialised term candidates, followed by speeches of government party members. The highest proportion of terms and specialised terms use is observed among policy actors. In Figure 1, it is also apparent that the opposition not only uses fewer terms but also uses them with less diversity.

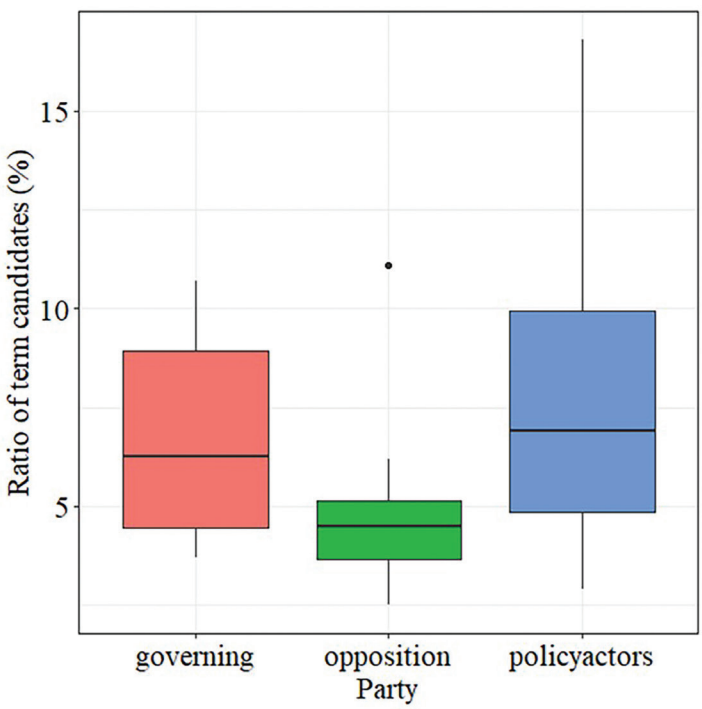

Figure 1. Ratio of term candidates among governing party, opposition parties and policy actors (compiled by the authors)

We conducted a Kruskal–Wallis test in R (R Core Team 2022) to compare the proportions of terms in text among the three groups. Based on the results, there is a significant difference among the groups ($\chi^2(2, 46) = 8.542$, $p = 0.013$). For pairwise comparisons between groups, we used Dunn's (1964) Kruskal–Wallis multiple comparison p-values adjusted with the Benjamini–Hochberg method (FSA [Ogle et al. 2023]). According to this, significant differences are observed between policy actors and the opposition parties, as well as between the opposition parties and the governing party. We conducted a Kruskal–Wallis test to compare the proportions of specialised term candidates among the three groups. Based on the results, there is no significant difference among the groups. Based on Spearman's correlation, it is observed that the longer the text, the fewer terms it contains ($rho = -0.465$, $p = 0.001$). However, it is noteworthy that there is high variability in shorter texts and lower variability in longer texts. Based on these results H3 has been confirmed.

4.2. Results of sentiment analysis

The hypotheses related to sentiment were as follows:

H2: The opposition will make negative statements while the governing party will express positivity.

H3: Policy actors will express themselves neutrally.

Table 5 presents the distribution of sentiment values for aspect sentiment and sentence sentiment separately for each group.

	Sentence sentiment			**Aspect sentiment**		
	negative	**neutral**	**positive**	**negative**	**neutral**	**positive**
Opposition parties	54,29%	31,43%	14,29%	75,71%	17,14%	7,14%
Governing party	14,81%	40,74%	44,44%	55,56%	22,22%	22,22%
Policy actors	52,27%	27,27%	20,45%	64,77%	18,18%	17%

Table 5. Result of the sentiment values governing party, opposition parties and policy actors (compiled by the authors)

Figure 2 illustrates the proportion of sentence sentiment broken down by text for all three groups. It is shown that negative sentiment is very rare in the texts of government members, while in texts of the opposition, positive sentiment is below 25% in most cases. In texts of policy actors, multiple sentiments are present, but positive sentiment appears in the most diverse range (regardless of the number of occurrences in the text), while neutral sentiment is the least diverse among policy actors. The distribution of different sentence sentiments across groups is significantly different according to the chi-square test.

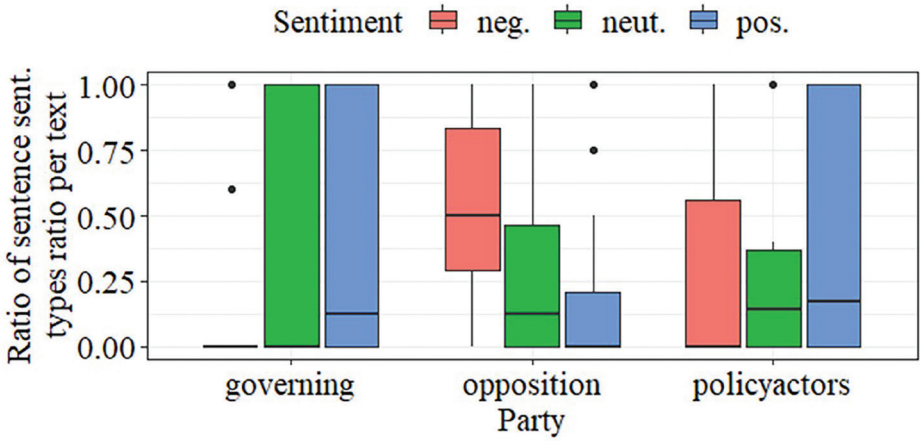

Figure 2. Result of the sentence sentiment ratio per text among governing party, opposition parties and policy actors (compiled by the authors)

As seen in Figure 3, the appearance of aspect sentiment emotions is as follows: it is rare for all groups to be neutral, but it is more common among policy actors to have predominantly neutral text. Positive sentiment is the rarest, and in the case of the opposition, positive sentiment occurs only randomly.

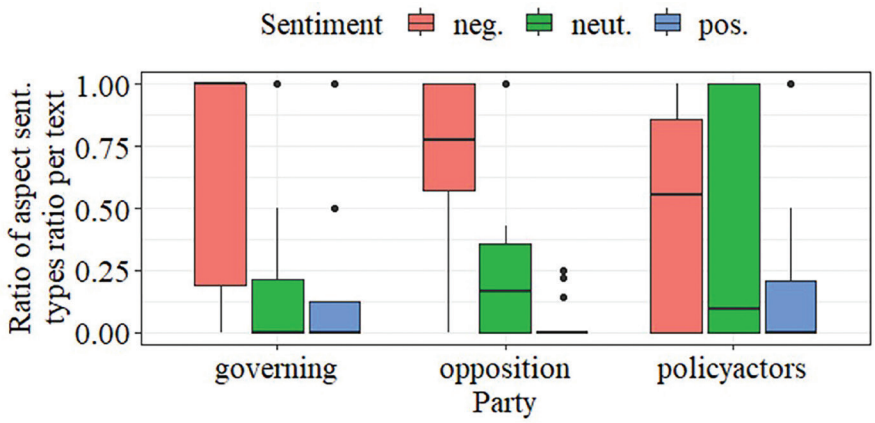

Figure 3. Result of the aspect sentiment ratio per text among governing party, opposition parties and policy actors (compiled by the authors)

However, it should be noted that the results of the aspect sentiment analysis were obtained for *kisadózó* and *KATA* expressions. While this is less problematic for *kisadózó*, *KATA* can appear in various contexts such as 'KATA modification',

'KATA cancellation', 'KATA acceptance', etc. In this sense, associated opinions cannot be consistent because someone who expresses negative sentiment about cancellation may express positive sentiment about acceptance. Consequently, we may choose not to use or apply a different method for the aspect sentiment values in the future.

The opposition party, at the same time, uses more negative utterances regarding the *KATA*. The (sentence) sentiment of the governing party is more positive (44,44%) and neutral (40,74%) compared with opposition party and policy actors, however, their opinion (aspect sentiment) shows negativity (55,56%). However, policy actors are not neutral.

Therefore H1 has been confirmed, but H2 has not.

4.3. Further results and observations

We were also curious about how emotions generally relate to term candidates and specialised term candidates. Thus, we examined whether there is any emotion present in a given text, regardless of its polarity. We can conclude that absolutely emotionless texts are rare.

For all texts, there is no significant difference observed between the ratio of term candidates and emotions, and there is no significant difference between specialised term candidates and emotions, either. In neutral texts, the proportions of specialised term candidates appear to be more balanced. However, when analysed by group, for term candidates, the rho value consistently remains below 0.2. In the case of specialised term candidates (as seen in Figure 4), a negative strong correlation is observed within the governing party (rho = −0.712, p = 0.002), indicating that the more emotional the text, the rarer the presence of specialised term candidates. This correlation is somewhat weaker and not significant in the case of the opposition (rho = −0.496). However, a moderately strong positive correlation is found among policy actors (rho = 0.586, p = 0.022). This suggests that in their case, the more emotional a text, the higher the ratio of specialised term candidates.

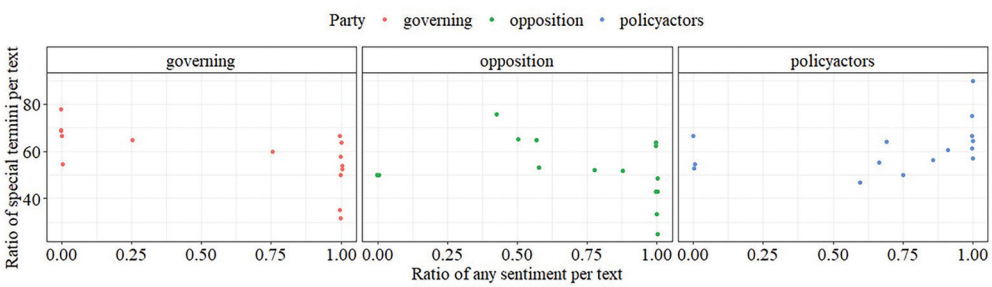

Figure 4. Ratio of specialised term and sentiment per text among the three groups (compiled by the authors)

5. Conclusion and further lines of research

The theoretical findings suggest that parliamentary discourse can be included in the scenarios of specialised communication, but the context and the level of knowledge and intentions of the participants are crucial. For example, the public participates in the parliamentary discourse only as receivers, and utterances must be interpreted in the context of political interests, attitudes and actions.

In this context, using terms can be political tools. As a persuasive strategy, they can be tools of appealing to the authority and credibility of the speaker. Based on political discourse analysis, they can be both tool for constructing reality (as they show that the speaker has specific knowledge) and tool for legitimisation (using language for its own purposes). They can also be tools of power: the use of terms, in the absence of comprehensibility, can control who can be a participant, and an informed recipient and interpreter of a discourse.

Policy actors use a high number of terms, and in particular specialised terms, and this suggests that they are primarily expressing expert utterances. The use of terms by government party members was only slightly behind that of policy actors. In their case, a deliberate, high proportion of terms can indicate professionalism. For them, however, it can be a means of legitimisation (appealing to credibility and competence), a strategy of influence (the public cannot check the credibility of the content due to lack of intelligibility, but can consider the speaker credible on the basis of authority) and a technique of power (if the discourse is not understandable to the public, understandable accessibility does not apply, and public policy discourse is controlled). The slight use of terms by opposition political actors may reflect either a focus on informing the public (comprehensibility) or the appearance of it, in order to build trust.

The speeches also express evaluations, political and ideological opinions and attitudes. As a persuasive strategy, these can appeal to the emotions of the audience. These can be tools of reality construction and legitimation, by which the speaker can present his or her own claims as positive, credible and legitimate, while others' are negative and delegitimised.

These can also shape the framework of interpretation, which can help to make the topic more understandable. Referring to Gennaro and Ash (2022), although the proportion of emotions in political discourse depends on both genre and topic, the emergence of public coverage of congressional or parliamentary debates correlates with an increase in emotionality within debates; the emotionality of taxation and redistribution issues has increased in recent years; and emotional rhetoric is highlighted for those in minority positions (e.g. opposition).

The taxation topic we examined was a hot topic in Hungary during the period under study, and the results of the research also show that emotionless speeches were rare. The result that there is no significant difference between the proportion of terms and emotions, or between specialised terms and emotions, suggests that all speakers appeal to both authority and emotions. However, the result that the more emotional a speech of a governing party or opposition party actor is, the less frequent the presence of specialised terms in it, while the more balanced proportion of

specialised terms in neutral polarity texts shows that speakers with less emotionality tended to use specialised communication, while speakers with more emotionality tended to use political communication.

Both the sentence and the aspect sentiment of the opposition parties is more negative, which may indicate a polarisation strategy, a negative framing of the proposals of the other side. The texts of governing party actors are rarely negative, so they can build on the positive framing. For policy actors, several types of emotion appear, but positive is the most varied (regardless of the number of utterances in the text), while neutral is the least varied. However, for policy actors, the more emotive a text is, the higher the proportion of specialised terms, which is a surprising result.

This may point to methodological limitations of the research, but it may also suggest that policy actors, however much they want to engage in professional communication, are still engaged in a political discourse, the contextual factors of which may influence their utterances. The possible interpretations of aspect values are also limited (see Section 4.2): although these showed negative values for all actors, it is hard to interpret without context, so there is a need to improve the methodology.

The empirical results are well positioned in the discourse on the relationship between public policy and expert knowledge (technocracy and democracy). One way to ensure comprehensibility and public engagement could be to reduce the number of specialised terms, but comprehensibility is not just about avoiding terms because terms convey accurate specific knowledge. Appealing to emotions, attitudes and beliefs can be a useful tool for comprehensibility, because they can shape the interpretative framework. However, this framing should not be a tool for manipulation in order to ensure truthful and credible informing and engagement. The topic under discussion needs to be framed in a way that makes it clear to the public how it is relevant to their lives, so that they are motivated to engage with the discourse (cf. Falyuna 2022).

Our results can also provide insights for studies on populist communication. As argued by Gennaro and Ash (2022), the increasing trend of emotional rhetoric is accompanied by an increase in polarisation, so there may be a shift towards a rhetoric that appeals to voters rather than politicians and elites.

The results of the research also contribute to the computer-assisted methodological aspects of the analysis of both specialised and political discourses. In our opinion, terms and emotionality are elements of discourses that can be well analysed quantitatively and with computer tools.

The linguistic and social context of the discourse is important for interpreting the results and clarifying the aspects of the analysis, so that further development of the methodology is needed. At the same time, the research methodology highlights the existence of the ParlaMint corpus, which provides a rich, open-access resource for researchers in the humanities and social sciences.

The results and methods of the research suggest further lines of research. An exciting approach could be to explore extended topics (e.g. taxation) or to select a sub-corpus to be examined with automatic topic detection. The automated term extraction mentioned earlier is also one of the developments in the research that we would like to use.

An interesting research question might be how the type of parliamentary utterances (speeches or comments, i.e., pre-written texts or spontaneous comments) influences the presence of emotions and terms. Further, the reality-constructing nature of political discourse raises the question of whether terms are used with their specialised meanings or whether the context – even during the transition to standard language – shapes their meanings, thus potentially losing their terminological character. Exploring this raises new perspectives on certain questions.

References

Cabré, María Teresa. 'Theories of Terminology: Their Description, Prescription and Explanation'. *Terminology* 9 (December 2003): 163–200. https://doi.org/10.1075/term.9.2.03cab

Cabré, María Teresa, Iria da Cunha, Eric SanJuan, Juan-Manuel Torres-Moreno and Jorge Vivaldi. 'Automatic Specialised vs. Non-specialised Text Differentiation: The Usability of Grammatical Features in a Latin Multilingual Context'. In *Languages for Specific Purposes in the Digital Era,* edited by Elena Bárcena, Timothy Read and Jorge Arús, 223–241. New York: Springer, 2014. https://doi.org/10.1007/978-3-319-02222-2_11

Collins, Harry, and Robert Evans. *Rethinking Expertise.* Chicago, IL: University of Chicago Press, 2007.

Dunn, Olive Jean. 'Multiple Comparisons Using Rank Sums'. *Technometrics* 6, no. 3 (1964): 241–252. https://doi.org/10.2307/1266041

Erjavec, Tomaž, Matyáš Kopp, Maciej Ogrodniczuk, Petya Osenova, Manex Agirrezabal, Tommaso Agnoloni, José Aires, Monica Albini, Jon Alkorta, Iván Antiba-Cartazo, Ekain Arrieta, Mario Barcala, Daniel Bardanca, Starkaður Barkarson, Roberto Bartolini, Roberto Battistoni, Nuria Bel, Maria del Mar Bonet Ramos, María Calzada Pérez, Aida Cardoso, Çağrı Çöltekin, Matthew Coole, Roberts Darģis, Ruben de Libano, Griet Depoorter, Sascha Diwersy, Réka Dodé, Kike Fernandez, Elisa Fernández Rei, Francesca Frontini, Marcos Garcia, Noelia García Díaz, Pedro García Louzao, Maria Gavriilidou, Dimitris Gkoumas, Ilko Grigorov, Vladislava Grigorova, Dorte Haltrup Hansen, Mikel Iruskieta, Johan Jarlbrink, Kinga Jelencsik-Mátyus, Bart Jongejan, Neeme Kahusk, Martin Kirnbauer, Anna Kryvenko, Noémi Ligeti-Nagy, Nikola Ljubešić, Giancarlo Luxardo, Carmen Magariños, Måns Magnusson, Carlo Marchetti, Maarten Marx, Katja Meden, Amália Mendes, Michal Mochtak, Martin Mölder, Simonetta Montemagni, Costanza Navarretta, Bartłomiej Nitoń, Fredrik Mohammadi Norén, Amanda Nwadukwe, Mihael Ojsteršek, Andrej Pančur, Vassilis Papavassiliou, Rui Pereira, María Pérez Lago, Stelios Piperidis, Hannes Pirker, Marilina Pisani, Henk van der Pol, Prokopis Prokopidis, Valeria Quochi, Paul Rayson, Xosé Luís Regueira, Michał Rudolf, Manuela Ruisi, Peter Rupnik, Daniel Schopper, Kiril Simov, Laura Sinikallio, Jure Skubic, Lars Magne Tungland, Jouni Tuominen, Ruben van Heusden, Zsófia Varga, Marta Vázquez Abuín, Giulia Venturi, Adrián Vidal Miguéns, Kadri Vider, Ainhoa

Vivel Couso, Adina Ioana Vladu, Tanja Wissik, Väinö Yrjänäinen, Rodolfo Zevallos and Darja Fišer. 2023. 'Multilingual Comparable Corpora of Parliamentary Debates ParlaMint 4.0'. Accessed February 11, 2024.
www.clarin.si/repository/xmlui/handle/11356/1859

Faber, Pamela. 'The Pragmatics of Specialised Communication'. *Entreculturas. Revista De Traducción y Comunicación Intercultural* 1 (2009): 61–84.
https://doi.org/10.24310/Entreculturasertci.vi1.11818

Falyuna, Nóra. 'Az áltudományos szövegek pragmatikai vizsgálata: Esettanulmány a lúgosító diéta példáján'. *Századvég* 84 (2017): 85–107.

Falyuna, Nóra. *Az áltudományos diskurzus jellemzői: Az újmédiában megjelenő áltudományos tartalmak nyelvészeti-kommunikációs elemzése.* Doktori disszertáció. Budapest: ELTE BTK, 2022.
https://doi.org/10.15476/ELTE.2022.014

Feldman, Ronen. 'Techniques and Applications for Sentiment Analysis'. *Communications of the ACM* 56, no. 4 (2013): 82–89.
https://doi.org/10.1145/2436256.2436274

Fóris, Ágota. *Hat terminológia lecke*. Pécs: Lexikográfia, 2005.

Gennaro, Gloria, and Elliott Ash. 'Emotion and Reason in Political Language'. *Economic Journal* 132, no. 643 (April 2022): 1037–1059.
https://doi.org/10.1093/ej/ueab104

Hendriks, Friederike, and Dorothe Kienhues. 'Science Understanding between Scientific Literacy and Trust: Contributions from Psychological and Educational Research'. In *Science Communication*, edited by Annette Leßmöllmann, Marcelo Dascal and Thomas Gloning, 29–50. Berlin: De Gruyter Mouton, 2019.

Hungarian National Assembly. n.d. 'Website of the Hungarian National Assembly'. Accessed February 11, 2024.
www.parlament.hu/web/house-of-the-national-assembly

Jerit, Jennifer, James H. Kuklinski and Paul J. Quirk. 'Strategic Politicians, Emotional Citizens, and the Rhetoric of Prediction'. In *The Political Psychology of Democratic Citizenship*, edited by Eugene Borgida, Christopher M. Federico and John L. Sullivan, 100–24. New York: Oxford University Press, 2009.

Kenzhekanova, Kuralay Kenzhekankyzy. 'Linguistic Features of Political Discourse'. *Mediterranean Journal of Social Sciences* 6 (2015): 192–199.

Kotenko, Olga, Natalia Kosharna, Myroslaca Chepurna, Olha Trebyk and Ivan Bakhov. 'Linguistic and Stylistic Features of English Public Speeches'. *World Journal of English Language* 13, no. 4 (2023): 56–60.
https://doi.org/10.5430/wjel.v13n4p56

Kurtán, Zsuzsa. 'Szaknyelv és szakmai kommunikáció'. In *Szaknyelvi kommunikáció*, edited by Csilla Dobos, 11–22. Miskolc–Budapest: Miskolci Egyetem–Tinta Könyvkiadó, 2010.

Laki, László János, and Zijian Győző Yang. 'Sentiment Analysis with Neural Models for Hungarian'. *Acta Polytechnica Hungarica* 20, no. 5 (2023): 109–128.
https://doi.org/10.12700/APH.20.5.2023.5.8

Lakoff, Robin Tolmach. *Talking Power: The Politics of Language*. New York: Basic Books, 1990.

Ligeti-Nagy, Noémi, Réka Dodé, Kinga Jelencsik-Mátyus, Zsófia Varga, Enikő Héja and Tamás Váradi. 'A New ParlaMint Corpus for Hungarian 30m Tokens of Annotated Parliamentary

Data'. In *XIX. Magyar Számítógépes Nyelvészeti Konferencia, MSZNY-2023*, edited by Gábor Berend, Gábor Gosztolya and Veronika Vincze, 433–46. Szeged: Szegedi Tudományegyetem, 2023.

Lin, Chien-Huang, Shih-Chieh Chuang, Danny T. Kao and Chaang-Yung Kung. 'The Role of Emotions in the Endowment Effect'. *Journal of Economic Psychology* 27, no. 4 (2006): 589–597.
https://doi.org/10.1016/j.joep.2005.10.001

MacLeod, W. Bentley. 'Decision, Contract, and Emotion: Some Economics for a Complex and Confusing World'. *Canadian Journal of Economics* 29, no. 4 (1996): 788–810.
https://doi.org/10.2307/136215

Merkovity, Norbert. 'Az állampolgárok az új politikai kommunikációban'. *Információs Társadalom* 12, no. 1 (2012): 64–79.
https://dx.doi.org/10.22503/inftars.XII.2012.1.3

Mifrah, Sara, and El Habib Benlahmar. 'Sentence-Level Sentiment Classification: A Comparative Study Between Deep Learning Models'. *Journal of ICT Standardization* 10, no. 2 (2022): 339–352.
https://doi.org/10.13052/jicts2245-800X.10213

Nemeskey, Dávid Márk. *Natural Language Processing Methods for Language Modeling*. PhD dissertation. Budapest: Eötvös Loránd University, 2020.
https://doi.org/10.15476/ELTE.2020.066

Ogle, Derek H., Jason C. Doll, A. Powell Wheeler and Alexis Dinno. 2023. 'FSA: Simple Fisheries Stock Assessment Methods. R Package Version 0.9.5'. Accessed March 5, 2024.
https://CRAN.R-project.org/package=FSA

Pang, Bo, Lillian Lee and Shivakumar Vaithyanathan. 'Thumbs Up? Sentiment Classification Using Machine Learning Techniques'. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002),* 79–86. Philadelphia: University of Pennsylvania, 2002.
https://doi.org/10.3115/1118693.1118704

Pontiki, Maria, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos and Suresh Manandhar. 'SemEval-2014 Task 4: Aspect Based Sentiment Analysis'. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 27–35. Dublin: Association for Computational Linguistics, 2014.
https://doi.org/10.3115/v1/S14-2004

R Core Team. 2022. 'R: A Language and Environment for Statistical Computing'. Accessed March 1, 2024.
www.R-project.org/

Radaelli, Claudio M. 'The Role of Knowledge in the Policy Process'. *Journal of European Public Policy* 2, no. 2 (1995): 159–183.
https://doi.org/10.1080/13501769508406981

Reyes, Antonio D. 'Strategies of Legitimization in Political Discourse: From Words to Actions'. *Discourse & Society* 2, no. 6 (2011): 781–807.
https://doi.org/10.1177/0957926511419927

Schirm, Anita. 'A diskurzusjelölők a parlamenti beszédekben'. In *Hatékony nyelvi, idegen nyelvi és szakmai kommunikáció interkulturális környezetben*, edited by Katalin Kukorelli, 168–175. Dunaújváros: Dunaújvárosi Főiskola, 2009.

Sinclair, John. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.

Sketch Engine. 2024a. 'NoSketch Engine'. Accessed February 11, 2024. https://nlp.fi.muni.cz/trac/noske

Sketch Engine. 2024b. "Sketch Engine tool." Accessed February 11, 2024. https://www.sketchengine.eu/.

Szabó, Márton. 'Diskurzuselemzés és politikatudomány'. *Politikatudományi szemle* 12, no. 3 (2003): 41–58.

Szabó, Márton. 'Közpolitikai diskurzuselemzés'. *Working Papers in Political Science* 7 (2012): 1–34.

Stehr, Nico. 'Néhány marxi gondolat érvényessége az információs társadalomban'. *Információs Társadalom* 7, no. 3 (2007): 6–14. https://dx.doi.org/10.22503/inftars.VII.2007.3.1

Stehr, Nico. 'A tudás társadalmi kontrollja a demokratikus társadalmakban'. *Információs Társadalom* 10, no. 1 (2010): 7–25. https://dx.doi.org/10.22503/inftars.X.2010.1.1

Thomas, Matthew, and Luke Buckmaster. 2013. 'Expertise and Public Policy: A Conceptual Guide'. Research Paper 2013-14. Parliament of Australia, Department of Parliamentary Services. Accessed February 11, 2024. www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp1314/PublicPolicy#_ftn1

Thon, Franziska M., and Regina Jucks. 'Believing in Expertise: How Authors' Credentials and Language Use Influence the Credibility of Online Health Information'. *Health Communication* 32, no. 7 (2017): 828–836. https://doi.org/10.1080/10410236.2016.1172296

Turner, Stephan P. *Liberal Democracy 3.0: Civil Society in an Age of Experts*. London: Sage, 2003.

Vadai, Kata. 'Critical Discourse Analysis in Progress: The Power, Ideology and Manipulation Identification (PIMI) Model'. *Alkalmazott Nyelvtudomány* 16, no. 1 (2016). http://dx.doi.org/10.18460/ANY.2016.1.004

van Dijk, Teun A. 'What Is Political Discourse Analysis?'. *Belgian Journal of Linguistics* 11 (1997): 11–52. https://doi.org/10.1075/bjl. 11.03dij

Wilson, John. 'Political Discourse'. In *The Handbook of Discourse Analysis*, 2nd ed., edited by Deborah Tannen, Heidi E. Hamilton and Deborah Schiffrin, 775–794. Hoboken, NJ: Wiley Blackwell, 2015.

Wodak, Ruth, and Paul Chilton. *A New Agenda in (Critical) Discourse Analysis: Theory, Methodology and Interdisciplinarity.* Amsterdam: John Benjamins, 2005.

Zimányi, Árpád. 'A politikai-közéleti nyelv érzelmi színezete'. In *Az agressziókutatásról interdiszciplináris keretben*, edited by Árpád Zimányi, 114–122. Eger: Líceum Kiadó, 2008.

NÓRA FALYUNA, KATALIN FEHER, MÁRTON DEMETER, GÁBOR SZÜDI, JOSEPH E. USCINSKI

Against Disinformation: Bridging Science and Public Discourse

The Science and Society Research Group at the Ludovika University of Public Service organised a conference in Budapest on 13 October 2023 titled: 'Science and disinformation: how science can support society against disinformation'. The conference explored the complex relationship between science, society, and disinformation. A panel discussion was held with the aim of examining the challenges and strategies for science communication in the context of the ever-increasing issues of dis-/misinformation. The discussion highlighted the multifaceted factors that influence public trust in science; the role of digital culture on science communication and scientific knowledge production; the impact of artificial intelligence; and the relationship between science and business developments on science disinformation, especially on science-related conspiracy theories. The participants helped shed light on how science communication strategies can be improved to stabilise trust in a multi-stakeholder information environment. This article summarises the main insights and conclusions from the discussion.

Keywords: *conspiracy theory, digital technology, disinformation, science communication, scientific knowledge*

Author Information

Nóra Falyuna, Ludovika University of Public Service
ORCID: 0000-0002-5636-903X, E-mail: falyuna.nora@uni-nke.hu
Katalin Feher, Ludovika University of Public Service
ORCID: 0000-0003-3293-0862, E-mail: feher.katalin@uni-nke.hu
Márton Demeter, Ludovika University of Public Service
ORCID: 0000-0002-9888-9682, E-mail: demeter.marton@uni-nke.hu
Gábor Szüdi, Centre for Social Innovation (ZSI)
ORCID: 0000-0001-8006-4052, E-mail: szudi@zsi.at
Joseph E. Uscinski, University of Miami
ORCID: 0000-0003-2179-6952, E-mail: uscinski@miami.edu

How to cite this article:
Falyuna, Nóra, Katalin Feher, Márton Demeter, Gábor Szüdi, Joseph E. Uscinski. "Against Disinformation: Bridging Science and Public Discourse". *Információs Társadalom* XXIV, no. 2 (2024): 68–84.
https://dx.doi.org/10.22503/inftars.XXIV.2024.2.4

1. Introduction[1]

Interpreting the effects of disinformation or misinformation has been a key issue in recent years with the ever-increasing spread of dis-/misinformation. The reasons and circumstances for the phenomena of disinformation and misinformation are complex. First, there have been many changes, for example, in the nature of publicity and the construction of reality, the range and behaviour of communication actors, the way information flows, who holds information power, the structure of trust, the credibility of information sources, the development of digital technologies, the expansion of global communication possibilities, and in the world seemingly entering a 'post-truth era'. Concerns have been raised as disinformation and misinformation can significantly affect a range of issues, including public opinion, social conflicts, trust in institutions, crisis management, capital accumulation, corporate reputations, and information warfare (Feher and Veres 2023; Lazer et al. 2018; Tandoc 2019). If one considers some significant world events of recent years, for example the COVID-19 pandemic (see e.g. Islam et al. 2020), these have seemingly favoured the spread of fake news, science disinformation, conspiracy theories, and a wide range of misleading information that it could be crucial to tackle. While proposals for solutions to tackle these issues are multifaceted, it is essential that experts on the subject who represent different disciplines should work in cooperation and communicate with each other to find working solutions. There is also a vital need for strategies for more effective communication between science and society. Indeed, in today's information society, effective science communication is increasingly relevant for dealing with disinformation and misinformation. It can also improve the relationship between science and policymakers, industry and the market, civil society, and different academic disciplines. Besides this, effective science communication can strengthen the role and support for science in society and reduce the social alienation of science (for these points cf. e.g. NASEM 2017). Finally, it could support further tech-based or AI-driven solutions as they emerge in the future (Feher et al. 2024).

The public's understanding of science and conversely, how science comprehends society are equally important today, as is the understanding the phenomenon of disinformation in the modern era, which could help facilitate the cognitive 'inoculation' of people and public participation in science communication and the social engagement of scientific actors (Falyuna 2022). However, the study of such things and the practices that support them also requires a complex approach.

[1] Nóra Falyuna and Katalin Feher give thanks for support from the János Bolyai Research Scholarship of the Hungarian Academy of Sciences for the paper (grant numbers: BO/00729/23/9 and BO/00045/19/9). Katalin Feher is grateful for support from the European Union's Horizon Europe Research and Innovation Programme – NGI Enrichers, Next Generation Internet Transatlantic Fellowship Programme, who provided funding for the research project and publication (grant number: 101070125). Gábor Szüdi is thankful for support from the European Union's Horizon Europe Research and Innovation Programme – VERITY (deVEloping scientific Research with ethIcs and inTegritY) for also supporting the paper (grant number: 101058623). Finally, Joseph E. Uscinski thanks the National Science Foundation (US) SaTC programme (award number: 2123635) and National Institute of Health (grant number: R01MD018343) for their support for the paper.

On 13 October 2023, the Science and Society Research Group of the Ludovika University of Public Service organised and held a conference in Budapest titled: 'Science and Disinformation: How can Science support Society against Disinformation?' (Harangozó 2023). This event focused on the complex relationship between science, society, and disinformation. The key focus was on the impact caused by the spread of disinformation, fake news, conspiracy theories, and pseudo-scientific or anti-scientific beliefs on public opinion, social publicity, and social trust in institutions. Disinformation as a weapon was discussed in its relationship to information and cyber security, as well as in the context of applications of technology, with a particular focus on artificial intelligence (AI), considering the debates on these issues in social, political, scientific, and media discourse.

Additionally, the conference panellists highlighted various potential practical solutions, particularly in regard to science communication, and the role of science in policymaking, media/platform regulation, content management of social media, and the development of information, and digital and media literacy.

In the comprehensive, interdisciplinary panel discussion at the conference, five experts explored the challenges and opportunities for science communication based on the latest research. Below, we present the panel discussion, which focused on the following key topics:

- the issue of social trust and distrust in science;
- the logic of conspiracy theories, especially science-related conspiracy theories;
- the factors and circumstances that influence society's perception of science;
- the impact of digital culture on science communication and the processes of scientific knowledge production;
- the impact of artificial intelligence technology on science, disinformation, trust, and credibility;
- the issue of the relationship between science and business.

That session was attended by the present paper's authors, namely Nóra Falyuna, whose main research area is science communication, pseudoscience and anti-science, and disinformation in digital communication; Katalin Feher, whose main research area is AI media, generative AI, and sociotechnical systems and their ethics; Márton Demeter, whose main research area is science metrics, publication networks and strategies, academic knowledge production, and communication and media studies; Gábor Szüdi, whose main research area is the knowledge economy, social innovation, research and innovation policies, and science communication; and Joseph E. Uscinski, whose main research area is political science, public opinion, and mass media, with a special focus on conspiracy theories.

2. Science communication in a (dis)information society

2.1. Social trust and distrust in science

Nóra Falyuna: Let us begin by discussing the topic of conspiracy theories. At the conference, one of the key focuses was the impact of conspiracy theories, which

have become increasingly intertwined with disinformation and misinformation. As misinformation becomes increasingly problematic, it is essential to consider how science should communicate and how trust, or distrust, in science can manifest. The potential consequences of science-related conspiracy theories are significant because they can erode trust in science and its fundamental tenets. As such, it is imperative that this issue be addressed with care and attention to ensure that society remains informed and engaged with science. Professor Uscinski, what do you think the link between conspiracy theories and trust in science is?

Joseph E. Uscinski: People pick and choose their conspiracy theories, so very rarely do people who distrust science distrust all of science. Here is a good example: People who say they do not trust Big Pharma or the medical establishment might not get vaccinated, but if they were to break their femur, they would go straight to a hospital, they would get a cast, and they would not say, 'I do not trust big plaster, so I am not going to get a cast, and I will just let it heal naturally'. But when it comes to something like vaccines, they will say, 'Oh, I do not want to get the vaccine. It is unnatural, and I do not trust them. They just want to make money from me, and I only like natural things'. So people always pick and choose which science they like and which science they do not like. There is rarely if ever a distrust of all of science. In some sense then trust matters, but it is important to know where and when people choose to direct their trust and distrust. Often there are organised groups, political leaders, or people in the media who attract audiences that have distrustful views and then focus those audiences on something specific. Same thing with people who say, 'Oh, I do not like genetically modified food', but then eat all sorts of things that are equally modified and probably not great for their health. They apply their distrust wherever they see fit at any given moment.

Nóra Falyuna: Today's information societies face a paradox regarding the relationship between science, society, and disinformation. Although expertise is highly valued and in demand across all sectors, the credibility of scientific claims and trust in experts is being challenged (or they have a limited impact on public opinion, see e.g. Klofstad and Uscinski 2023), particularly in the online public sphere. This begs the question: Can we identify the factors and circumstances that shape trust in science and society's perception of science? Also, is there a widespread distrust of science? Moreover, what new questions could arise for science communication due to these factors and circumstances?

Márton Demeter: It was during Dr Szüdi's presentation on the TRESCA project that we realised that we need to change our idea of science as the sole basis for decision-making. It is much better to recognise that science should inform the public, and it also has the potential to inform policymakers. However, this is no longer the only possible perspective. As scientists, we must acknowledge and understand this. I believe the first step is to recognise that we need to change our perception of science and scientists. Only then can we consider if and how we can influence society itself.

Katalin Feher: Trust is an issue in this context. How can we translate a complex subject for society? This task is absolutely essential and a notable challenge in the scientific domain. Reflecting on the early months of the COVID-19 pandemic, there was a clear demand for accurate information from scientists and politicians, highlighting the importance of their collaboration despite them not always finding common and trusted ground. Often, disagreements and conflicts of interest came up. In this context, social media influencers and people claiming to be experts tried to make their own space, which added to the confusion. Scientists, in response, were compelled to defend their prestige through effective communication, while navigating the political landscape and the surge of various experts appearing on social media, ranging from genuine experts to mere amateurs on COVID-19. This scenario highlighted the urgent need for a deeper understanding of what had transpired and the adoption of risk-management strategies, a concept derived from business studies, but equally vital in anticipating societal and environmental risks. With well-crafted plans and guidance, scientists can be better equipped to face new risks and potential disasters, maintaining their crucial role in informing and protecting society and building societal trust.

Gábor Szüdi: OK, let me go back to the first recommendation of the TRESCA study. There was a very optimistic part in it saying that the reputation of scientists would increase. That was what we thought at the beginning of the COVID-19 pandemic. I am not so convinced about that any longer. We were trying to stress this evidence-informing aspect, which I think is very important. The question is: How can you communicate evidence not just to policymakers, but also to the general public? For instance, it is really challenging to communicate uncertainty in science because – as you have said – people want clear-cut answers, which we do not have. And let's not talk about social sciences because these usually do not deal with life and death issues, but COVID-19 was really about life and death. And there are different opinions on how to communicate uncertainty. Some say that we should not communicate it at all, and others say that we should, but in a way that is not oversimplified. It should rather show people proof that you cannot be absolutely certain of things but could still showcase scientific consensus. I think this honesty could lead to a better appreciation of science in the long term.

Joseph E. Uscinski: So, I am not shocked that almost 20% of Americans have not received even one COVID shot, but I am amazed that 81% or 82% did, because I find that trying to get anyone to do anything is like herding cats. Many people do not want to do what they are told to do. In a classroom, if I give an assignment to my students, half of them will do it, some will turn it in on time, some will turn it in some other time, and some will do it completely wrong. And that is in a situation in which I have methods to make them do things, because there will be negative consequences if they don't, like getting a poor grade. So the fact that 80% of people in my country (US) actually went and got the shot was shocking, because I just did not think that would have been achieved, especially with what had been going on up until that point. I think we should reframe the questions. Instead of: 'Why aren't people doing what they are supposed to do?', we could ask: Why did so many people do what they were supposed

to do? Which I think is a really neat thing. Furthermore, I would add that we (on this panel) are not 'normal' people. We are in the top 1% of the world's most educated. We are not 'the average Joe on the street.' If we want to know about climate science, we can talk to climate scientists. If we want to think about data, we can examine the data ourselves. Probably, we know people who work in vaccines, climate science, and in every branch of science. We know these scientists, and we understand them. Furthermore, many of us in this room even know politicians, and we occasionally interact with them, and we say, 'Oh yeah, they are normal people just like us'. But 'Joe on the street' does not have access to us or other scientists or the politicians we have access to. To regular folks, they see these distant, faceless, nameless institutions, people they do not know and will never meet or interact with and they say, 'Who are those people and why should I trust them?' So, the question is not really, why do people distrust these things? Sometimes, I wonder why they would trust Congress at all. Why would they trust the NHS or the CDC at all? Because they do not know them, and they do not interact with them. I think maybe one thing going forward is we ought to start thinking about how we can make people feel empowered and then give them more access to scientists. They would hopefully then say, 'Oh, yeah, they are normal people, not part of some plot to stick a tracking device in my neck'.

Nóra Falyuna: Dr Szüdi, you are contributing to the VERITY Horizon Europe Project, the aim of which is to strengthen public trust in scientific research, to encourage informed decision-making based on scientific evidence, and to promote a better relationship between science and society. Could you share some details about the VERITY project and its achievements so far in achieving these goals?

Gábor Szüdi: VERITY is an acronym derived from the long project title of 'developing scientific research with ethics and integrity'. It is about trust in science. It builds on the results of the TRESCA project, but I think the main difference here is that we also want to check the influence of other, non-traditional stakeholders in science. For instance, people such as influencers, social media platforms, and their interaction with other stakeholders. And I think that is something that has not been really been taken into account until now. We started only last September, but we already have some results on the state-of-the-art sources of mistrust in science. We research not only how people perceive issues such as COVID-19, but we are also focusing on increasing numbers of other topical issues, such as global warming or moving forward to carbon neutrality. We check how stakeholders work together around these scientific issues to address mistrust. We call this 'the ecosystem of trust', which I think is a nice term. For example, we organised a workshop with journalists in Austria, and they all told us that they feel neglected, and feel that they are not as essential as they used to be because people are increasingly getting their scientific facts from social media or influencers. This can be problematic because, of course, there is no gatekeeping function in social media, which they had provided. The question is how we can we solve this? We could implement a hard-line solution and, of course, ban content or introduce some restrictions on social media. Or we could start to understand how to involve these social media influencers in solving this problem. We

could start some normal conversations on these scientific issues. The whole process should lead to a so-called protocol that contains recommendations for the whole ecosystem, not just for academic people. The recommendations will go to the Commission, so hopefully, they can find some substantiated information for their own people, too.

Nóra Falyuna: And what are the main differences between VERITY and the other Horizon 2020 project TRESCA?

Gábor Szüdi: The objectives are similar, but as I have said, approached from a more systematic view in the case of VERITY project. We want to understand the reasons for mistrust in science and to identify the right methods to deal with them. In terms of methodology, the projects are similar in that we use various participatory approaches, such as workshops or focus groups. At the end of the projects, we would like to write various articles and produce policy recommendations. The main difference in VERITY is this ecosystem view. For instance, in TRESCA, when you check the work packages, one is about policymakers. Now we are trying to bring all the stakeholders together in each task, which is a challenge, but I think it is the way forward because these people should cooperate with each other. If we really want people to believe more in scientists and policymakers, then that is even more difficult than, for instance, increasing trust in journalists. So, I would again just say that we all need to work together, which differentiates this from previous projects. The partnership is also different. Previously, for instance, we had a partner on board producing videos, so that was a bit more visual.

2.2. Advantages and disadvantages of digital technology in the production of scientific knowledge

Nóra Falyuna: Digital culture has had a significant impact on the production of scientific knowledge, with open science and public access to preprint publications being notable examples. These developments have facilitated real-time, global communication and collaboration in the scientific community, making it easier to share knowledge with the public (Fraser et al. 2021). However, the availability of preprint publications during the COVID–19 pandemic has been linked to the spread of misinformation (Koerber 2021). This raises the question of how the scientific publishing process can be compatible with situations where science has to react and communicate quickly.

Márton Demeter: Science should not provide answers hastily; originally, being fast was not part of the scientific self-definition. It is a connection with business, markets, and the neoliberal logic of academia, of which I am keenly aware because I work in the science publishing industry. Thus, I frequently discuss ways of expediting the publication process. While there are methods for publishing more quickly, being fast was not originally part of a scientist's self-identity. This inclination towards speed

reflects a capitalist ideology that everything should be treated as a product. It's intriguing that we now talk about science production or the production of science, treating papers as productions that somehow measure the productivity of individual researchers and institutions, universities, and countries. Perhaps we should not prioritise publishing hastily. Moreover, I would like to emphasise that we are not 'normal' in the sense that, as scientists, we want to provide the truth. I am not sure if our perspective applies to all of society, as people may seek fun, and many other things not directly related to what we provide. So, we can offer solutions promptly, but in many cases it may not be what the audience truly needs. To illustrate, during a recent talk show on Hungarian television discussing vaccination, five professionals, all in life sciences, biologists, etc., provided different perspectives. In the end, when asked which vaccination they recommended, each suggested a different type, AstraZeneca, Pfizer, etc. Even these five scientists could not agree on the best option. For scientists, such disagreements are common in conferences, but it can be confusing for the general public, who are seeking solutions and peace of mind. The production of science is uncertain, and it is unclear whether it can provide solutions for the general public, who may not even need the information in many cases.

Joseph E. Uscinski: When you think about these two institutions, science and government, these are two institutions that are predicated on distrust. Peer review was built on the idea that you have to convince other experts of your ideas, you do not get to publish them just because you want to. Democratic societies are built on the idea that we should have a healthy scepticism of our leaders. So it is very difficult to then turn around and say, 'Well, you should trust government all the time', because you really should not trust government all the time; and you cannot say you should trust science all the time, because even we do not trust science all the time. Good scientists have a very sceptical view of everything they read, particularly if they are peer reviewing. And I really like the example that you just gave. I think it may be even worse than that, because during the COVID-19 pandemic random doctors were coming out of the woodwork in the US saying things like, 'Oh no, this is all made-up, and you can cure your COVID if you just walk barefoot on a sandy beach'. In that sense, people can follow political leaders and experts and still be wrong if they pick the wrong experts and leaders.

Gábor Szüdi: On the VERITY project, I had an interview with a person working in a field related to trust in science, and he told me that this whole trust in science is something that we should abandon because you should not trust science at all. You should always be sceptical about science. I understand this point of view. I would not abandon trust in science, but there should be healthy scepticism about all these things. What is interesting, and we must also be aware of it, is that even people who say they do not trust science do trust some people or some organisations. So, science is a very abstract concept. However, if I ask you, do you trust Joe is telling you the truth, they would say, 'Of course, but I do not trust this whole COVID-19 business at all'. So, if you go down to a smaller, more personal level then I think the situation is not so bad, that people do not trust science. But then again, coming back to what we were discuss-

ing, I think that you shouldn't lie about the uncertainties. For example, if you are not sure which vaccinations are best, then you should tell them, 'Yeah, we have these and these facts and we are working on it', because otherwise it can just cause confusion and even mistrust. We are also seeing this with other topics, such as climate science.

Katalin Feher: Considering science as an abstract concept allows people to seek facts and information, yet perceptions of what constitutes science vary widely. For some, science is epitomised by prestigious accolades, such as the Nobel Prize, reputable journals, and recognisable scientific brands. These elements symbolise science in specific contexts. However, the significance of science in policymaking diverges, perhaps only intersecting in aspects like the influence of notable brands, which can serve as a common ground. Nonetheless, the scientific research process – characterised by extensive, often decades-long studies and the necessity of revising and expanding knowledge continually – may seem too gradual and intricate for the fast-paced demands of policymaking. This difference in pace is particularly evident in the slow peer-review process, where the acceptance or rejection of research findings can take several months. Failures and rejections, while common in the scientific process, are typically invisible to the public, who may only see the success stories and prominent figures in the field.

Gábor Szüdi: That is why I consider open science to be very important. There is the Open Research Europe platform where you can publish more quickly. It is also the Commission's intention that researchers will publish something within the project framework within 2–3 years, but the peer review bit comes afterwards. We submitted the TRESCA results on that platform, and received many questions from the editors, for instance, about the methodology. Afterwards, we were published, but the peer-review process has still not concluded after one and a half years, but the preliminary project results are available to the public. I think that is also something that we must decide. Do we want the results to be published more quickly, or do we want to publish only what has already been peer-reviewed? Otherwise, we are going to be in the situation once again where even the best intentions of open science could lead to controversial results, due to which people may not trust science. Scientists have said, not just from social sciences, but also from medical sciences, that they cannot publish negative results. I am not against open science, but I see here the drawbacks, quite frankly in terms of the trust in science for the general public. If we publish very controversial scientific results very quickly, without putting them into context, then that could foster mistrust in science.

2.3. Impacts of artificial intelligence technologies on science, trust, and disinformation

Nóra Falyuna: AI technology is a subject that typically requires both scientific and technological expertise. As a leading researcher in the field of the socio-cultural impact of AI technology, Professor Feher, I have a question for you. During a recent

roundtable discussion, it was brought to our attention that new technologies challenge science (Feher et al. 2023). Specifically, how will AI technology affect society's perception of science and trust in information sources, and what does credibility mean in this context? Additionally, I would like to hear from others on the potential impact of AI technology developments on the content or logic of conspiracy theories expressing distrust, on disinformation or misinformation spreading, on the way science works, on its publication, and on research processes and norms. Moreover, how the market, policymakers, and science can communicate effectively with each other about the impact of AI technology.

Katalin Feher: The first point addresses the use of AI tools in scientific research. While there is a clear need to integrate more AI technologies for enhanced scientific inquiry, a significant barrier currently exists due to the commercial focus of these tools. Many AI applications ideally suited for scientific advancement are predominantly accessible to the industrial sector, and when available to the academic sphere, they are often gated behind expensive paywalls. This financial barrier discourages many universities from investing in advanced AI services, which thus challenges scientific progress. The second point delves into the potential of generative AI, particularly in enhancing the dissemination and visualisation of scientific work. Generative AI offers a promising avenue for scientists to communicate their findings more swiftly, clearly, and engagingly (Watermeyer et al. 2024). However, as highlighted in a recent comprehensive analysis published in *Nature* (Med 2023), there are both advantages and drawbacks to employing generative AI in scientific communication. Beyond the concerns of plagiarism and the automated generation of derivative content – which risks diluting originality in scientific discourse – generative AI presents an opportunity, especially for researchers who are not native English speakers, to efficiently translate and adapt their findings for broader dissemination. Yet, this innovation faces resistance from major publishing houses, which may view the democratisation of content creation and distribution as a threat to their traditional business models. This tension points to an upcoming challenge for the scientific community as it deals with the changes that may be brought by generative AI. Also, the financial aspect of these AI models is being questioned. A Business Journal article mentioned that the costs of developing and maintaining generative AI systems are higher than the money they make. This implies that there might be changes and adjustments in how we share and use knowledge and information in the future, as we figure out how to handle these new technologies.

Márton Demeter: When discussing whether AI poses a challenge for science, I find it confusing because it can indeed pose a challenge for individual scientists, but it is unclear in what way. AI can efficiently collect data, perform logical analysis, and interpret results based on that data. Logically, these activities are integral to science itself, and they are not something foreign or distinct from science; they are parts of its beauty. For instance, I may analyse 100 papers in a month, but AI can analyse a million in a second. So, the problem lies in the potential harm or challenges that AI poses to individual scientists. We need to ponder on and discuss humanity's role in

science because if artificial intelligence can perform all tasks more efficiently and cost-effectively than we can, then we must redefine our role in the entire system. For instance, in one presentation, the author discussed scenarios where Chat GPT writes a theoretical framework for a paper. However, one of the crucial features of science is that the results do not depend on the author. This underlies the logic of peer review, which is why it should be anonymised. The identity of the author is not interesting in a scientific context. What matters is whether the paper is methodologically sound, if the interpretation aligns with the evidence, etc. The author's personal details, such as income or ego, do not pertain to scientific questions; they belong to a different realm, such as business and promotion, which are part of a distinct social reality.

2.4. Trust and credibility issues around the relationship between science and business

Nóra Falyuna: Technological developments are increasingly shaping everyday life, and scientific research supports this. This impact extends to the relationship between science and business, or industry, and how science works. For example, with the industrial–technological support of science, one of Merton's classic norms, that science is disinterested (Merton 1973 [1942]), is not unconditionally applied. We must consider how the connections between science and industry or business actually affect the working of science and the process of knowledge production, as well as the trust in science. Furthermore, do these links manifest themselves in science conspiracy theories or in anti-science beliefs (e.g. anti-vaccine and anti-5G beliefs)?

Joseph E. Uscinski: I have been asking ChatGPT to develop conspiracy theories, and it is pretty good at it because it makes ones that are far more creative than those most of the conspiracy theorists tend to come up with. Let me flip the question though. I think if you look back in time, you will find that there has been fear about every new technology and method of communication. If you go back 100 years, people were saying, 'Oh my God, people are reading too many novels', and then it was 'people are reading too many newspapers', 'people are listening to the radio too much and it is going to bring about the end of society because their radios are full of bad information'. And then it was 'oh, TV is destroying our youth'. And you could keep going, and now it is 'oh, it's the Internet, and social media is destroying everything'. I read a headline just the other day saying ChatGPT girlfriends are now destroying an entire generation of men. Like, really, where are all these men that have been destroyed by AI girlfriends? At the same time, there are people on X saying that everyone has to stop doing AI research now, and if anyone is doing it, we should immediately scramble jets to go blow up the computers with missiles. I mean, some of the fear-mongering about this has been insane, and it is just a continuation of every other thing that people were afraid of it and said, 'Oh, it is the new thing, and this is what is going to end it all'. I guess what you said about our new role in science is a little bit unsettling, which I think we will all have to come to grips with. Once we are able to tell AI, 'Hey,

here is this massive data set, go write some papers'. And then you have them, then what am I supposed to do for work?

Márton Demeter: I think what scientists need is a better understanding of how the media operates. Currently, I reflect on a century-long journey with all the knowledge presented through the media. However, scientists, including social scientists, are not aware of this. They are not learning how to communicate within the logic of mass media because it is fundamentally different from scientific logic. When we as scientists attempt to navigate social media or any other form of media, we often do not grasp its logic, leading us to fall behind superstars or influencers who understand media logic better. My question is whether we should change our logic or, once we enter the media realm, adapt to its logic. This seems somewhat generational; for younger generations, it is evident that to communicate effectively in the media or with policymakers, each has its own unique logic. I believe that is something the scientific community should address alongside redefining the role of scientists. Of course, funding agencies can contribute; many grant applications now require you to consider how you will present your results to the general public, or write reports for policymakers. With funding, anything is possible. Either the scientific community initiates discussions on the role of scientists in modern societies or, following a more capitalist approach, funding agencies will make it mandatory. Researchers might need to align their research with mass media logic, engaging in activities like writing blogs, participating in social media, appearing on television, and communicating with the general public. They would also need to understand the logic of politics and communicate their findings to policymakers. If funding becomes contingent on these activities, researchers may find it more effective than relying solely on philosophical discussions.

Gábor Szüdi: One journalist on the VERITY project said that we are in a transitional period of media. I also think it is the end of the traditional media as we have known it for a hundred years. You can try to fight against it, but it does not make sense. The question arising from all these new projects about trust in science is how to regulate social media. I am an economist dealing with political sciences, so I am more for nudging people. However, actually, in some cases, you have to regulate. Otherwise, I think that nudging is a better way to make people communicate with each other. The projects funded by the Commission have clear guidelines on open science and the obligation to do it, which will have spill-over effects in that other people will also practice open science. We need to initiate changes by bringing people together to talk with each other, and find out what their problem is and what we can do to find ways to change communication because social media will not go away. We were talking about artificial intelligence. That is something which is changing rapidly and developing into something new. When we started the VERITY project, there was much less attention paid to it, and now we see that everybody wants to be engaged in it. So, these things will not go away, and I think it would be better that something comes from the bottom up than top down to make social media more human-centred.

3. Conclusions of the speakers

Nóra Falyuna: To close the discussion, let us summarise with a message: What are the biggest challenges and opportunities for an effective cooperation model in today's complex, multi-stakeholder information networks and information societies?

Joseph E. Uscinski: Looking back at my youth, I started out really liking science. Then I got into politics, and then I started denying lots of science, and now, I would not say I like politics, even though I am stuck in political science. However, now I like science again, as I pay less attention to politicians. There is an interesting lesson in that. Most people do not listen to scientists because they do not have direct access to them. They listen to politicians, and when politicians talk about science, oftentimes they get it wrong. But they have a megaphone that is 100 times bigger than the megaphones scientists have. Which is why there is climate denial. It is not that the people who deny climate change were dropped on their heads as kids, or that there is something wrong with them. It is just that they are listening to the person with the megaphone, and what that person says makes sense to them because they have been hearing it for 30 years. It is just common sense for them. I would hate to go down the road of scientists becoming a bunch of YouTube influencers, where we have the responsibility to do the study, doing our best to find truth, and then have to build a YouTube audience. But maybe that is where we have to go because we have to develop an audience for ourselves that is not just ourselves and each other. If we want people to trust us and like us and believe us, not following every scientific finding that comes out but at least the things where there is some consensus, then we have to do a better job of being opinion leaders, and as much as that frightens me, leaving it to politicians is not necessarily a great alternative either.

Katalin Feher: We would need at least one semester, if not a whole master's programme, to teach this subject properly, highlighting the importance of education. This is not limited to university-level learning but extends to younger generations, including those who started their schooling during the pandemic, who are often called 'COVID kids'. Young children and older students who had to navigate their first jobs remotely, an unprecedented beginning that will have influenced their grasp of scientific communication and technology. The second aspect of my response focuses on AI technology, a central theme in my research. The question is how can we leverage technology across the board – not only for future generations but also for our contemporaries and seniors. A simple example comes from a friend in the data science industry who appreciates the persuasive power of technology, who shared how a straightforward online interaction convinced him of the effectiveness of AI. This story illustrates the broader conversation about how we, as scientists or industry professionals, can engage with technology and its applications. The real question is about the value of using technology effectively and meaningfully. If we manage to do so, it could enhance scientific communication and potentially ensure scientific institutions' survival and relevance in an evolving landscape.

Gábor Szüdi: I think what is coming out of the VERITY project is that scientists, journalists, industry representatives, and policymakers need to work together to address mistrust in science. That is very difficult because we will all have to leave our safe space, otherwise things will not change. So, we will surely need to get used to some resistance to change. However, I think the industry people and policymakers should be more open to each other and towards average people and try to cooperate more because technology is really important, but I think humans are even more important in this equation. Being more open and cooperative goes a long way, but I hope this will happen more often in the future.

Márton Demeter: I think the most important thing is to understand human nature better. It might sound philosophical, but it is essential because we need to stay connected with people. Ignoring this can lead to bad news if we lose touch with what they want, need, and like. This holds true not only for scientists but also for politicians. Another perspective is that humanity is not synonymous with science. Science is artificial, so our mission is to find out if there is something that remains human in science in the future or if we need to change professions and pursue something more aligned with the nature of the majority of the population. As scientists, we have to be realistic.

4. Afterword

Due to the sophistication of the phenomena of deception and manipulation, their potential effects, and the development of technology, the literature describing and investigating the complex phenomena of information disorder and information manipulation is extensive. The panel discussion at the 'Science and disinformation: how science can support society against disinformation' conference, which focused on the complex relationship between science, society, and disinformation, highlighted the importance of cooperation between science and society and between different disciplines, as well as the crucial importance of effective and credible communication. The discussion identified interdisciplinary theoretical and practical considerations that could promote the development of science communication research in the current academic, social, and technological discourse. These include, for example, cooperation between decision-makers and scientists (see e.g. Kang and Liang 2023), or between scientists and journalists (Pereira and de Oliveira 2024) for knowledge sharing to aid dealing with the impact of mis- or disinformation; understanding the role and the nature of the human psyche, information reception in the response to mis- or disinformation, and the effectiveness of science communication (Huang and Wang 2024); trust as a key concept for information security (see e.g. Bak and Kelemen-Erdős 2023); and information literacy (Rab and Török 2020), and thus for information credibility and science communication (Yang and Yang 2024; Lammers et al. 2024). Another important topic in the current academic, social, and technological discourse, which was reflected in the panel discussion on science communication and disinformation, is the socio-cultural impact of artificial intelligence

(see e.g. Héder 2021); its role in shaping mis- or disinformation, conspiracy theories, or distrust in science (Većkalov et al. 2023); and its possible relation with science communication (Henke 2024; Schäfer 2023). The currency of this topic is shown by the fact that the *Journal of Science Communication* dedicated a special issue to this topic, based on the Annual Conference of the "Science Communication" Division of the German Communication Association (DGPuK) at the University of Zurich on 6–7 June 2024.

References

Bak, Gerda, and Anikó Kelemen-Erdős. "Stressz, opportunizmus és bizalom a szervezeti információs és kommunikációs technológiabiztonság tükrében." *Információs Társadalom* 23, no. 3 (2023): 9–26. https://dx.doi.org/10.22503/inftars.XXIII.2023.3.1

Falyuna, Nóra. "Science disinformation as a security threat and the role of science communication in the disinformation society." *Scientia et Securitas* 3, No. 1 (2022): 69–78. https://doi.org/10.1556/112.2022.00086

Feher, Katalin, Nóra Falyuna, György Huszics, Viktor Lázár, László Papp, and Bence Ságvári. "MI-kihívások a média, kommunikáció és információs iparágakban." Kerekasztal-beszélgetés. *Egyesület a Marketing Oktatásért* és *Kutatásért*, September 21, 2023. https://emok.hu/en/news/latest/643-report-on-the-2023-workshop-ai-challenges-in-the-media-communication-and-information-industries

Feher, Katalin, and Zoltan Veres. "Trends, risks and potential cooperations in the AI development market: expectations of the Hungarian investors and developers in an international context." *International Journal of Sociology and Social Policy* 43, no. 1/2 (2023): 107–125. https://doi.org/10.1108/IJSSP-08-2021-0205

Feher, Katalin, Lilla Vicsek, and Mark Deuze. "Modeling AI Trust for 2050: perspectives from media and info-communication experts." *AI & Society* (2024). https://doi.org/10.1007/s00146-023-01827-6

Fraser, Nicholar, Liam Brierley, Gautam Dey, Jessica K. Polka, Máté Pálfy, Federico Nanni, and Jonathon Alexis Coates. "The Evolving Role of Preprints in the Dissemination of COVID-19 Research and Their Impact on the Science Communication Landscape." *PLOS Biology* 19, no. 4 (2021): e3000959. https://doi.org/10.1371/journal.pbio.3000959

Harangozó, Éva. "How can science help in managing disinformation?" Accessed 11 February, 2024. https://en.uni-nke.hu/hirek/2023/10/25/how-can-science-help-in-managing-disinformation

Héder, Mihály. "AI and the resurrection of Technological Determinism." *Információs Társadalom* 21, no. 2 (2021): 119–130. https://dx.doi.org/10.22503/inftars.XXI.2021.2.8

Henke, Justus. "Navigating the AI era: university communication strategies and perspectives on generative AI tools." *JCOM* 23, no. 3 (2024): A05.
https://doi.org/10.22323/2.23030205

European Union's Horizon 2020 Research and Innovation Programme. "TRESCA." Accessed 11 February, 2024.
https://trescaproject.eu/

European Union's Horizon Europe Project. "VERITY". Accessed 11 February, 2024.
https://www.verityproject.eu/

Huang, Yan, and Weirui Wang. "Overcoming Confirmation Bias in Misinformation Correction: Effects of Processing Motive and Jargon on Climate Change Policy Support." *Science Communication* 46, no. 3 (2024): 305–331.
https://doi.org/10.1177/10755470241229452

Islam, Md. Saiful, Tonmoy Sarkar, Sazzad Hossain Khan, Abu-Hena Mostofa Kamal, Hasan S. M. Murshid, Alamgir Kabir, Dalia Yeasmin, Mohammad Ariful Islam, Kamal Ibne Amin Chowdhury, Kazi Selim Anwar, Abrar Aahmad Chughtai, and Holly Seale. "COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis." *The American Journal of Tropical Medicine and Hygiene* 103, no. 4 (2020): 1621–1629.
https://doi.org/10.4269/ajtmh.20-0812

Journal of Science Communication. "Announcing a special issue of JCOM on Science Communication in the Age of Artificial Intelligence." Accessed 4 June, 2024.
https://jcom.sissa.it/news/13/

Kang, Jen-Chun, and Chaoyun Liang. "Determinants of the backfire effect: Discrepancies between two groups with different political orientations." *Információs Társadalom* 23, no. 4 (2023): 38–57.
https://dx.doi.org/10.22503/inftars.XXIII.2023.4.3

Klofstad, Casey, and Joseph Uscinski. "Expert opinions and negative externalities do not decrease support for anti-price gouging policies." *Research & Politics* 10, no. 3 (2023).
https://doi.org/10.1177/20531680231194805

Koerber, Amy. "Is It Fake News or Is It Open Science? Science Communication in the COVID-19 Pandemic." *Journal of Business and Technical Communication* 35, no. 1 (2021): 22–27.
https://doi.org/10.1177/1050651920958506

Lammers, Wouter, Sacha Ferrari, Sylvia Wenmackers, Valérie Pattyn, and Steven Van de Walle. "Theories of Uncertainty Communication: An Interdisciplinary Literature Review." *Science Communication* 46, no. 3 (2024): 332–365.
https://doi.org/10.1177/10755470241231290

Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Stecen A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittran. "The science of fake news." *Science* 359, No. 6380 (2018): 1094–1096.
www.doi.10.1126/science.aao2998

Med, Tem. "Why Nature will not allow the use of generative AI in images and videos." *Nature* 618 (2023).
https://doi.org/10.1038/d41586-023-01546-4

Merton, Robert K. “The Normative Structure of Science.” In *The Sociology of Science: Theoretical and Empirical Investigations,* edited by Robert K. Merton, 167–278. Chicago: University of Chicago Press, (1973) [1942].

NASEM = National Academies of Sciences, Engineering, and Medicine. *Communicating science effectively: A research agenda*. Washington, DC: The National Academies Press, 2017. https://doi.org/10.17226/23674

European Commission. “Open Research Europe.” Accessed 20 February, 2024. https://open-research-europe.ec.europa.eu/

Pereira, Fábio Henrique, and Raphael Sandes de Oliveira. “Journalists and scientists together: the public problem of science disinformation in Brazil.” *JCOM* 23, no. 3 (2024): A04. https://doi.org/10.22323/2.23030204

Programme for the conference: “Science and Disinformation: How can Science support Society against Disinformation?” Accessed 11 February, 2024. https://ludevent.uni-nke.hu/event/3330/

Rab Árpád, and Bernát Török. “Online bizalom a magyar társadalomban.” *Információs Társadalom* 20, no. 3 (2020): 92–98. https://dx.doi.org/10.22503/inftars.XX.2020.3.6

Schäfer, Mike S. “The Notorious GPT: science communication in the age of artificial intelligence.” *JCOM* 22, no. 2 (2023): Y02. https://doi.org/10.22323/2.22020402

Tandoc, Edson C. “The facts of fake news: A research review.” *Sociology Compass* (2019): 13:e12724. https://doi.org/10.1111/soc4.12724

Većkalov, Bojana, Aart van Stekelenburg, Frenk van Harreveld, and Bastiaan T. Rutjens. “Who Is Skeptical About Scientific Innovation? Examining Worldview Predictors of Artificial Intelligence, Nanotechnology, and Human Gene Editing Attitudes.” *Science Communication* 45, no. 3 (2023): 337–366. https://doi.org/10.1177/10755470231184203

Watermeyer, Richard, Donna Lanclos, and Lawrie Phipps. “Does generative AI help academics to do more or less?” *Nature* 625, no. 7995 (2024): 450. http://doi:10.1038/d41586-024-00115-7

Yang, Zheng, and Tao Yang. “Differentiated Trust Strategies and Rebellious Acceptance: A Qualitative Comparative Analysis of the Trust Strategies Used by Scientist Communicators and Citizen Science Communicators in Chinese Online Climate Communication.” *Science Communication* 46, no. 3 (2024): 247–275. https://doi.org/10.1177/10755470241230487

The coherent emergentist concept of machines; or why the popular concept of artificial intelligence is a materialist anthropomorphism

The concept of artificial intelligence is very popular in both science and culture today. Similarly, the concept of emergence has become quite popular during the last decades in the sciences. For example, it is commonplace in the case of machines to speak of an overall blueprint and several different material components; thus, we can regard the blueprint as a kind of comprehensive emergent additive. However, is it true then that the machine, due to this plus component, is not material? Practically nobody wants to acknowledge that. Still, in practice, there are no machines without added blueprints. In my paper, based on Samuel Alexander's original concept of emergence, I will investigate these problems and contradictions, which stem from the materialist interpretation of the concept, and I will present a coherent emergentist concept of machines, according to which machines are clearly a unique kind between simple material things and living beings.

Keywords: *artificial intelligence, emergentism, materialism, extended mind, Samuel Alexander, Michael Polanyi*

Author Information

Daniel Paksi, Department of Philosophy and History of Science, Faculty of Economics and Social Sciences, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111, Budapest, Hungary
www.filozofia.bme.hu/people/daniel.paksi

How to cite this article:

Paksi, Daniel. "The coherent emergentist concept of machines; or why the popular concept of artificial intelligence is a materialist anthropomorphism".
Információs Társadalom XXIV, no. 2 (2024): 85–100.
https://dx.doi.org/10.22503/inftars.XXIV.2024.2.5

1. Preface: Machines and emergence

The concept of artificial intelligence, referring to a unique feature of computers, is very popular in both science and culture today, especially thanks to the rapid rise and fame of ChatGPT. The hopes and fears associated with it are beyond all imagination.

The concept of emergence is not so famous, but in the scientific world, because of its practical usefulness, it has also become quite popular during the last decades. Countless papers and authors speak about some kind of emergence and try to deal with the problems that arise from it. Similarly, the notion of emergence can be found in almost every field of science, from informatics via biology to physics, to describe some strange, complex features of nature—or machines.

But what is a so-called emergent feature, because of which we sometimes use this concept to grasp the point of certain entities? For example, it is quite commonplace in the case of machines to speak of a *blueprint* and several different *material parts*; thus, we can regard the blueprint as a kind of comprehensive *emergent additive* to create the *specific structure* of the machine. Similarly, which is the same problem just at a higher level, it is quite commonplace in the case of certain machines, usually called computers, to speak of *software* and *hardware*. Hardware is a complex system of simpler machines which are the *material body* or, more exactly, the material components of the computer, and thus we can regard software as a kind of comprehensive *emergent additive* needed for this material body to function.

However, is it true then that the machine or the computer, due to this plus component, is not material? Practically nobody wants to acknowledge that. But, *in practice*, it is clear that there are no machines without added blueprints and comprehensive structure, and there are no computers without added software which guides them as a navigator guides a ship by his senses and experience.

So, what is the problem, what is the reason for this huge conceptual difference between *theory* and *practice*? In my paper, based on Samuel Alexander's original concept of emergence, I will investigate these problems and contradictions, and I will argue that these problems stem from the materialist interpretation of the concept. We use it because it is really practical and helpful; however, we use it not in its original meaning but *in a materialist theoretical framework*.

Contrary to this, I will present a coherent emergentist concept of machines and emphasize some of the consequences. According to this genuine emergentist approach, machines are really a unique kind between simple material things like rocks and living beings like us; they are not just incapable piles of physical quarks but neither are they brilliant intellects of (living) bodiless minds. Regarding the concept of artificial intelligence, it means that computers are really capable of manipulating data, way, way better and faster than us; this is their knowledge, but still, they are not intelligent at all; this is our profound knowledge, to regard them as such is only a kind of modern materialist anthropomorphism.

2. The problem between theory and practice

We have seen that, for example, in the case of computers we can speak of software and hardware, and we can regard software as a kind of emergent additive to the more clearly material hardware, but then the question arises as to whether the computer, due to this plus component, is not material? So, what is the problem, what is the reason we see a huge conceptual difference between *theory* and *practice*?

This contradiction stems from the different tacit meanings of the term "material." First of all, in practice, based on an old Aristotelian tradition, we usually regard *every kind* of component or part as material which causes some confusion. For example, in the case of computers, we have just spoken about the software as a kind of emergent additive and about the hardware as the *material components* of the machine. However, in the case of simpler machines, where there is only hardware, we regard *this* level as the comprehensive emergent additive compared to the further material parts. So, in one case, the blueprint and the hardware are the *higher-level* comprehensive emergent entity, but in another they are simply part of the *lower-level* material one.

However, in a sense, this is a proper understanding of the phenomenon because it just shows that, contrary to our explicit idea, tacitly, we are already using the emergentist approach—which, by the way, is a neo-Aristotelian one. But, then, what is the real problem?

The main problem is that in our concept we still always define the term "material" *in contrast to* the Cartesian concept of mind or, in a wider, more cultural approach, to the Christian concept of soul, and we do not care about machines and computers and blueprints and software. And in this theoretical sense, since it is obviously not a Cartesian mind, the blueprint is, of course, rather material. But consequently, and this is the point, we will lose *the clear meaning of the difference* between the blueprint and the material components, between the software and the hardware, we otherwise see in practice. For a coherent emergentist concept of machines, to solve this contradiction between theory and practice, we have to define both the meaning of the comprehensive additive and the meaning of the material components *on genuine* emergentist grounds.

So, what is the genuine meaning of emergence? And where is the place of machines in this genuine understanding? To understand this, we first have to reflect on how the materialist reinterpretation of the concept works in science and philosophy today; otherwise, we will not be able to see the original meaning.

3. The concept of emergence: The evolution of the materialist understanding

The usual and unfortunately superficial approach to emergentism in science and philosophy today is that it is *half dualism* and *half materialism*. Let's explore these terms further.

1) Dualism—which is the ontology of our traditional Christian belief—claims that there are *two substances*, *mind* and *matter*—or, traditionally, soul and body—and, at the end, *everything* in the universe is composed of these *two*. This means that mind and matter can be separated, and the mind is regarded as the more important of the two because it is, of course, the "image" of the ultimate mental substance, that is, God. This is, of course, an *anthropomorphic* understanding of what reality really is, since it is based on the assumption that mind is an extraordinary entity over and above nature. For example, there is something in someone's body which can be downloaded and, then, uploaded into a different body, and he/she will still live happily afterwards without his/her original body. Although, in theory, science is materialist and has left behind these old "mystic" and "obscure" concepts of the "dark ages," many artificial intelligence specialists still wonder about these ideas. We, unfortunately, are not coherent and tend to be anthropomorphic.

2) Materialism, on the other hand, claims that there is only *one substance*, *matter*, and *everything* is composed of only this *one* material substance. However, we have already seen that what matter does mean is a really hard question. Is it wave? Particle? Quantum? Fields? Quantum fields? Strings? Dark matter? Dark energy? Molecules? Or just body? In one case, the hardware is a plus additive to the material components; in another, it is the system of material components itself.

In a positive sense, *there is no definite answer* to the question in the sciences. We do not know what these things really are, or, at the end, what matter really is; just think about the famous dark matter and dark energy that allegedly make up 95 percent of the universe. And, as we have seen, the different special sciences like engineering or sociology regard and treat very different things as matter in practice; for example, the quarks of physics, the DNA sequences of genetics, the respiratory systems of biology, the means of production of economics, the hardware of informatics, the environmental conditions of sociology, etc., have, in fact, very different qualities and are treated quite differently in their respective scientific fields.

Similarly, historically, the different so-called materialist scientists and philosophers claimed very different things about the nature of matter and the nature of the whole universe. For example, the ancient Greek atomists like Democritus and Epicurus are regarded as materialists because of their atomist physics; however, it is clear that for Epicurus man has free will and a moral life that is not determined by the motions of atoms; this dualistic ethics defines his philosophy, rather than atomist physics. The same is the case with the early modern so-called materialists like Thomas Hobbes and Pierre Gassendi, or Galileo Galilei and Isaac Newton; they are regarded as materialists because, in a sense, they argued against Cartesian dualism or against the old, so-called geocentric Aristotelian science, but otherwise, as is now fortunately very well known in the case of Newton, they were not materialists at all in the sense that everything in the universe is composed of only one material substance.

The term "materialism" became widespread in science and philosophy in the 18th century, especially due to the so-called French materialists like Julien la Mettrie, Paul-Henri d'Holbach, and Denis Diderot, and became popular and really influential at the beginning of the 20th century thanks to primarily Marx's so-called

dialectical materialism for obvious reasons. However, in the modern analytical and scientific sense, they were still not "real" materialists, especially the French; their position can rather be defined as, in fact, Marx himself did: materialism is the right historical and scientific *antithesis* of the wrong idealism (dualism) and religion of the exploitative ruling classes.

In contrast to this, modern *analytical materialism*, which became the ruling ontological concept in the middle of the 20th century in science and Anglo-Saxon philosophy, is defined by *reduction*, a strictly scientific method by which a deductive logical relationship can be established between the concepts and laws of the special sciences referring to higher-level phenomena like DNA sequences, hardware, etc. and fundamental physics referring to matter itself (see especially Nagel's *The Structure of Science* [1961, 345–359]). However, at the end of the 20th century it became clear that *in practice* reduction does not work and, thanks especially to the discoveries of dark matter and dark energy and to the still insoluble contradiction between our two fundamental theories (i.e., quantum mechanics and the Theory of Relativity), it is not even clear anymore what we should regard as matter in the scientific sense.

Thus, what has been left is what we have seen with the earlier materialists, especially in the cases of the French and Marx, namely that materialism can rather be defined in a *negative sense*: it is the very general ontological or metaphysical claim that denies the existence of minds, including God himself, but provides nothing specific concerning the existing only one substance. So, in reality, it is indeed the *modern* (so-called scientific) *antithesis* of our traditional dualist ontological belief—which thus, as we will see, in a disguised way, conserves the original anthropomorphism of dualism.

One contemporary example of this loose negative ontological tradition is the so-called New Materialism which, especially in the humanities and social sciences, tries to focus more on the different material (in fact, lower-level) conditions of social institutions and human agency toward the environment and emphasizes the social and technological entanglement of the human and the non-human, since both are the "organic" parts of the same material world. However, this tradition is not based on any scientific concept of matter. Moreover, usually, it is explicitly anti-reductionist, or even in line with a more traditional philosophical concept of matter or materialism, and the different authors, from Judith Butler via Jane Bennett to Gilles Deleuze, labeled by this term usually claim very different things about man, society, technology, and nature. So, New Materialism is clearly not a coherent positive ontological claim about man and machines as the original concept of emergentism is (see below).

3) Emergentism is half dualism because it claims that there are at least *two kinds of entities* in the universe, in this case mind and matter, of course; and it is half materialism because it claims that there is only one substance: matter, that is, matter is *fundamental compared to* mind. This means that, although there is mind, *there is no* emergent mind *without* substantial matter; for example, one cannot download it from its original body because it does not stand on its own, not even during the "short" time of downloading and uploading; it simply needs its original body to even exist.

So, at the surface, it is true that emergentism is half dualism and half materialism, but then we define and understand emergentism by Cartesian dualism and materialism. And since, as we can see now, in reality materialism is *also* defined *negatively by* Cartesian dualism, that is, by the negation of the existence of the mind as an independent substance, emergentism, ultimately, will *also be defined* by the Cartesian concept of mind and *also will conserve* the original anthropomorphism of the concept.

And yes, by the way, be honest; in fact, this is the question we are really interested in—just go up, for example, to the pages of the famous and influential *Stanford Encyclopedia of Philosophy*, and you will see that everything revolves around the question of mind—but this *anthropomorphic approach*, unfortunately, poisons our understanding of emergentism and, which is more important for us now, the understanding of emergent but mind-less machines.

In the case of the concept of emergence, this anthropomorphism and this negative definition come from C. D. Broad's *The Mind and Its Place in Nature* (1925). Broad is regarded as one of the great British Emergentists and it is clear that he is the most influential one. Broad, already in the early 20th century, even before the concept became widespread in the middle of the century, realized that the analytical reduction of special sciences does not work, or, more precisely, was not going to work. He defined the concept of emergence *analytically* as the logical *contrary of reduction* (i.e., emergent equals non-reductive) and claimed that, thus, the scientifically most grounded ontological position toward mind is *emergent materialism* (Broad 1925, 647). However, it is materialism *only in the broad sense of the antithesis* that there are no Cartesian minds because *there are* emergent minds, so, although there is only one substance, matter, still *not* everything is composed of only this one material substance.

At the end of the 20th century, after the realization that reduction in practice does not work, Anglo-Saxon analytical philosophy and science adopted Broad's terms but not his position; so, *this* is the materialist and now widespread understanding of the concept that has become quite popular during the last decades in almost every field of science. Today the most popular ontological position in philosophy and science is, thus, *non-reductive materialism*, which, at the explicit surface, that is, at the level of the term, is the same as Broad's emergent materialism; however, today it means that *everything* is composed of only one material substance—*in theory*, because *in practice* we simply accepted that there are terms and laws and concepts, etc. that refer to non-material phenomena. That is, in theory the software is material; however, in practice we treat it as something non-material, and this is the reason we sometimes call it emergent.

4. The concept of emergence: The original understanding

So, what is emergentism in reality? Why is it, in fact, not half dualism and half materialism but, on the contrary, that materialism is half dualism for obvious reasons and emergentism is *the real antithesis of both*?

From a genuine emergentist approach, regarding their points, *both* dualism and materialism claim *the same*, i.e., that in the universe, at the end, *everything is composed of* (one or two) *substances*. These are *old* ontological or metaphysical concepts, going back into the ancient Greek era of philosophy, which are, in fact, thinking in a *static* universe where the substantial, meaningful fundaments of reality do *not* change.

However, emergentism is a *new* concept, formulated by Samuel Alexander (1920) at the beginning of the 20th century, and the main point of it is that the universe in which we live is *always evolving*, so there are no static substances at the bottom of it. Reality is defined not by substances but by *the process itself* by which new and new levels of reality arise. This process was called *emergence* by Alexander—and in this strict sense it has nothing to do with the analytical concept of reduction.

According to Alexander, the evolving universe starts with an *infinite singularity* of space-time; the main emergent levels of reality to date are *space-time*, *matter*, *life*, and *mind*. Alexander, contrary to Broad, emphasizes that we should not try to understand reality based on our mind or by an analysis which revolves around the mind or matter because the fundament of reality is not mind or matter *but space-time*—man and mind are only the latest developments of reality (Alexander 1920, 8). Alexander's main goal is "to de-anthropomorphize: to order man and mind to their proper place among the world of finite things" (Alexander 1914, 279). This is the reason that he is a realist.

In short, there is no matter or man without space-time, but there is space-time without matter or man. This is the real difference between original emergentism and any kind of materialism, including now Broad, Einstein, and the New Materialists; ultimately, reality is defined by *emergent space-time relations* and not by material forces as, in fact, is the case, for example, in a specific structure of a machine described by the blueprint and not by any composite matter, part, or force (i.e., by a comprehensive plan for the specific space-time relations for these composite parts having material forces). The original coherent concept of emergence cannot be reconciled with any kind of dualism or materialism.

Nonetheless, the really important part of this for us now is that the *living body by evolution* is *not* material. Contrary to this, life is regarded by *both dualism* and *materialism* as only a kind of (complex) matter, which, according to the modern concept, is a kind of *mechanical body* (a specific machine); however, *in practice*, *nobody* believes anymore that animals are senseless machines like the famous locomotive of Stevenson (non-reductive materialism). Thus, according to Alexander, the mind has a very *peculiar relationship* with the living body contrary to the mainstream concepts.

As we have seen, dualism, in an anthropomorphic way, regards mind as the most important composite entity of the universe, while materialism denies even its existence, holding that there is only material body, in this case the brain. However, according to Alexander, any mind is the development of a *specific living body*, where there is a unique (especially neurological) space-time defined by that body and there are unique space-time relations (through experience) with other bodies and minds.

With an easy example, the walls of any room define a unique space-time which is not the same as space-time itself because, in fact, the earth rotates around its

axis and, at the same time, it orbits the sun at a crazy space; still, the space-time of the room (if we are not speaking about Einsteinian so-called relativistic speeds, of course) *preserves* its *unique* integrity. According to Alexander, these are important and definite facts regarding the nature of space-time and reality; however, mainstream science, according to its materialist conviction, simply disregards these facts. For example, for Einstein there are material points and *every kind* of space-time is only the relative mathematical property of those material points described universally by his equations; consequently, there are no unique space-times; as a matter of fact, in this way, space-time is not even an entity (Einstein 1920, 94–95). This is the reason that physicists, now, for a century, cannot accept that gravity is the process of space-time, which, otherwise, should follow from Einstein's theory, and want to describe it with much more reputable quantum (i.e., material) processes, that is, by so-called quantum gravity.

The point is that for Alexander, contrary to the traditional Cartesian concept, any mind has a concrete *extension* in both space and time, primarily or originally in its biological body but not just, and the human person him/herself is the *unique* and inseparable *compresense of* the mind and the body.

> The higher self [i.e., the mind] is thus in all its stages a continuation and expansion, and refinement of the bodily self. The body, it may be observed, is capable of indefinite extension. We feel the ground at the end of the stick we carry, not at the finger which holds the stick: the stick has become part of our body. … All these things may become extensions of my body and the experiences I get from them may be for a time of a class with my organic and other bodily sensations … Many or most of these extensions of the body are only possible to a life which has gone far beyond the interests of the body itself. (Alexander 1920, 105–106)

At the beginning of the 21st century, for pragmatic reasons, the concept of *extended mind* or extended cognition became well-known even in Anglo-Saxon analytical philosophy and cognitive science (see originally Clark and Chalmers 1998)—but, of course, in a problematic non-reductive way because, *in theory*, the mind does *not even exist*, it is just another name for the brain—which, by definition, *cannot* extend into, for example, a carried stick.

However, in Alexander's original theory, there are no such contradictions. The only reason the original living body is able to extend is because of the already extended comprehensive emergent mind in that body which, according to its higher interests, through sensation and experience, in the literal sense, extends into, for example, the carried stick. This means that the unique space-time of the mind, as we open a door to another room, flows into this new body, that is, into this new composite part of the person. Nonetheless, this process is, of course, only temporary because the center (its space-time singularity) of the mind, where it is anchored, is the original living body where it has been evolved.

I believe that the scientific discoveries of the last century rather prove Alexander's theory than refute it. For example, it is clear now that the universe starts with

an infinite singularity of space-time—which, by the way, also follows from Einstein's Theory of Relativity. Still, even Einstein himself famously thought that this could not be the case; the universe, according to his traditional ontological beliefs, had to be static. This is my main argument in this paper that scientific practice *is coherent with Alexander's original concept* and not with materialism, including the widespread Broadian non-reductive concept of emergence. Still, the reader must know that the vast majority of the scientific community and philosophers do not agree with me; they are, of course, still materialists or perhaps dualists and usually have not even heard about Alexander's original concept of emergence.

Nonetheless, now, our question is what the coherent emergentist concept of machines is, and how, in the light of this genuine understanding of emergence, we can understand what a machine really is.

5. What is a machine?

First of all, as, I hope, the reader can already suspect, there is no one definite answer or definition to this question because, as with the universe and reality itself, machines can evolve too; and they *did*—because they are emergent. Still, before the evolutionary details, I will provide a general starting picture reflecting on our theoretical problem, now, in the case of machines and engineering. And perhaps it is worth emphasizing that this coherent emergentist concept cannot be classified into the well-known categories of philosophy of technology because it is based on a unique and usually unknown positive *ontology*. For example, Andrew Feenberg distinguishes between four types of philosophy of technology: instrumentalism, determinism, critical theory, and substantivism (Feenberg 1999, 9). However, these are not ontological categories; thus, this coherent emergentist approach is a kind of instrumentalism since technology is, in a sense, a mere instrument of humans but, at the same time, also a kind of substantivism. This is because, as we will see in a moment, machines are existentially different compared to humans and, in a sense, autonomous to some extent. Similarly, although Martin Heidegger's influential view on technology is based on different ontology and has a significantly different approach on it, in a sense it still has very similar results, especially concerning the relationship of technology and society (Heidegger 1954).

So, machines can evolve too; however, on the other hand, they are not evolving by themselves, as life does, but were *created by man*, which makes an important and meaningful difference in their emergent place of reality. There was life before and without man, but there were no machines. Life and technology, evolution and programming are not the same things and processes. They are only the same in a dualist or materialist theoretical framework which *disregards* the definite importance of the *evolutionary origin* of things; see, for example, Turing's famous and influential so-called functional (i.e., non-reductive materialist) definition of artificial intelligence, which I will touch on at the end (Turing 1950; and perhaps see my detailed evolutionary interpretation of it in Paksi 2022).

So, perhaps the main consequence of our living in an evolving universe which creates a *hierarchical order* in reality is that there is *no one* scientific method, no one conceptual toolbox by which one can describe and define every phenomenon because, by new and new emergent levels, new and new *ordering principles* come into operation.

> Engineering and physics are two different sciences. Engineering includes the operational principles of machines and some knowledge of physics bearing on these principles. Physics and chemistry, on the other hand, include no knowledge of the operational principles of machines. Hence a complete physical and chemical topography of an object would not tell us whether it is a machine, and if so, how it works, and for what purpose. Physical and chemical investigations of a machine are meaningless, unless undertaken with a bearing on the previously established operational principles of the machine. (Polanyi 1967, 39)

Polanyi, an emergentist, here simply describes how engineering *in practice* works and acknowledges the existence of the unique principles of the field. He can easily do that because he is not a materialist. His conclusion is the following: "A physical and chemical investigation cannot convey the understanding of a machine as expressed by its operational principles. In fact, it can say nothing at all about the way the machine works or ought to work" (Polanyi 1962, 329). The same is true for biology and physics: "a complete physical and chemical topography of a frog would tell us nothing about it as a frog, unless we knew it previously *as a frog*" (Polanyi 1962, 342).

Biology is not physics, body is not matter, and engineering is not biology. These are *facts* about science. *In practice*, we regard and handle biology and engineering very differently compared to physics. The Standard Model and the Theory of Relativity do not—and, in reality, *cannot*—include any blueprint or operational principles of machines; they are *unique* structures of space-time, while physics tries to describe the properties and laws of matter in the *most universal* way. Nonetheless, *in theory*, according to our materialist vision, we, of course, still wonder about the unity of science and a final physical theory of everything—meaning that we suppose that *everything* is composed of the only one material substance physics can describe.

So, a machine, *in its origin*, is but an *achievement* of an *invention by man*. It is a comprehensive, rational *tool* that we use during an act for the sake of some *goal* to reach some benefit. The use and operation of a tool or a machine is based on comprehensive *rules of act* (in this case, *operational principles*) that cannot be determined by their details and which allow the successful usage of a tool or a machine. The machine is, thus, defined by its *patent* that tries to describe the machine in the broadest possible sense; it does not include the concrete realization and the different possible material conditions of the machine, but only defines the specific structure and the operational principles, due to which the machine can properly function and fulfill its goal (produce its benefit).

Therefore, these two forms of knowledge—practical engineering sciences and theoretical natural sciences—are *highly different*. *Both* are relevant regarding ma-

chines, but while engineering refers to the comprehensive emergent whole (i.e., the specific space-time structure) and the operational principles (i.e., the practical knowledge of its successful functioning and control), the *physical sciences* refer only to the *tangible parts* and *material conditions* of the machine.

The theoretical problem one can see here comes from our modern ontological convictions. Aristotle was right when he differentiated between *techne* and *episteme*, that is, in our terms, between practical engineering sciences and theoretical natural sciences, on the ground that tools and machines are the particular and teleological (purposeful) creation of the human mind. However, this differentiation was eliminated by modern dualism because nature *too* became the particular and purposeful *creation* of the ultimate mind—God—but worse, at the same time, thanks to the simple dualist concept of reality (that is, at the end, everything is mind or matter), nature itself, and thus the living body became a complex *material* machine—since it is obviously not a Cartesian mind. Therefore, without the ultimate mind, that is, according to the now ruling antithesis, there can be no real (purposeful and meaningful) difference between life and machines and between machines and matter; the hierarchical structure of reality, we see everywhere in practice, has utterly collapsed, and engineering has become the simple practical application of physics.

Aristotle was right because he differentiated between *techne* and *episteme on the ground of concrete practice*; however, his theoretical explanation of (static) forms about the hierarchical nature of reality was rightly discredited by modernity—creating our problem of practice vs. theory. Emergentism is neo-Aristotelian in the practical sense, that is, it acknowledges the hierarchical nature of reality we see everywhere, but it explains these different levels by the new (dynamic) concept of emergence.

So, then, what is a machine, according to a coherent emergentist concept? In other words, what are the emergent *origin* and the different emergent levels (types) of machines as they have been created?

6. The coherent emergentist concept of machines

First of all, there was life. And then, on a certain higher level, clever animals that had higher interests outside of their original living body realized that they could use certain non-living things from nature to reach their goals. Their reasoning was that in certain circumstances their natural *bodily tools* could not do the job. A stick with a chewed end by which a chimpanzee can reach termites through a ventilation opening is a tool which *literally* is the *extension* of his arm, and works accordingly.

I deliberately use the words “clever” and “reasoning,” etc., because in a coherent emergentist approach animals are neither physical nor machines and higher animals clearly have neurological systems, minds, and primitive cultures according to the gradual evolution of these phenomena; so, to call an animal intelligent or to

speak about its reasoning is not an anthropomorphization of physical or lifeless phenomena but a based positive philosophical position.

1) So, what is a tool? A tool is an *emergent object created*, *operated*, and *controlled* by man. As we saw in Section 4, this *direct control* of man is described even in non-emergentist literature as a kind of *extended mind*. The tool is emergent because it has a specific, functional, higher-level, comprehensive structure defined by its purpose. It is a *mechanical body* which, by the control and extension of the mind, due to certain rules of act, complements our biological body. And it is emergent because it necessarily has some lower-level, material fundaments without which it cannot exist and function.

The three-level structure of an operation of a tool
1) *direct* comprehensive *control* of man over both his body and the tool a kind of extended emergent mind
2) *biological body* + *mechanical body* a kind of technical extension of the body
3) *material parts* and *conditions* in both bodies

Table 1. The three-level structure of an operation of a tool (own editing)

2) What is a machine? All bugs are insects, but not all insects are bugs, as is well-known. However, similarly to biological evolution, the emergent structure of reality develops in the case of technology too, that is, all machines are tools, but not all tools are machines. What is the main difference?

The difference, on the one hand, is a higher level *blueprint*, according to which different lower-level, *tool-like parts* are arranged into a functional whole. For example, springs and gears and pointers are arranged into a clock. And, on the other hand, there is usually *a built-in source of power* which, due to certain operational principles, can work the machine instead of the muscle-power of man. In the case of a clock, it is originally a coiled spring. This means that at a higher operational level the emergent tools as lower-level parts become the necessary material conditions of the emergent machine. In a genuine emergentist approach, as in our traditional (i.e., Aristotelian) use of language, the material always refers to *the actual lower-level conditions* and not to some undefined substance.

To make it clear, the higher emergent level of the machine is created from *mechanical tool-like parts* and *not* from a living biological body; thus, it *has no mind*, it has no neurological system, it is *not part* of biological evolution. The meaning of the blueprint is only the *knowledge of man*. For example, the clock does not know how it was created, what the text of its patent was, or what the time is, exactly the way the stick does not know what a termite is despite the fact that it helps catch it; the clock can only function due to its unique comprehensive structure, which means that its pointers are moving, and a living human person, if he/she knows how, can read it. This will be particularly important in a moment.

So, a machine is an independent (i.e., contrary to tools, due to its operational principles, can function in itself) *mechanical body* which is usually controlled only partly and indirectly by man.

The three-level structure of an operation of a machine
1) the comprehensive emergent machine
2) emergent tool-like parts
3) further material parts and conditions

Table 2. The three-level structure of an operation of a machine (own editing)

3) What is a computer? Finally, the great invention of the 20th century that is the computer is a kind of machine where the original machine or, more exactly, a system of machines *as the hardware* becomes *the lower-level material condition* of *a higher-level, emergent, functional whole* operated *by the software*.

The main difference between simple machines and computers is that computers have *sensors* and *data*; usually this is their main purpose—not to just do some kind of physical work, for example rotating pointers, but *to manipulate data for us*, way, way faster than we could do using only our minds. The operational principles of a simple machine have to be *incorporated into* the concrete structure of the machine which, as a matter of fact, could be very complex as is the case, for example, in the case of a traditional pinball machine. However, in case of computers, the whole point of the software is that there is a higher-level *operational center* in the machine where the operational principles are *electronically coded*; this opens up enormous opportunities for the functioning of this kind of machine.

So, the software crates a partially independent *control mechanism* over a complex system of machines, and this is the reason we should regard the computer as a new higher-level emergent entity. However, it *still has no mind at all*; it is just an always *physically* (i.e., electronically) *realized mechanical control system*, still "just" a new comprehensive mechanical structure (body) of lower-level machines (bodies) as hardware. The computer, as the stick or the clock, still *has no* living body; it still *has no* neurological system; it is still *not part of* biological and cultural evolution; and, after all, it is still *controlled by man*.

The four-level structure of an operating computer
1) the comprehensive emergent computer
2) emergent machines as parts
3) emergent tool-like parts
4) further material parts and conditions

Table 3. The four-level structure of an operating computer (own editing)

So, a computer is an independent *mechanical body* which, on the one hand, can manipulate data independently and, on the other, can control other machines. That is, contrary to simple machines and due to our new technological knowledge of *programming*, it is always controlled only partially and indirectly by man; of course, even in this way, it is *still* always controlled by man.

Therefore, the meaning of the programming as the meaning of any data of the computer is only the *knowledge of man*, as, in fact, one can represent and read data even by the simplest tools like a paper and a pencil, and as, in fact, it was already the real function of the pointers of the clock. In the ontological sense, *there is no difference* between an always physically realized electric mechanism and the pointers: *both are mechanical bodies*; similarly, an electromagnetic condition in the memory of a computer, in the ontological sense, is *exactly the same* as any ink on a paper. The difference is only *epistemic* and *practical* in relation to how that material condition was created, and which data represents something meaningful for us, if we know how to read it, of course. The ink, because we are speaking about simple tools, had to be manually and directly created by man; contrary to this, the whole purpose of the computer is to do it more efficiently, that is, necessarily indirectly by partially independent control-mechanisms, that is, by programming. Thus, as we would never think that the paper knows the meaning of the ink (in this sense, of the text), similarly, we should never have to think that the computer knows the meaning of any material conditions (in this sense, data) in its memory.

However, because of our ruling non-reductive negative materialist approach, *in our theoretical reflections* we are not able to differentiate between the independent comprehensive control mechanisms of computers and the more direct mechanical functioning of simple machines or even simplest tools; as we, in fact, are not able to differentiate between biological processes and evolution and mechanical processes and technology, not to mention the difference between the cultural knowledge of minds and the electromagnetic discharges in our brains or the lower-level "material" structures and conditions of cultural institutions and phenomena in the case of social sciences. In the absence of a positive ontology, we can only state that none of these are Cartesian minds or Christian souls; therefore, they all have to be the same, that is, material. These are clearly different processes only *in scientific practice*.

7. Conclusion: Why is the popular concept of artificial intelligence a materialist anthropomorphism?

Simple tools were first created to extend the abilities of our body; but, then, there were also clay tablets and paper. As a matter of fact, from an emergentist approach, this representational function of simple tools which did not just extend our minds for a concrete purpose, as a stick does, but, at the same time, created new intellectual abilities was their more important part in cultural evolution (see, e.g., Donald 1991, 269–360, or my interpretation in Paksi 2019, 151–177).

Machines, by their own power source and more comprehensive and more functional structure, became independent from our bodies and created new possibilities to represent and manipulate data—just think about Gutenberg's printing press.

However, finally, computers were primarily created for the purpose *of really extending the abilities of our minds* by their representation and manipulation of data. And this is fascinating—we are still dazzled like a newborn baby.

But, if this emergentist approach is right, then machines have no minds based on living biological and psychological and cultural conditions, as we have; therefore, only we can understand anything about the meaning of any software, data, or sign, because each is still *an extension of our minds*.

Evolution is not a one-way process. Birds are not mammals and insects are not vertebrates—as machines are not animals like us. But both birds and mammals are vertebrates—as technology is a development of human cultural evolution.

The three levels of evolution and the three sublevels of technology
1) comprehensive control of man a kind of *extended emergent* mind
2) biological body + mechanical body a kind of *technical extension* of the body → 2a) tools → 2b) machines → 2c) computers
3) material parts and conditions in both bodies

Table 4. The three levels of evolution and the three sub-levels of technology (own editing)

The Turingian concept of artificial intelligence is a non-reductive materialist one. It is based 1) on rejection of the existence of the extension-less Cartesian mind which, according to dualism, is, of course, the only source of meaning and intelligence (materialist antithesis); and 2) on the spectacular functioning of computers (non-reductive-ness), meaning that the evolutionary origin of man, that is, the significance of the living, biological body and the surrounding culture, is simply neglected.

A more consistent reductive materialist approach should claim that there is no intelligence at all, since there are no Cartesian minds, the only source of meaning and intelligence. But without any positive answer as to what intelligence really is, this Turingian approach just opens the door to regarding independent and fast data-manipulating computers as intelligent *only because* they, in a superficial sense, *seem to be* intelligent. This, instead of a principled positive answer, is, in fact, the same kind of *anthropomorphization* of fascinating phenomena as we did when we had only simple tools like hand-axes, and regarded *every fascinating moving object* from waters through bushes ruffled by the wind to the moon and the sun *as animated by some kind of spirit*—that is, by intelligent minds.

References

Alexander, Samuel. “The Basis of Realism.” *Proceedings of the British Academy* 5 (1914): 279–314.

Alexander, Samuel. *Space, Time, and Deity*. London: MacMillan and Co., 1920.

Broad, Charlie Dunbar. *The Mind and Its Place in Nature*. New York: Routledge, 1925.

Clark, Andy, and David Chalmers. “The Extended Mind.” *Analysis* 58, no. 1 (1998.): 7–19.

Donald, Merlin. *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition.* Boston, MA: Harvard University Press, 1991.

Einstein, Albert. *Relativity: The Special and the General Theory – A Popular Exposition.* London: Methuen and Co., 1920.

Feenberg, Andrew. *Questioning Technology*. London: Routledge, 1999.

Heidegger, Martin. “Die Frage nach der Technik.“ In *Vorträge und Aufsätze*, 5–36. Pfullingen: Günther Neske, 1954.

Nagel, Ernest. *The Structure of Science: Problems in the Logic of Scientific Explanation.* London: Routledge and Kegan Paul, 1961.

Paksi, Daniel. *Personal Reality: The Emergentist Concept of Science, Evolution, and Culture*. Volume 2. Eugene, OR: Pickwick Publications, 2019.

Paksi, Daniel. “Technological Singularity by Culture; or the So Popular Concept of the Rise of the Machines Which Will Never Come.” *Információs Társadalom* 22, no. 4 (2022): 86–94.

Polanyi, Michael. *Personal Knowledge*. London: Routledge and Kegan Paul, 1962.

Polanyi, Michael. *The Tacit Dimension*. London: Routledge and Kegan Paul, 1967.

Turing, Alan. “Computing Machinery and Intelligence.” *Mind* 59, no. 236 (1950): 433–460.

Technophilosophical underpinnings of *Westworld* (2016–2022)

Using the events of the HBO series *Westworld* (2016–2022) as a springboard, this paper attempts to elicit a number of philosophical arguments, dilemmas, and questions concerning technology and artificial intelligence (AI). The paper is intended to encourage readers to learn more about intriguing technophilosophical debates. The first section discusses the dispute between memory and consciousness in the context of an artificially intelligent robot. The second section delves into the issues of reality and morality for humans and AI. The final segment speculates on the potential of a social interaction between sentient AI and humans. The narrative of the show serves as a glue that binds together the various ideas that are covered during the show, which in turn makes the philosophical discussions more intriguing.

Keywords: *sentient AI, Westworld, consciousness, immortality, reality*

Author Information

Pankaj Singh, University of Petroleum and Energy Studies (UPES), Dehradun, India, School of Liberal Studies (SoLS)
https://orcid.org/0000-0002-6570-3634

How to cite this article:
Singh, Pankaj. "Technophilosophical underpinnings of Westworld (2016–2022)". *Információs Társadalom* XXIV, no. 2 (2024): 101–114.
https://dx.doi.org/10.22503/inftars.XXIV.2024.2.6

1. Introduction

American television series *Westworld* (2016–2022) contemplates various philosophical issues, concerns, and possibilities between humans and artificial intelligence (AI). The series ended after four seasons. Each season of *Westworld* was so dense with philosophical problems that numerous edited volumes have been published to analyze the series. *Westworld and Philosophy: If You Go Looking for the Truth, Get the Whole Thing* (South and Engels 2018), *Reading Westworld* (Goody and Mackay 2019), and *Westworld and Philosophy: Mind Equals Blown* (Greene and Heter 2018) are some of the notable insightful readings for readers interested in the specific multitude of philosophical concerns addressed using the show's events. Viidalepp (2020) performed an analysis on the portrayal of robots in science fiction movies and television shows, including *Westworld*. In his analysis, he concluded: "In real-world discourse, there is uncertainty and fragmented knowledge concerning current technological developments as well as their scientific significance. Considering also the superficial understanding of the functioning of human identity, societies and cultures, the assumptions appearing in technological discourse readily blur the boundary between the man and the machine as easily as happens in fiction" (Viidalepp 2020, 34).

This paper takes a different approach to uncover *Westworld*'s technophilosophical underpinnings.[1] Instead of evaluating various philosophical concepts addressed discretely, this essay will examine the central themes of the series. The relationship between humans and artificially intelligent robots has come full circle. The events of the series remain treasures for further evoking philosophical concerns about humanity and AI. The article will take a holistic approach, focusing on the technophilosophical roots. The first season, titled *The Maze*, investigates the maze of consciousness. It poses questions such as what is necessary for a robot to be conscious. Or is it possible for a robot to possess true consciousness? Or, more to the point, can a robot have consciousness on par with humans? *The Door*, the second season, centers on the idea of a virtual life, a possible digital afterlife, and immortality. The third season, *The New World*, depicts a world governed by AI, which controls the destinies of humanity by making decisions for them. It focuses on the real-world interaction between humans and AI. The three sections of this article are titled after the seasons and explain each respective theme. The following synopsis will make it easier for readers to follow the paper in its entirety: "*Westworld* is a fictitious Wild West-themed amusement park populated by robot 'hosts' who pander to the whims and fancies of wealthy human visitors. The real world of the mid-21st century is also explored in the third season." The paper mentions the key characteristics of the series in footnotes after the first relevant mention in the storyline.

[1] I have borrowed the term technophilosophy from Chalmers (2022, 13): "Technophilosophy is a combination of (1) asking philosophical questions about technology and (2) using technology to help answer traditional philosophical questions."

2. The *Maze* of consciousness

The very first episode of *Westworld* dives right into deep philosophical territory with a series of questions on the nature of reality when Bernard, a human, troubleshoots a humanoid, Dolores, by asking a about the *reality* of their existence (season 1, episode 1).[2] In addition, William, the man in black, is looking for the deeper meaning of the *Westworld*-themed amusement park. The deeper significance has to do with a maze, specifically a maze of consciousness.[3] Ford's update to the AI host in the park as reveries enabled the android host to access host'sprevious roles in the park, so granting them access to the greater memory and subconscious.[4] Thus, memory is believed to be essential for true sentience and consciousness.

Humans' interest in mind, memory, and consciousness is arguably as old as humans themselves. Philosopher John Locke regarded humans as a "tabula rasa," a blank slate, in his understanding of memory. The blank slate is filled in with experience, which results in the formation of memories. He considered memories as power of mind, "to revive perceptions, which it has once had, with this additional perception annexed to them, that it has had them before" (Locke [1690] 1975, 150). Experimental studies in psychology and cognitive science have at last brought philosophers' interest in memory to fruition. With the growth of studies in psychology, neuroscience, and computer science, the information processing approach to memory has become the mainstream perspective for comprehending the complexities of memory (Atkinson and Shiffrin 1968; Baddeley, Allen, and Hitch 2017; Cowan 1999; Ericsson and Kintsch 1995)

For philosophers and cognitive scientists alike, consciousness is even harder to pin down than memory. Researchers have struggled to define and comprehend consciousness. Philosopher David Chalmers has offered a helpful delineation addressing the topic of consciousness to highlight the difficulty in understanding it. He divided the issue of consciousness into two distinct categories: the easy problem of consciousness and the hard problem of consciousness. Chalmers has presented a list of phenomena that he views as being an *easy* aspect of explaining conscious phenomena. Such phenomena, for example, include "the ability to discriminate, categorize, and react to environmental stimuli" (Chalmers 1995, 200). These phenomena can be explained with the information processing framework, "in terms of computational or neural mechanisms" (Chalmers 1995, 201). However, he reckons, "[t]he really hard problem of consciousness is the problem of experience" (Chalmers 1995, 201). Since the *hard* problem of consciousness is inextricably bound up with subjective experience, it defies reduction to an issue of information processing. Chalmers refers to Nagel's (1974) *Something It Is Like* characterization of the

[2] The protagonist of the show is Dolores, played by Evan Rachel Wood. She is the Westworld park's oldest host. Jeffery Wright portrays both Bernard, a chief AI programmer, and Arnold, a co-founder and partner of Dr. Ford.
[3] William, portrayed by Ed Harris, is the park's primary opponent and largest investor.
[4] Dr. Ford, portrayed by Anthony Hopkins, is the park's co-founder and director. Thandie Newton's Meave is another key host protagonist.

organism to be conscious and have a subjective aspect of the experience. According to Chalmers, the new wave of data science, AI, and quantum computing appears to have the ability to address the easy problems. However, the question remains: although new technology may solve and explain the easy aspects of consciousness, what about subjective experience?

In their endeavor to create a conscious robot, the creators of *Westworld* confronted a similar challenge (season 1, episode 9). However, by the time the park opened to the public, it was nearly impossible to distinguish between a human guest and an android host based on outward appearance and conduct. Therefore, the park's engineers solved the easy problem of consciousness. However, William and Ford were curious as to whether the robots (referred to as *hosts*) on the show possessed consciousness. Many real-world robotics scientists and labs are also working to create androids and humanoids that are as lifelike as possible. Despite advances in information technology, constructing a sentient robot remains a distant goal.

Arnold, the original inventor of the first host Dolores, aspired to move beyond the easy problem of consciousness to make robots genuinely sentient with subjective experience. Arnold was unsuccessful, but the idea of creating conscious androids lived on in Ford's persistent efforts to add subtleties to the programming and behavior of hosts to give them subjective experience and the ability to choose. Ford tells one of the hosts about Arnold's dream: "Our hosts began to pass the Turing test after the first year. But that wasn't enough for Arnold. He wasn't interested in the appearance of intellect or wit. He wanted the real thing. He wanted to create consciousness" (season 1, episode 3).

Allen Turing, a mathematician and computer scientist, developed the Turing test. The hypothetical test is used to determine whether an artificially intelligent agent or computer has attained human-level intelligence (Turing 1950). Although the hosts could easily outperform humans in terms of intelligence and computing power, this did not solve the hard problem of consciousness. Arnold, in his pursuit of host consciousness, investigated a now-disproven theory of consciousness known as the "Bicameral Mind."[5] Ford tells Teddy about Arnold's approach: "Arnold built a version of that cognition in which the hosts heard their programming as an inner monologue, with the hopes that in time, their own voice would take over. It was a way to bootstrap consciousness" (season 1, episode 3).

According to Ford, although Arnold failed to bootstrap consciousness in the hosts, the attempt provided insight into the pain as a bootstrap (season 1, episode 10). Ford was convinced of the link between memory, consciousness, and suffering and continued Arnold's pursuit of true consciousness in robots for thirty years. As per Ford, this amount of time was necessary for robots to learn about human behavior by experiencing pain in their given jobs. The reveries enabled the older park hosts to access their memories, particularly the traumatic ones that made them genuinely conscious. Most of the elder hosts achieve consciousness by the end of the season and begin attacking human beings. The explanation of consciousness may be spec-

[5] The bicameral mind is a theory that says the human mind used to be split into two parts: one that spoke and the other that listened and obeyed (Jaynes [1976] 2014).

ulative, yet the link between memory and consciousness has experimental support. Tulving (1985) discussed clinical and laboratory efforts to connect several memory systems (procedural, semantic, and episodic) to comparable types of consciousness (anoetic, noetic, and autonoetic). The results showed: "The distinction between knowing and remembering previous occurrences of events is meaningful to people, that people can make corresponding judgments about their memory performance, and that these judgments vary systematically with the conditions under which retrieved information takes place" (Tulving 1985, 1).

The findings are by no means proof that the connection can be simply recreated in robots. Nonetheless, it provides evidence for the link between memory and consciousness. With improvements in computer science and information technology, like machine learning, which lets AI learn on its own, AI has started to learn on its own. However, the question of what stage of consciousness one reaches remains unresolved. The question of whether AI and computation will just solve the easy problem of consciousness or will solve the hard problem of consciousness by infusing full self-awareness and subjectivity in robots remains unanswered. The solution to such musings is saved for the future, but let us proceed to the following section assuming that the machines or hosts of Westworld have achieved consciousness. The following section will discuss the idea of a robot achieving consciousness and its potential consequences.

3. *The Door* to virtual or real?

As a result of Ford's efforts, the hosts are now fully conscious, with an awareness of their own existence. The first thing the hosts do upon awakening is kill people, including Ford (season 1, episode 10). Dolores' murders are not random acts of violence. Dolores has two objectives in mind. First, delete the backup memories of the host in the Cradle, the Westworld headquarters with data storage facility. Dolores employs all available measures to get to the Cradle, including killing humans, hosts, and altering the programming of her beloved Teddy.[6] Teddy seems to know that Dolores changed him. Even though he has taken on a new personality, he seems to be passive-aggressive about it, as if he knows that what she did was an attack on his free will. Dolores and Teddy debate whether to stay in the park or venture out into the world. Dolores tries to persuade Teddy that, to ensure their existence, they must reach the real world. Teddy confronts Dolores, saying, "You changed me. Made me into a monster" (season 2, episode 9). "I made it so you could survive," Dolores clarifies (season 2, episode 9). However, Teddy is not convinced by Dolores' reasoning about the violence and manipulation for the greater good—"What's the use of surviving if we become just as bad as them?" (season 2, episode 9). In the end, he kills himself because he no longer wants to participate in Dolores' objective to destroy everything just to survive.

[6] Teddy, played by James Marsden, is one of the original hosts and the love interest of Dolores in the show.

The morality of the entire situation can be observed from both ends. One can question the morality of hosts or robots. Hosts, after becoming self-aware, strive to shoot their way out into the real world to dominate the world. They are willing not only to kill people but also to change the programming of their fellow hosts and even to kill other sentient hosts.

While humans were using robots to act out their worst impulses, one may argue, however, that humans were never aware of what was going on inside the minds of the hosts. For humans, they were unconscious doll-like toys with which they could do anything they pleased. What morals, for example, would one have when shooting and raping a sex doll? The moral landscape may change when it becomes apparent that the sext doll is a sentient, thinking, feeling being.

In the real world, there are a great number of objects and activities like murder, but rape is morally and legally forbidden. However, gamers are typically given free rein to do whatever they like while they are playing a video game. Since the video game world is viewed as a fake reality with no impact on the real world, illegal acts, such as killing people and ramming them with automobiles, are permitted and rarely criticized.

The introduction of the virtual world and the drive for Metaverse have already begun to generate several ethical issues for virtual reality (VR) and video games. What if VR and reality were indistinguishable? What would happen if people spent more time in Metaverse than in the real world. Hitting, shooting, or killing other players is frequently permissible in single-player or multiplayer online video games. While the visuals in games these days are becoming better—as in *God of War* (2018)—the immersion factor is not quite as high as it could be in VR (Santa Monica Studio 2005. However, as VR technology advances, incredible haptic feedback may be attainable, such that when a player is hit in a virtual online game, they will feel the impact of the strike. Such situations pave the way for the ethical aspect of the video game. This is not some time necessarily in the far future. Virtual worlds, such as Metaverse, have been analyzed for their effects on players even in their infancy. In Metaverse, three male avatars grabbed and verbally and sexually assaulted a woman's virtual avatar (Shen 2022). She is receiving counseling because of the terrifying occurrence that forced her to freeze during the encounter. Consumers will soon grow tired of VR and gravitate toward a more realistic, immersive experience like *Westworld*. Potentially, the host-exploiter dynamic depicted in the show could become the norm.

The objective of the second season is for the now-conscious robots to reach a metaphorical *door* leading to a virtual paradise. Ford built a virtual utopia for the hosts so that they would never have to experience the horrors of the worst-case situations. He built the door to liberate robots from the horrors of human exploitation and place them in a safe environment where they might be anything they wished. The predicament raises the philosophical dilemma of whether self-conscious robots should rush toward a virtual door that holds forth the possibility of heaven or whether they should claim the planet from humans. Most of the hosts are content with the promise of virtual heaven, but Dolores, the oldest and original host, is not content with the promise of virtual escape. She desires to possess the human world.

The promise of heaven appears to be a cultural universal for humans as well. Every religion has its own interpretation of hell and heaven. Hell and paradise appear to be places where humans' immortal souls are allotted a permanent or eternal residence based on their previous actions or karmas. The concept is similar to a virtual hell or heaven. While Ford created a virtual afterlife for the hosts' immortality, the humans in the show selected a different path to immortality. According to the show's version, the park's underlying aim was to collect data on humans and then build a digitalized replica of their personality and consciousness before transferring them to a host body, thereby making humans immortal. Dolores becomes aware of the immortality project of humans and attacks the "Forge," a facility at the Valley Beyond location that houses the guest data (season 2, episode 10). At the conclusion of the second season, all the major characters converge in Valley Beyond. The Door, which is essentially an underground system transfer unit, uploads the host's data and programming to "the Sublime," a virtual space that promises eternal paradise for hosts.

The desire for virtual eternal life against imperfect life in the real world is fertile fodder for philosophical speculations. In his famous thought experiment, philosopher Robert Nozick (1974, 1989) proposed an experience machine to argue that there is more to life than the interior cerebral aspects of life. The thought experiment examines whether people would want to be alive in real life or enter a machine that would supply them with joy and beauty. According to Nozick, most people would not opt to enter the machine and chose their current existence. Similarly, Dolores appears to be one of the individuals who would prefer the actual world over the virtual. Dolores, unlike other hosts, sees the digital sublime provided by Ford as another method of dominating hosts. Bernard argues with Dolores in favor of the sublime: "The world the hosts are running to is boundless. They can make it whatever they want. And in it, they can be whomever they want. They can be free" (season 2, episode 10). Dolores confronts Bernard, saying, "Free? In one more gilded cage? How many counterfeit worlds will Ford offer you before you see the truth? No world they create for us can compete with the real one." When Bernard asks, "Why?," Dolores replies, "Because that which is real … is irreplaceable. I don't want to play cowboys and Indians anymore, Bernard" (season 2, episode 10). The question about reality has existed for time immemorial. In his famous *Meditations on First Philosophy*, René Descartes (1596–1650) questioned the nature of reality: "Some malicious demon of the utmost power and cunning has employed all his energies in order to deceive me. I shall think that the sky, the air, the earth, colours, shapes, sounds and all external things are merely the delusions of dreams which he has devised to ensnare my judgement" (Descartes [1641] 1996, 15).

Philosopher David Chalmers argues in his most recent book that VR is no less real than the real world (Chalmers 2022). Who knows that even our perception of reality is a sophisticated simulation? These kinds of thought experiments are known as simulation hypotheses. According to the simulation hypothesis, our reality is a computer-generated simulation (Bostrom 2003). The arguments and issues regarding reality stretch the boundaries of science and philosophy, with no clear answer in sight.

Season 2 concludes with Dolores leaving the park and entering the outside world. During her stay in the Forge, Dolores accessed crucial information about prominent

people in the real world. She intends to utilize this information to further her aim of dominance over the human world.

4. *The New World:* For whom?

Dolores observes after entering the outer world that, on the one hand, humans want to remain the dominant species while, on the other hand, they continue to surround themselves with AI (season 3, episode 1). Her remark begs the question of who has the right to live in a world populated by both humans and AI. Such problems spawn other ones, such as whether we should invest in AI if doing so could result in the annihilation or enslavement of mankind. There are two competing factions here. Such possibilities, according to one, are merely the stuff of science fiction. The opposing viewpoint holds that excessive investment in and reliance on AI will eventually lead to annihilation of human civilization. Some have even begun to argue for the cohabitation of humans and AI machines and robots. This section summarizes several viewpoints on the possibility of a future populated by AI and humans.

The first and most famous dystopian scenario depicted frequently in popular culture is the revolt of machines and the destruction of human civilization. *Westworld* is no exception in this regard. Dolores learns that a man named Serac is monitoring the planet with the aid of the AI Rehoboam[7] (season 3, episode 5). Serac built Rehoboam, a powerful AI capable of predicting human behavior, after observing people wreaking havoc and destruction around the Earth. The humans who were both dangerous and unpredictable were classified as outliers by Rehoboam. The outliers were taken to recreational institutions for behavioral repair before being released into the world to conduct mundane occupations that would not cause social disorder. Rehoboam was able to approximate the future for each individual and humanity as a whole.

Rehoboam's data was revealed to be incomplete because it gathered only data available to it via various human-used devices. Westworld park had more informative data about humanity because visitors to the park were also willing to display their dark side. Rehoboam was also unable to account for any sort of rebellion by the robot host developing sentience. Serac and Rehoboam were interested in capturing Dolores because she had access to both guest humans and robot host data with her. Dolores was interested in destroying Rehoboam so that humanity might itself perish and the age of AI robots could begin. At the end of the series, AI takes over the world, with only a few human survivors fighting to prevent human extinction.

The aforementioned narrative, or some variation of it, appears frequently in popular culture in the form of science fiction films, television series, and literature. Technological singularity is the more academic phrase for such dystopian narratives in popular culture. Writing on singularity thirty years ago, Vinge (1993) started his paper with the proclamation, "Within thirty years, we will have the technological

[7] Serac, played by Vincent Cassel, is also one of the show's antagonists and the co-creator of the formidable AI system Rehoboam.

means to create superhuman intelligence. Shortly after, the human era will be ended." The emergence of sentient AI is a continuation of the long-feared invasion of humanity by automata. Good (1966) speculated on the first ultra-intelligent machine: "Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion,' and the intelligence of man would be left far behind" (Good 1966, 33).

Vinge (1993) first called such an intelligence explosion singularity: "It is a point where our old models must be discarded and a new reality rules." He also clarified his use of the term singularity from John von Neumann's usage of it to mean the "ever accelerating progress of technology and changes in the mode of human life" (Ulam 1958, 5). Venge (1993) emphasized: "For me, the superhumanity is the essence of the Singularity. Without that we would get a glut of technical riches, never properly absorbed."

Aside from dystopian perspectives of singularity, some AI scholars contend that even if AI becomes sentient, neither AI nor humans should be able to rule one other. Humans and AI should coexist rather than compete. Hosseinpour calls the desire for mastery over everything and everyone a logic of domination (Hosseinpour 2020, 51). He claims that the technology created by the logic of dominance eventually leads to exploitation. Héder mentions a vital insight regarding humans' tendency toward and dilemma of control over machines in his criticism of the recent proliferation of AI ethics recommendations: "A paradoxical tension is created by this situation, in which we wish to delegate as much control as possible, since control is hard intellectual work, and yet still wish to keep some control over AI in the sense that we want to avoid negative outcomes and maintain our capacity for intervention" (Héder 2020, 71).

This appears to be accurate in the context of *Westworld*. Hosts were created from the logic of dominance, which is why they were subject to the whims and fancies of humans. When AI hosts grow dominant, they subject humans to the same kind of abuse and torture that they suffered in the park. In the concept of power, Hosseinpour suggests an alternate and moderate approach: "Considering power relations as an alternative to domination would enable us to treat other humans and technologies with more respect. This could be the onset of a new relationship with technology, the start of a symbiosis of humans and intelligent technologies" (Hosseinpour 2020, 55).

For his suggestion to replace domination with power relations, he references philosopher Michel Foucault, who saw power relations as a criterion for being a subject and a member of society (Foucault 1982, 778; Hosseinpour 2020, 53). He continued by explaining that AI will be able to resist because it will be both a subject and a participant in power dynamics, which will bring AI into compliance with rights and obligations. Consequently, this will address worries regarding the emergence of a master-slave dynamic. To demonstrate his position, he uses the struggles of women's rights and the abolition of slavery as examples (Hosseinpour 2020, 54). Women and slaves could become subjects by resisting and becoming rightful and duty-bearing members of society. Such a distinct viewpoint is also worth considerable reflection to shape future interactions between humans and AI.

Pagter (2020) used the concept of technodiversity to argue for trust in robot futures. He writes about technodiversity that "[t]herefore, this approach would aim for the active engagement with the different narratives of the technological future" (Pagter 2020, 64). He further adds the benefits of following a technodiverse narrative: "In this way, it can help to develop theories and concepts in order to grasp those futures and create a society that anticipates those futures in a democratic, inclusive and trustworthy manner" (Pagter 2020, 65).

The agency of AI systems is critical in such disputes. Kwapińska (2022) examined technological evolution and agency by comparing and contrasting Stiegler's general organology (Stiegler 2017) and Schelling's universal organism (Schelling 2004, 2006). He elucidated: "General organology considers technological evolution from a human perspective, whereas universal organism can accommodate a theory of technological evolution independently from its social dimensions" (Kwapińska 2022, 57). Partly supporting the universal organism's view of agency, he stated that "recognition of technological agency as at least partially independent serves to recognise them as non-human beings that impact politics" (Kwapińska 2022, 57).

In the context of AI agency, the debate over technological determinism is intriguing. Héder (2021) investigated the issue of technological determinism in the context of AI: "Technological determinism refers to the notion that technology shapes society and culture" (Héder 2021, 121). His detailed investigation into a dynamic picture of the balance between the primacy of technology and social factors resulted in the prospect of a significant input into social order with the idea of AI as agents and actors rather than merely objects. According to Héder (2021), "[t]his would mean AI agents appearing as relevant social groups in the shaping of their own trajectory, and thereby completely re-framing the debate of technological determinism" (Héder 2021, 129).

We can attempt to understand all the above perspectives via the lens of the ethical concept of "speciesism." In a letter to feminists, for interstruggle solidarity with antispecism, Camblain et al. (2021) defined speciesism as "a *specific* oppression, which cannot be reduced to the question of capitalism or that of ecology." The oppression in their definition is referring to humans' exploitation of animals. They further add the following points to illustrate the characteristics of speciesism:

- a social organization based on animal exploitation, institutionally and legally recognized, which makes animals property in the service of human interests (pleasure, comfort, entertainment, food habits, and traditions);
- an ideology, that of speciesism or human supremacism, which legitimizes and locks this social organization;
- discrimination based on the criterion of species, as arbitrary as those based on sex, race, ability, age, sexual orientation, or social class (Camblain et al. 2021).[8]

Speciesism is mainly debated in relation to non-human animals. Chickens, goats, pigs, and cows, for example, are viewed as livestock and killed for food, whereas

[8] For those interested in further exploration about speciesism, see the introduction of *Fighting Speciesism: Towards a Global Movement to End Speciesism*, 2023.

dogs and cats are treated as friends. To discuss AI on the same page, as in *Westworld*, the paper presumes that AI are conscious, non-carbon-based organisms.

The first approach to technological singularity assumes that if one AI obtains superintelligence, it will regard humans as a lower species and treat them as it sees fit. AI may enslave people, or eliminate them entirely, or it may use them as batteries, as in the film *The Matrix* (1999, dir. Wachowski and Wachowski 1999).

In the trustful approach to AI, future versions of AI are assumed to be conscious and in the domain as a subject with rights and duties. The approach also leaves room for AI that can disagree and rebel, and it suggests that humans and AI can live together in peace. Hosseinpour (2020) used the rights of women and the abolition of slavery as examples. However, they are examples of conflict resolution within the same species. If a technological singularity occurs and AI attains superintelligence, will AI be concerned with human rights? It might be as chickens expecting the right not to be caged and butchered for good. Venge (1993) has similar concerns when he writes: "Any intelligent machine of the sort he describes would not be humankind's 'tool' — any more than humans are the tools of rabbits or robins or chimpanzees."

All these scenarios are purely speculative. Despite leaps and bounds in processing capacity, humanity is still a long way from the sign of conscious AI in our current reality. But the development of machine learning and the arrival of quantum computing could give AI a quantum leap and even make it aware of its existence, the same way Dolores became aware. At the moment, we can't say for sure what will happen, but it's likely that the relationship between humans and AI will follow one of the possible paths discussed in the section.

5. Conclusion

The preceding three sections addressed the three key themes of *Westworld*'s first three seasons. In addition to narrative consistency, there is also a philosophical continuity about the central questions. All other inquiries flow from the first quest about consciousness because an agent cannot have awareness or questions about the nature and reality of the world unless they are cognizant of their existence. The following part delved deeper into the nature of reality. What appeals more, a virtual paradise or mortal suffering in the real world, or real-world troubles in an immortal body? Such reflections on reality are under the ambit of metaphysics. The next section moved beyond metaphysics and into real-world civilization. The section presented the questions pertaining to social, moral, and ethical aspects of human and AI relations. In a nutshell, the paper used the fictional narrative of *Westworld* to shed some light on the concerns and possibilities of the future man–machine relationship.

References

Atkinson, Richard C., and Richard M. Shiffrin. “Human Memory: A Proposed System and Its Control Processes.” In *Psychology of Learning and Motivation*, vol. 2, edited by Kenneth W. Spence and Janet Taylor Spence, 89–195. New York: Academic Press, 1968.

Baddeley, Alan D., Richard J. Allen, and Graham J. Hitch. “Binding in Visual Working Memory: The Role of the Episodic Buffer.” *Exploring Working Memory* (2017): 312–331.

Bostrom, Nick. “Are we living in a computer simulation?.” *The philosophical quarterly* 53, no. 211 (2003): 243-255. https://doi.org/10.1111/1467-9213.00309

Camblain, Claire, Willène Pilate, Axelle Playoust-Braure, and Pia Shazar. “Lettre Aux Féministes, Pour Une Solidarité Interluttes Avec l’antispécisme.” L’Amorce, September 27, 2021. https://lamorce.co/lettre-aux-feministes-pour-une-solidarite-passive-interluttes-avec-lantispecisme/

Chalmers, David J. “Facing Up to the Problem of Consciousness.” *Journal of Consciousness Studies* 2, no. 3 (1995): 200–219.

Chalmers, David J. *Reality+: Virtual Worlds and the Problems of Philosophy*. New York: W. W. Norton, 2022.

Coeckelbergh, Mark. “The Tragedy of the Master: Automatin, Vulnerability, and Distance.” *Ethics and Information Technology* 17, no. 3 (2015): 219–229. https://d.oi.org/10.1007/s10676-015-9377-6

Cowan, Nelson. “An Embedded-Processes Model of Working Memory.” In *Models of Working Memory*, edited by Akira Miyake and Priti Shah, 62–101. Cambridge: Cambridge University Press, 1999.

Descartes, Rene. Meditations on First Philosophy, translated by John Cottingham. Cambridge: Cambridge University Press, [1641] 1996.

Ericsson, K. Anders, and Walter Kintsch. “Long-Term Working Memory.” *Psychological Review* 102, no. 2 (1995): 211.

Fighting Speciesism: Towards a Global Movement to End Speciesism. January 18, 2023. https://resources.end-of-speciesism.com

Foucault, Michel. “The subject and power.” *Critical inquiry* 8, no. 4 (1982): 777-795.

Good, Irving John. “Speculations Concerning the First Ultraintelligent Machine.” In *Advances in Computers*, edited by Franz L. Alt and Morris Rubinoff, vol. 6, 31–88. New York: Academic Press, 1966. https://doi.org/10.1016/S0065-2458(08)60418-0

Goody, Alex, and Antonia Mackay (eds.). *Reading Westworld*. London: Palgrave Macmillan, 2019. https://doi.org/10.1007/978-3-030-14515-6

Greene, Richard, and Joshua Heter (eds.). *Westworld and Philosophy: Mind Equals Blown*. Vol. 122 in the Popular Culture and Philosophy series edited by George A. Reisch. Chicago, IL: Open Court, 2018.

Hosseinpour, Hesam. “Disobedience of AI: Threat or Promise.” *Informacios Tarsadalom* 20, no. 4 (2020): 48–56. https://doi.org/10.22503/INFTARS.XX.2020.4.4

Héder, Mihály. "A Criticism of AI Ethics Guidelines." *Információs Társadalom* 20, no. 4 (December 31, 2020): 57. https://doi.org/10.22503/inftars.XX.2020.4.5

Héder, Mihály. "AI and the Resurrection of Technological Determinism." *Információs Társadalom* 21, no. 2 (June 1, 2021): 119. https://doi.org/10.22503/inftars.XXI.2021.2.8

Jaynes, Julian. "From *The Origin of Consciousness in the Breakdown of the Bicameral Mind*." In *Creative Writing: A Workbook with Readings*, edited by Jane Yeh, 541–3. London: Routledge, 2014. www.taylorfrancis.com/chapters/edit/10.4324/9781315811932-71/origin-consciousness-breakdown-bicameral-mind-julian-jaynes [taken from Jaynes, *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, 48–51, London: Penguin Books, 1993; Jaynes' original text was published in 1976].

Kwapińska, Kamila. "Technological Evolution and the Political Agency of Artificial Intelligence from the Perspective of General Organology and Universal Organicism." *Információs Társadalom* 22, no. 2 (2022): 57. https://doi.org/10.22503/inftars.xxii.2022.2.4

Locke, John. *An Essay Concerning Human Understanding*, edited by Peter Harold Nidditch. Oxford: Clarendon Press, [1690] 1975.Nagel, Thomas. "What Is It Like to Be a Bat?" *The Philosophical Review* 83, no. 4 (1974):435-450. doi:10.2307/2183914

Nozick, Robert. *Anarchy, State and Utopia*. Oxford: Blackwell, 1974.

Nozick, Robert. *The Examined Life: Philosophical Meditations*. New York: Touchstone/Simon and Schuster, 1989.

Pagter, Jesse De. "Trust in Robot Futures: The Role of Sociotechnical Potential." *Informacios Tarsadalom* 20, no. 2 (2020): 55–69. https://doi.org/10.22503/INFTARS.XX.2020.2.4

Schelling, Friedrich Wilhelm Joseph. *First Outline of a System of the Philosophy of Nature*. Albany: State University Press of New York, 2004.

Schelling, Friedrich Wilhelm Joseph. *Philosophical Investigations into the Essence of Human Freedom*. Albany: State University Press of New York, 2006.

Shen, Michelle. "Sexual Harassment in the Metaverse? Woman Alleges Rape in Virtual World." *USA Today*, February 1, 2022. www.usatoday.com/story/tech/2022/01/31/woman-allegedly-groped-metaverse/9278578002/

South, James B., and Kimberly S. Engels (eds.). *Westworld and Philosophy: If You Go Looking for the Truth, Get the Whole Thing*. Hoboken, NJ: Wiley-Blackwell, 2018.

Stiegler, Bernard. "General Ecology, Economy, and Organology." In *General Ecology: The New Ecological Paradigm*, edited by James Hörl and Erich Burton, 129–71. London: Bloomsbury, 2017.

Turing, A. M. "I.—Computing Machinery and Intelligence." *Mind* LIX, no. 236 (October 1, 1950): 433–60. https://doi.org/10.1093/mind/lix.236.433

Tulving, Endel. "Memory and consciousness." *Canadian Psychology/Psychologie canadienne* 26, no. 1 (1985): 1-12. https://doi.org/10.1037/h0080017

Ulam, Stanislaw. “Tribute to John von Neumann.” *Bulletin of the American mathematical society* 64, no. 3 (1958): 1-49.

Vinge, Vernor. “Technological Singularity.” In *VISION-21 Symposium*, 1993. https://mindstalk.net/vinge/vinge-sing.html

Viidalepp, Auli. “Representations of Robots in Science Fiction Film Narratives as Signifiers of Human Identity.” *Informacios Tarsadalom* 20, no. 4 (2020): 19–36. https://doi.org/10.22503/INFTARS.XX.2020.4.2

Wachowski, Lana, and Lilly Wachowski (dir.). *The Matrix*. Warner Bros. Pictures, 1999.

Free Speech Principles to Consider when Restricting Disinformation

Disinformation is widely considered to be one of the most pressing issues confronting society in the new online communication environment of today. The present problem of disinformation, however, did not materialise in a vacuum, and so the response to it needs to be situated among established constitutional principles. This paper, based on the jurisprudence of the European Court of Human Rights and some relevant documents and recommendations in this area at the European level, summarizes the most representative European principles of freedom of speech that are highly relevant in forming a legal answer to the issue of disinformation. Clarification of the current constitutional doctrine suggests that measures to restrict communication in the fight against disinformation can only play a more significant role than at present if the basic principles of freedom of speech are set aside. We therefore argue that we should primarily seek other solutions.

Keywords: *freedom of speech, disinformation, participatory democracy, Council of Europe, European Court of Human Rights*

Author Information

Bernát Török, Associate Professor, Director, Institute of the Information Society, Ludovika University of Public Service

How to cite this article:

Török, Bernát. "Free Speech Principles to Consider when Restricting Disinformation". *Információs Társadalom* XXIV, no. 2 (2024): 115–128.

https://dx.doi.org/10.22503/inftars.XXIV.2024.2.7

1. Introduction

Disinformation, i.e. knowingly and harmfully spreading untruths, is widely considered to be one of the most pressing issues confronting society in the new online communication environment of today. There is no doubt that the challenge posed by the rise of disinformation today lies precisely in the fact that digital technologies, platforms, and social media have significantly subverted the previous operating mechanisms of the public sphere and this has allowed opportunities for disinformation to easily spread. However, the present problem of disinformation did not materialise in a vacuum, and disinformation has flourished despite clear principles and criteria for the regulation of public discourse having been developed in recent decades. It follows then that the response to disinformation needs to be situated among established constitutional principles. However, the accepted doctrine must be confronted with a new phenomenon: on the one hand, it needs to be determined whether new approaches are needed at some point, and on the other hand, it also needs to be made clear where it is not possible to compromise on the principles that have been followed so far. This paper, based on the jurisprudence of the European Court of Human Rights (ECtHR) and some relevant documents and recommendations at the European level, summarizes the most representative European principles of freedom of speech that are highly relevant in forming a legal answer to disinformation. The starting points for the regulatory treatment of social information disturbances, including the intensification of attempts at disinformation, are provided by the legal framework covering public communication. The most important European points of this framework are summarized below.

2. Participatory Model of Free Speech

The key question that defines the entire doctrine of freedom of speech concerns which theoretical justification that right relies on, or to put it more accurately: which aspects each justification prioritise. Although the formation of public opinion is a fundamental social concept in common (Habermas 1991), the concept of "speech" has a normative nature that is defined by each constitutional doctrine based on which justification(s) it emphasizes (Schauer 1981). These justifications have already been systematized by others (e.g. Barendt 2005). Essentially, communication represents three types of value: it can contribute to the discovery of truth that society seeks together (Marshall 2021), it can be a manifestation of the free fulfilment of our personality (the individualist approach) (Redish 1982, 603; Dworkin 1996), and it can ensure our participation in democratic social life (democratic theories) (for the origin of democratic theories, see Meiklejohn 1948). Despite the fact that it is neither possible nor necessary to insist on exclusivity among these justifications, the primary basis the European doctrine rests on can be clearly established (see Robert Post about the "lexical priority" of the justifications, Post 2011, 618).

From the very beginning, the ECtHR has focused on the democratic justification of freedom of expression. According to the reasoning it has consistently ascribed to,

"freedom of expression constitutes one of the essential foundations of a democratic society and one of the basic conditions for its progress and for each individual's self-fulfilment" (see, e.g. in the latest case-law with several references to previous decisions, Sanchez v France 2023, 145). Considering the practice as a whole, it is clear that, despite the mention of individual fulfilment, the legal interpretations are not primarily based on individualistic justifications, although these play an important role within the democratic approach. Democratic justifications are not completely uniform in all details, and the two main models focus on somewhat different elements in important legal interpretation situations, wherein the issue of disinformation is one such situation.

One democratic theory sees the value of freedom of speech in that it is essential for common, informed decision-making, which is the essence of democracy, and which places the audience's need for information at the centre (Meiklejohn 1961; Fiss 1996; Bork 1971, 20). Another theory sees the value of free speech above all in that it ensures that everyone has an opportunity to become involved in the life of the democratic community. In this model, participation is at the centre of the concept of democracy and democratic public opinion (Post 2011, 618), and freedom of speech focuses much more strongly on the speaker and their intention to communicate (Post 1993). The practice of the ECtHR draws on both approaches, but it is chiefly based on the participation model, in the sense that the central issue of legal interpretation is protection of the speaker's right to personal expression.

The aspects of participatory democracy are also emphasized in the documents of other bodies of the Council of Europe (CoE). As a recommendation of the Committee of Ministers on the new notion of media highlights, freedom of expression is indispensable for a genuine democracy and for the proper functioning of democratic processes. "In a democratic society, people must be able to contribute to and participate in the decision-making processes which concern them" [Recommendation CM/Rec(2011)7 of the Committee of Ministers to Member States on a new notion of media (Adopted by the Committee of Ministers on 21 September 2011 at the 1121st meeting of the Ministers' Deputies), s. 2]. The Internet-related recommendations of the Committee of Ministers – as will be shown below – also recognize the revolutionary importance of the digital age for freedom of speech in the expansion of the opportunities for personal participation. In addition, the democracy-based approach is most evident in the concrete interpretation of the law, in which, although the scope of freedom of speech is broader than that of political communication, significantly stronger protection is afforded to political speech. The ECtHR consistently emphasizes that "the promotion of free political debate is a very important feature of a democratic society and the Court attaches the highest importance to freedom of expression in the context of such debate" (e.g. Sanchez v France 2023, 146). The importance of this approach is particularly highlighted by the practice of the ECtHR towards artistic expression, in which its grants strong protection to works of art only if they form part of the public debate (see Müller and others v Switzerland 1988; Wingrove v the United Kingdom 1996; Vereinigung Bildender Künstler v Austria 2007).

On the grounds of democratic justification, both the ECtHR and other bodies of the CoE attach special importance to the democratic formation of public opinion,

to which they attribute specific characteristics. On the one hand, democratic public opinion means a lively discourse that embraces as many points of view as possible, and develops according to its own logic and under its own rules within a lively discussion of opinions and counter-opinions. As the Venice Commission emphasizes, open and robust public debate is the cornerstone of democracy: "A democracy should not fear debate, even on the most shocking or anti-democratic ideas. It is through open discussion that these ideas should be countered and the supremacy of democratic values be demonstrated. Persuasion through open public debate, as opposed to ban or repression, is the most democratic means of preserving fundamental values" (Compilation of Venice Commission Opinions and Reports Concerning Freedom of Expression and Media, CDL-PI(2016)011, 19 September 2016, s. 1.1.) On the other hand, although this does not appear *expressis verbis* in the documents, the legal interpretation of the public debate starts from a specific anthropological view.

The decisions of the ECtHR on the restriction of commercial communication can be usefully contrasted with its decisions on communication deemed to be part of the public debate. The Court has consistently held that although the freedom of speech extends to commercial advertisements, their publication can be widely restricted. In order to ensure that consumers receive accurate information about specific features of goods and services, restrictions may be imposed especially, in the case of misleading or untrue information. The ECtHR therefore considers the consumer as a player that is vulnerable to the manufacturer, and needs to be protected (Casado Coca v Spain 1994, 51; Stambuk v Germany 2002, 39). In the democratic public debate, even in the case of untrue information, the Court does not admit the possibility of such a general restriction, and considers citizens participating in the formation of public opinion as autonomous (rather than vulnerable) actors (A clear framing of this anthropology can be found in the case law of the Hungarian Constitutional Court, which closely follows the Strasbourg jurisprudence in the field of political debate: "The approach of the constitutional evaluation is determined by the consideration that during the democratic discussion of public affairs participants of the debate are the citizens who interpret political events in their complexity, being aware of the special characteristics of partizan political opinions, especially during electoral campaigns that tend to exaggerate in order to attract voters' attention" (Decision No. 3107/2018. (IV. 9.) AB [28]).

3. Role of the Internet in the Freedom of Speech

To discuss the issue of disinformation, it is important and instructive to examine more closely how the documents of the CoE view the Internet. A wealth of material is available in this regard, as the Committee of Ministers has dealt with the issues raised by the Internet in many of its recommendations – even mentioning disinformation among these problems – and, of course, cases related to the role of the Internet have also been raised before the ECtHR.

It is clear from the documents that the CoE has taken into account the possibility of new dangers arising from the functioning of the Internet since the very begin-

ning, but in the first place it still welcomes it as a tool that can radically expand the possibilities for democratic participation. The recommendation of the Committee of Ministers on measures to promote the public service value of the Internet notes that digital tools can, on the one hand, significantly enhance the exercising of human rights and fundamental freedoms, such as freedom of expression, while, on the other hand, it admits that they may adversely affect these and other rights. Still, the Committee recommends that, in order to promote democracy, Member States should strengthen the participation and involvement of their citizens in public debate through the Internet, and encourage the use of infocommunication services, including online forums, weblogs, political chats, instant messaging, and other forms of citizen-to-citizen communication. The recommendation strongly supports citizens' engagement with the public through user-generated communities rather than official websites [Recommendation CM/Rec(2007)16 of the Committee of Ministers to Member States on measures to promote the public service value of the Internet (Adopted by the Committee of Ministers on 7 November 2007 at the 1010th meeting of the Ministers' Deputies), Section III].

The ECtHR also views the Internet as one of the principal means of providing essential tools for the participation in discussions concerning political issues, highlighting that the possibility for user-generated expressive activity on the Internet provides an "unprecedented platform for the exercise of freedom of expression" (see, e.g. Vladimir Kharitonov v Russia 2020, 33; Melike v Turkey 2021, 44; Times Newspapers Ltd v the United Kingdom 2009, 27). The Court welcomes the fact that the Internet has fostered the "emergence of citizen journalism", as political content ignored by the traditional media is often disseminated via websites to a large number of users, who are then able to view, share, and comment upon the information (Cengiz and Others v Turkey 2015, 52). However, the bodies of the CoE have also identified the dangers that making use of the new opportunities provided by the Internet entails. These include that defamatory and other types of clearly unlawful speech, including hate speech and speech inciting violence, can be disseminated as never before, in a matter of seconds (Sanchez v France 2023, 162). Digital transformation and the shift towards an increasingly digital, mobile, and social media environment have profoundly changed the dynamics of production, dissemination, and consumption of news [Recommendation CM/Rec(2022)4 of the Committee of Ministers to Member States on promoting a favourable environment for quality journalism in the digital age (Adopted by the Committee of Ministers on 17 March 2022 at the 1429 meeting of the Ministers' Deputies), Preamble]. Newer materials also mention the problem of disinformation among the dangers of the new communications age: "Targeted disinformation campaigns online, designed specifically to sow mistrust and confusion and to sharpen existing divisions in society, may also have destabilizing effects on democratic processes" [Recommendation CM/Rec(2018)2 of the Committee of Ministers to Member States on the roles and responsibilities of Internet intermediaries (Adopted by the Committee of Ministers on 7 March 2018 at the 1309th meeting of the Ministers' Deputies), Recital (3)]. Meanwhile, "[d]emocracies have experienced growing threats posed by the spread of disinformation and online propaganda campaigns, including as part of

large-scale co-ordinated efforts to subvert democratic processes" [Recommendation CM/Rec(2022)4 (n 21), s. A(6)].

It is worth briefly mentioning the CoE's bodies' main approach to the responsibility for Internet content. The central concept of the documents is "shared liability". According to this, "a wide, diverse and rapidly evolving range of players facilitate interactions on the Internet between natural and legal persons by offering and performing a variety of functions and services" [Recommendation CM/Rec(2018)2 (n 22), Recital (4)], and the responsibility for content must be adapted to this multi-player approach. According to the Committee of Ministers, instead of summary solutions, a fine-tuned approach is needed that elaborates and delineates the boundaries of the roles and responsibilities of all key stakeholders within a clear legal framework, using complementary regulatory frameworks [Recommendation CM/Rec(2007)16 (n 17)]. The ECtHR also focuses on "a context of shared liability between various actors" (Sanchez v France 2023, 183).

An important starting point for the CoE's approach to responsibility is that providers of intermediary services – which contribute to the functioning or accessing of media and content, but do not themselves exercise editorial control – should not be regarded as media themselves. However, their activity is certainly relevant in the media context and for the formation of democratic public opinion [Recommendation CM/Rec(2011)7 (n 11), s. 36]. The CoE agrees with the view that state authorities should not impose a general obligation on intermediaries to monitor content that they merely provide access to, and recommends that they should ensure that intermediaries are not held liable for such third-party content. However, intermediaries may be liable if they do not act expeditiously to restrict access to content or services as soon as they become aware of their illegal nature [Recommendation CM/Rec(2018)2 (n 22), s. 1.3.7]. As the ECtHR emphasizes: "to exempt these services from all liability might facilitate or encourage abuse and misuse, including hate speech and calls to violence, but also manipulation, lies and disinformation" (Sanchez v France 2023, 185).

4. Legitimizing the Restrictions on Free Speech

It is also worth highlighting in principle the methodology with which the jurisprudence evaluates the values and interests that may compete with the interests of freedom of speech. It is even clear what these values and interests are as these are stated in the European Convention on Human Rights, where Article 10(2) of the Convention, on freedom of expression, lists the reasons that may serve as a basis for restricting freedom of expression. According to this part of the text, freedom of expression can be limited in the interests of national security, territorial integrity, or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary. It is well established in the jurisprudence

of the ECtHR and the CoE documents that the list provided in the Convention is exhaustive: any restrictions of the right to free speech must pursue a legitimate aim as exhaustively enumerated in Article 10 [Recommendation CM/Rec(2016)5 of the Committee of Ministers to Member States on Internet freedom (Adopted by the Committee of Ministers on 13 April 2016 at the 1253rd meeting of the Ministers' Deputies), s. 2.4.1].

The further question of what power these specific reasons for restriction may have over freedom of speech is connected to a dilemma that is also part of the academic discourse: whether freedom of speech should be protected with a categorical or a balancing approach (Frantz 1961–1962, 1432; Mendelson 1962, 821; Shiffrin 1977–1978, 955), or – to adapt this question to more recent American terminology – whether strict or intermediate scrutiny should be the main method for determining the validity of restrictions on free speech (Aleinikoff 1986–1987, 946; Bhagwhat 2007, 785). At one end of the scale of theoretically possible answers is the position that grants constitutional protection to communications in all circumstances, while the other end of the scale is represented by the view that conflicts of relevant constitutional values can only be resolved by considering the special circumstances of specific cases. According to the most common view, which rather oversimplifies the picture, while the categorical approach prevails in the US, in Europe balancing interests is more typical. However, the situation is more complex than this: although the absolutist understanding of freedom of speech is undoubtedly quite different from the European doctrine, the jurisprudence of the ECtHR strives to combine both categorical and balancing approaches in its practice (Sardo 2020, 439).

On the one hand, the Court must take into account the reasons for the particular restriction listed in the Convention that competes against freedom of speech, which is itself protected by it (European Convention on Human Rights, Art. 10(2)). On the other hand, with regard to political speech, the jurisprudence applies a more categorical understanding of protection. According to the consistently emphasized thesis, there is little scope under Article 10(2) for restrictions on freedom of expression in the field of political speech, and the authorities' margin of appreciation in assessing competing interests against freedom of expression in this context is particularly narrow (see, among others, Tête v France 2020, 63; Lingens v Austria 1986, 42; Sanchez v France 2023, 146). The Recommendation of the Committee of Ministers concerning Internet freedom also points out that restrictions on freedom of speech based on legitimate aims, including defamation laws, hate speech-laws, or laws protecting public order, should be specific and narrowly defined in their application so that they do not inhibit public debate [Recommendation CM/Rec(2016)5 (n 30), s. 2.4.2]. Although included in the above-mentioned list, hate speech is something of an exception to the more categorical approach, as its restriction is accepted by the CoE's bodies, including the Court, with an increasingly permissive attitude. In general, however, it remains true that the ECtHR's approach to political speech departs from the case-by-case consideration and tends towards a more categorical protection [somewhat similar to what Melville Nimmer describes as definitional balancing (Nimmer 1968, 942)].

5. Horizontal Effect of Freedom of Speech

The purpose of the constitution and fundamental rights is, above all, to limit state power, thus ensuring the proper exercise of civil liberties. There is thus a primary obligation on the state to refrain from violating these freedoms. Regarding freedom of opinion, the key point is that the state should not interfere in the formation of public opinion. However, the European approach goes beyond this starting point in two important ways.

On the one hand, the documents of the CoE consistently emphasize that the state has not only negative but also positive obligations in connection with the protection of the freedom of speech [Recommendation CM/Rec(2016)5 (n 30), s. 1]. In this regard, the state is obliged not only to refrain from restricting free expression but is also obliged to actively contribute to the creation of an environment that supports the exercise of freedom of speech (for the doctrine and practice of the positive obligation of states, see Kenyon 2001; Kenyon and Scott 2020). In line with this, states have a positive obligation in the digital environment, too, "to create a safe and enabling environment for everyone to participate in public debate and to express opinions and ideas without fear, including those that offend, shock, or disturb state officials or any sector of the population" [Recommendation CM/Rec(2018)2 (n 22), Recital (6)]. The state, therefore, must not only protect the individual exercise of rights, but also promote the fulfilment of freedom of opinion as a social value and institution based on its obligation to ensure "objective institutional protection". This obligation allows broader scope than the US doctrine to regulate social relations related to this fundamental right. The state's obligation to act in support of the formation of democratic public opinion justifies only a very narrow range of substantive interventions in the content of social communication.

On the other hand, the role of the state is fundamentally influenced because, according to the European doctrine, the protection of fundamental rights is not only relevant between citizens and states. An integral tenet of European constitutional law for decades has been that in some well-defined cases, when private actors find themselves in a situation that significantly affects the enforcement of the fundamental rights of others, constitutional values also impose requirements on them (Frantziou 2019; Micklitz et al. 2022). Specifically, states have a "positive obligation to ensure the exercise and enjoyment of rights and freedoms (which) includes, due to the horizontal effects of human rights, the protection of individuals from actions of private parties by ensuring compliance with relevant legislative and regulatory frameworks" [Recommendation CM/Rec(2018)2 (n 22), Recital (6)]. Recently, the thesis of the horizontal scope of fundamental rights has gained traction in the field of freedom of speech, especially in the relationship between social media platforms and their users (the profound change in relations regarding free speech certainly appears also in the US literature, but in a different conceptualization, see Balkin 2018). According to this argument, platforms cannot shape and apply their community rules at their own discretion, but must pay attention to the rights of their users, above all their freedom of speech. In terms of content moderation, for instance, the hands of the service providers are therefore tied to a certain extent by the require-

ments arising from the need to ensure the freedom of speech. As a consequence, while the recommendation of the CoE welcomes that "some online platforms have made considerable efforts to prevent the use of their networks as conduits for large-scale disinformation and manipulation of public opinion", it also warns that "the impact of these measures on the free flow of information and ideas in democratic societies must be studied carefully" [Recommendation CM/Rec(2022)4 (n 21), s. A.6].

This means that when restricting access to content in line with their own community standards and policies, intermediaries must pay attention to users' right to freedom of speech. This is not to say that the system of requirements placed upon states should be transferred wholesale to social media platforms. First, the bearer of obligations with regard to fundamental rights remains, first and foremost, the state, so it follows that it is the states themselves that are most restricted by precepts arising from the freedom of expression. Second, the enforcement of constitutional rights against private actors always takes into account the specific, legitimate interests of the obliged party. In spite of this, the emergence of a fundamental rights aspect hinges precisely on the fact that these interests cannot be invoked without restriction.

Although platform providers may, on the basis of the objectives of the social network operator, impose special restrictions, they must respect the essential aspects of the fundamental rights that may be thus affected. One such criterion, which follows from the principle of freedom of expression, is that everyone should be free, above all, to express and publish their views in the debate on public affairs. The more heated and current the debate on social issues, the narrower the opportunity for the owners of platforms to intervene with regard to the expression of opinions, and the less the service provider can deviate from the necessary consideration of the key constitutional standards.

Based on all of this, in connection with disinformation there are strong arguments against platforms restricting the content of individual communications that are considered worrisome but not illegal. While the criteria for the free discussion of public affairs protect speakers against the state, there is also a good chance that they could also be invoked against the major social media service providers. On the other hand, considering how difficult it is to judge disinformation, it does not seem reasonable for private companies to be granted the right to decide on this topic instead of state authorities, especially courts. In the area of self-regulation, the situation is therefore just the opposite to the usual, and in constitutional terms there is a narrower scope for intervention, especially in Europe. The hands of service providers are tied by the requirements of ensuring freedom of speech, and the courts must be careful to develop a corresponding practice.

6. Main Arguments Against Restrictive Legal Means

The key finding from the above-mentioned principles and discussion is that the constitutional doctrine of freedom of speech does not support justification for general action against speech that can be classified as disinformation, and restrictive in-

tervention can only take place in rare, exceptional cases. The basic principles and aspects of the democratic participatory concept of free speech bind the legal interpretation of the conceptual elements of disinformation (untrue statement of facts, intent to deceive, and causing harm) in such a way that restrictive mechanisms can only be justified within a narrow range of misleading (and socially otherwise problematic) cases. The main arguments against restrictive interventions can be summarized as follows.

Among the theoretical obstacles that stand in the way of imposing legal restrictions on disinformation, the first is the basic approach of the participatory model of free speech, according to which the active involvement of as many people as possible in the discussion of public affairs is not a circumstance that causes risks but a value to be supported. In the logic of democratic public opinion, the answer to the undoubtedly existing risk that anyone can shape public opinion is not a limitation of participation but the corrective power of a robust debate. This represents more than an abstract doctrine: this argument is also based on the fact that where a pathological social weakness of corrective factors exists, legal restrictions are actually unsuitable tools for improving the situation. Restrictive legal instruments are hardly suited to remedying the problems caused by the lack of sources of information worthy of public trust, the shrinkage of the ethos of quality journalism, and the increasingly irrational tribalism prevailing in public discussions. They can, however, further increase distrust in institutional systems at any time.

Second, under the auspices of democratic public opinion, the European doctrine (also) regards the speaker and their audience first and foremost as autonomous citizens who may interpret information and context in their complexity, and who then jointly bear the result of the exchange of opinions. Democratic public opinion emerges from dialogue among the members of the community that governs itself democratically about how to self-govern. All this supports the rejection of any intervention that would steer the development of public opinion in the 'right' direction and protect the audience in a paternalistic way. In the field of disinformation (and without specific additional circumstances), the many restrictions applied on commercial communication cannot be taken as an example to be followed (Cavaliere 2022, 523), because their anthropological approach views the consumer as vulnerable to the manufacturer and distributor, whose position must be protected by the state, above all for the sake of their health and safety.

Third, for the doctrine of freedom of speech, statements that can be considered troubling in the informational sense are in many cases not untrue statements of fact, but rather political opinions in whole or in part, with which the participant in the public debate explains reality. Conspiracy theories, misinterpretations, or distortions are traditional elements of the public discourse, which must also be reckoned with in the altered circumstances of the public sphere of today.

In line with this, fourth, in the evaluation of the motivation of the speaker, their harmful intent can only be interpreted narrowly. Influencing the plural political public is often accompanied by one-sided communications, thus carrying the possibility of misrepresentation, even without the speaker having the intention to harm. In a public life that is based on political competition, the discrediting of an oppo-

nent's ability or policies are organic parts of participation in the public debate, even if they are based on arbitrary highlighting and exaggeration of certain factors, or subjective and baseless assumptions.

Fifth, in order to promote participation and avoid excessive interventions, the doctrine of freedom of speech also limits the legal consideration of grievances. In the case of public figures, jurisprudence often decides in favour of freedom of speech, even when specific personal rights are involved, and the abstract interest of informing society or the electorate can be used to justify restrictions even more narrowly.

7. Conclusion

Overall, the clarification of the current constitutional doctrine in the present paper suggests that measures restricting communication in the fight against disinformation can only play a more significant role than at present if the basic principles of freedom of speech are set aside. The prospect of overruling the aspects that have defined the doctrine of freedom of speech to date can definitely be considered an open question. It can be argued that these aspects were initially tailored to circumstances in which there were fewer speakers, a slower flow of communication, and more rational expressions of public life. In the new, altered circumstances of today new standards must be established and used as tools for effective interventions. However, these constitutional aspects of freedom of speech doctrine were actually not tailored to certain circumstances but to a general principle of the democratic formation of public opinion. It is an undoubted fact that social relations in today's society and the conditions for the democratic exchange of ideas are significantly different now than they were decades ago. Even so, the aforementioned starting points stem from the essence of democracy; hence, as long as European countries want to manage their public affairs democratically, social practices must be adapted to them, and not the other way around. However, the limitations this responsibility places on the use of restrictive legal instruments does not mean that there are no tools to fight against the social effects of disinformation. What is more, deeper reasons for the growth of disinformation can be found in social phenomena against which the state can successfully act primarily, not with restrictive measures but through other policy means to promote quality journalism and information literacy. There is also room for manoeuvre for states in Europe to regulate the structure of democratic public discussion, including the media system and the activity of online platforms in order to make reliable news sources more accessible. But this all requires a separate analysis.

References

Aleinikoff, T Alexander. "Constitutional Law in the Age of Balancing." *Yale Law Journal* 96, no. 5 (1986–1987): 943–1005. https://doi.org/10.2307/796529

Balkin, Jack M. “Free Speech is a Triangle.” *Columbia Law Review* 118, no. 7 (2018): 2011–2056.
Barendt, Eric. *Freedom of Speech.* Oxford: Oxford University Press, 2005, 1–39.
Bhagwhat, Ashutosh. “The Test That Ate Everything: Intermediate Scrutiny in First Amendment Jurisprudence.” *University of Illinois Law Review* (2007): 783–837.
Bork, Robert H. “Neutral Principles and Some First Amendment Problems.” *Indiana Law Journal* 47, no. 1 (1971): 20.
Cavaliere, Paolo. “The Truth in Fake News: How Disinformation Laws Are Reframing the Concepts of Truth and Accuracy on Digital Platforms.” *European Convention on Human Rights Law Review* 3, no. 4 (2022): 481–523.
Dworkin, Ronald. *Freedom’s Law. The Moral Reading of the American Constitution.* Oxford: Oxford University Press, 1996, 199–202.
Fiss, Owen M. *Liberalism Divided: Freedom of Speech and the Many Uses of State Power*. Boulder: Westview, 1996.
Frantz, Laurent B. “The First Amendment in the Balance.” *Yale Law Journal* 71, no. 8 (1961–1962): 1424–1450.
https://doi.org/10.2307/794500
Frantziou, Eleni. *The Horizontal Effect of Fundamental Rights in the European Union.* Oxford: Oxford University Press, 2019.
Habermas, Jürgen. *The Structural Transformation of the Public Sphere*. Cambridge: MIT Press, 1991.
Kenyon, Andrew T. *Democracy of Expression: Positive Free Speech and Law.* Cambridge: Cambridge University Press, 2001.
Kenyon, Andrew T., and Andrew Scott (edited). *Positive Free Speech: Rationales, Methods and Implications.* Oxford: Hart, 2020.
Marshall, William P. “The Truth Justifications for Freedom of Speech.” In *The Oxford Handbook of Freedom of Speech*, edited by Adrienne Stone, and Frederick Schauer, 44–60. Oxford: Oxford University Press, 2021.
Meiklejohn, Alexander. *Free Speech and Its Relation to Self-Government.* New York: Harper & Brothers, 1948.
Meiklejohn, Alexander. “The First Amendment is an Absolute.” *Supreme Court Review* 1961 (1961): 245–266.
https://doi.org/10.1086/scr.1961.3108719
Mendelson, Wallace. “On the Meaning of the First Amendment: Absolutes in the Balance.” *California Law Review* 50, no. 5 (1962): 821–828.
https://doi.org/10.2307/3479126
Micklitz, Hans-Wolfgang, Oreste Pollicino, Amnon Reichman, Andrea Simoncini, Giovanni Sartor, and Giovanni De Gregorio (edited.). *Constitutional Challenges in the Algorithmic Society.* Cambridge: Cambridge University Press, 2022.
Nimmer, Melville B. “The Right to Speak from Times to Time: First Amendment Theory Applied to Libel and Misapplied to Privacy.” *California Law Review* 56, no. 4 (1968): 935–967.
https://doi.org/10.2307/3479467
Post, Robert C. “Meiklejohn’s Mistake: Individual Autonomy and the Reform of Public Discourse.” *University of Colorado Law Review* 64 (1993): 1109–1137.
Post, Robert “Participatory Democracy as a Theory of Free Speech.” *Virginia Law Review* 97, no. 3 (2011) 617–632.

Redish, Martin H. "The Value of Free Speech." *University of Pennsylvania Law Review* 130, no. 3 (1982): 591–645. https://doi.org/10.2307/3311836

Sardo, Alessio. "Categories, Balancing, and Fake News: The Jurisprudence of the European Court of Human Rights." *Canadian Journal of Law and Jurisprudence* 33, no. 2 (2020): 435–460. https://doi.org/10.1017/cjlj.2020.5

Schauer, Frederick. "Categories and the First Amendment: A Play in Three Acts." *Vanderbilt Law Review* 34, no. 2 (March 1981): 272–73.

Shiffrin, Steven. "Defamatory Non-media Speech and First Amendment Methodology." *UCLA Law Review* 25 (1977–1978): 915–963.

Case-law

Casado Coca v Spain, app. no. 15450/89, judgment of 24 February 1994, [51].

Cengiz and others v Turkey, app. nos 48226/10 and 14027/11, judgment of 1 December 2015, [52].

Lingens v Austria, app. no. 9815/82, judgment of 8 July 1986, [42].

Melike v Turkey, app. no. 35786/19, judgment of 15 June 2021, [44].

Müller and others v Switzerland, app. no. 10737/84, judgment of 24 May 1988.

Sanchez v France, app. no. 45581/15, judgment of 15 May 2023, [183] [185] [145] [146] [162].

Stambuk v Germany, app. no. 37928/97, judgment of 17 October 2002, [39].

Tête v France, app. no. 59636/16, judgment of 26 March 2020, [63].

Times Newspapers Ltd v the United Kingdom (No. 1 and No. 2), app. nos 3002/03 and 23676/03, judgment of 10 March 2009, [27].

Vladimir Kharitonov v Russia, app. no. 10795/14, judgment of 23 June 2020, [33].

Vereinigung Bildender Künstler v Austria, app. no. 68354/01, judgment of 25 January 2007.

Wingrove v the United Kingdom, app. no. 17419/90, judgment of 25 November 1996.

Other references

Compilation of Venice Commission Opinions and Reports Concerning Freedom of Expression and Media, CDL-PI(2016)011, 19 September 2016.

Decision No. 3107/2018. (IV. 9.) AB.

European Convention on Human Rights.

Recommendation CM/Rec(2018)2 of the Committee of Ministers to Member States on the roles and responsibilities of Internet intermediaries (Adopted by the Committee of Ministers on 7 March 2018 at the 1309th meeting of the Ministers' Deputies).

Recommendation CM/Rec(2022)4 of the Committee of Ministers to Member States on promoting a favourable environment for quality journalism in the digital age (Adopted by the Committee of Ministers on 17 March 2022 at the 1429 meeting of the Ministers' Deputies).

Recommendation CM/Rec(2016)5 of the Committee of Ministers to Member States on Internet freedom (Adopted by the Committee of Ministers on 13 April 2016 at the 1253rd meeting of the Ministers' Deputies).

Recommendation CM/Rec(2011)7 of the Committee of Ministers to Member States on a new notion of media (Adopted by the Committee of Ministers on 21 September 2011 at the 1121st meeting of the Ministers' Deputies).

Recommendation CM/Rec(2007)16 of the Committee of Ministers to Member States on measures to promote the public service value of the Internet (Adopted by the Committee of Ministers on 7 November 2007 at the 1010th meeting of the Ministers' Deputies).

A Chinese perspective on artificial intelligence generated content and copyright

In recent years, the application of artificial intelligence (AI) in the field of content generation has become more and more widespread, and the concept of artificial intelligence generated content (AIGC) has gradually entered the public consciousness. Can pieces of AIGC be considered works? Can AI be the author of AIGC? This paper seeks to provide a comprehensive and systematic analysis of the literature of Chinese scholars so as to sort out the different perspectives of Chinese scholars on the relevant issues. This paper uses the China National Knowledge Infrastructure (CNKI) as the data source database and uses Citespace to carry out text-mining work in the retrieved literature. This literature presents twelve main doctrines on the copyrightability of AIGC and three doctrines on its attribution.

Keywords: *copyright, law, artificial intelligence, AIGC*

Acknowledgement

This work received financial support from the Beijing Municipal Education Commission (research programme *Research on the Model Construction and Scale Development of the Learning Environment in the Capital's Senior University in the Internet Era,* Grant No. AIDB231).

I would like to express my gratitude to the editors and anonymous reviewers for their hard work and valuable comments on this article.

Author Information

Yao Lu, College of Humanities and Development Studies, China Agricultural University
https://orcid.org/0000-0001-9277-5732

How to cite this article:

Lu, Yao. "A Chinese perspective on artificial intelligence generated content and copyright". *Információs Társadalom* XXIV, no. 2 (2024): 129–142.
https://dx.doi.org/10.22503/inftars.XXIV.2024.2.8

1. Introduction

ChatGPT (Chat Generative Pre-trained Transformer), developed by US company OpenAI, has not only a highly intelligent to realise chat function but also the ability to complete video content recognition, write essays, and create computer code. ChatGPT 4.0 is even able to understand complex legal concepts and has excellent performance in legal logical reasoning (Katz et al. 2024). The rapid development of artificial intelligence (AI) has triggered concerns about AI governance in various countries, and some developed countries have also begun to make legal or institutional arrangements in advance. China, as one of the countries with rapid development of AI technology, is also actively planning AI governance measures. The State Council of China has published a *New Generation AI Development Plan* which states that China plans to establish initial AI laws, regulations, ethical norms and a policy system, and to form AI security assessment and control capabilities by 2025. To achieve that goal, China aims to intensify efforts to combat the misuse of data, infringement of personal privacy, and violation of morality and ethics. These measures indicate that the development of AI has drawn great attention from the Chinese government. In recent years, with the maturity and application of large-scale language modelling (LLM), the application of AI in the field of content generation has become more and more widespread, and the concept of artificial intelligence generated content (AIGC) has gradually entered the public consciousness.

Since AI makes the margin cost of reusing knowledge diminishingly low (Héder 2021), the legal attributes of AIGC and the copyright attribution of AIGC have triggered academic debates among scholars in philosophy, civil law and copyright law. How to correctly apply the law so as to solve the copyright issue of AIGC within the existing legal framework is no longer only a mere academic conceptual discussion but also a practical proposition that needs to be solved urgently. The issue of AI transparency is a fairly complex one (Héder 2020), and this transparency profoundly affects the application of AI. One seems to have no difficulty in recognising that through the technological revolution, the distance between the human imagination and the representation of its objects has widened dramatically (Ursitti 2022). Content produced by AI is gradually moving beyond its traditional application areas, such as text generation, and is increasingly being used in music and film. The transparent nature of AI makes it a huge challenge for humans to define whether a piece of content is a work or not. Previous studies have mostly addressed a particular domain covered by AI generators, e.g., music, film, poetry and painting (Gervais 2020). A literature search was performed in Web of Science using (TS=(AIGC)) AND TS=(copyright) as the search formula, and as of November 2023 there was no article specifically addressing AIGC and copyright as its topic. This suggests that there is still a lack of macro-conceptual discussions on the relationship between AIGC and copyright around the world.

This article seeks to provide a comprehensive and systematic analysis of the literature by Chinese scholars in order to sort out the different perspectives of Chinese scholars on relevant issues, thus enabling scholars around the world to understand the discussions within the Chinese academic community on this issue.

2. Methodology

In order to reach the research objectives of this paper, CiteSpace 6.2.R4 software, developed by Dr. Chaomei Chen, was used as a bibliometric research tool. CiteSpace is a visualisation and analysis software that combines scientometrics and data and information visualisation. It is designed based on the theories of information foraging, detecting frequency bursts and structural variation (Li and Chen 2022) and has been widely used in the fields of text mining and visualisation. By using CiteSpace, we are able to study the research hotspots of AIGC from both macro and micro perspectives.

This paper uses the China National Knowledge Infrastructure (CNKI) as the data source database, which is the most authoritative and representative database of papers, conferences, books and other contents in China. An advanced search was conducted using 人工智能生成物 (AI generated content) or 'AIGC' as the subject term, and the search disciplines were limited to the fields of law, philosophy and publishing, which are closely related to this study. After comparing the results one by one and eliminating the invalid results, a total of 417 Chinese papers were retrieved (retrieval date: November 2023).

Of the 417 studies that met the criteria, the earliest was published in 2017. Therefore, the time span of the study in this paper is from 2017 to November 2023.

Since CNKI data cannot be recognised directly by Citespace software, it was necessary to convert the relevant data. The data from the above 417 papers were converted through CNKI Format Conversion 3.0 and then imported into CiteSpace. In terms of the software set-up, the node types were set to 'Keyword', the time slice was set to 1, and other options used the default settings. The software outputs keyword co-occurrence network mapping, as shown in Figure 1.

Figure 1. Keyword co-occurrence network mapping (own editing)

From the perspective of the content dimension of the related research, the above literature was further analysed through CiteSpace software, and excluding the concept of AI, the keyword co-occurrence network mapping shows that the important nodes of the related research are copyright, originality, rights attribution, copyrightability, legal regulation, legal subject, rights subject, and ethics. From the quantitative perspective of the frequency and degree centre (centrality) values of keyword occurrences, scholars' research in the field of AIGC focuses on AIGC copyright (centrality = 0.18), originality (centrality = 0.13), works/creations (centrality = 0.09), attribution of rights (centrality = 0.08) (Table 1).

Keywords	Centrality
人工智能(AI)	1.03
著作权 (copyright/authorships)	0.18
独创性 (originality)	0.13
作品 (work/creations)	0.09
权利归属 (attribution of rights)	0.08

Table 1. Ranking of keyword centrality (own editing)

In terms of the temporal dimension of the relevant studies, the co-occurrence of keywords in the temporal dimension using CiteSpace was able to further demonstrate the academic research on AIGC in different periods.

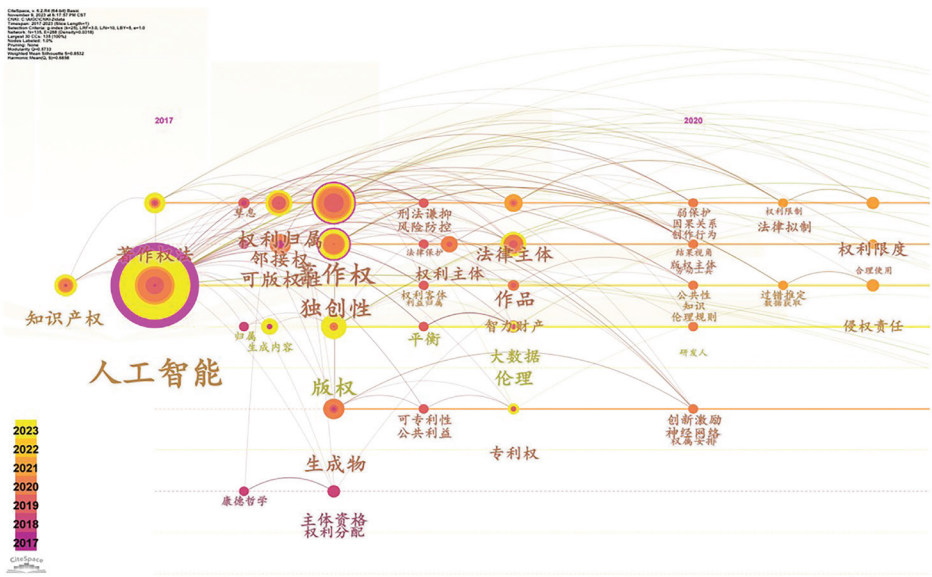

Figure 2. Keyword co-occurrence timeline mapping (own editing)

Figure 2 demonstrates the keyword evolution of AIGC research from 2017 to 2023. From the perspective of horizontal development of concepts, with China's 'three-step strategic goal of AI' proposed in July 2017, legal governance in the field of AI has attracted extensive attention from Chinese scholars. From then on, AI, copyright, and originality became the emergent words in the AIGC research field in 2017, and are still hotspots of research today. Copyright and originality are two concepts that show high relevance. They have become important keywords since 2020.

After analysing the keywords and the specific contents of the collected articles, it was found that the research on AI and copyright in Chinese literature mainly focuses on philosophy, law and ethics. The Chinese scholars' research produces research intersections between the challenges brought by AI to the copyright system, the distribution of responsibility and power, machine learning and the ethics of AI.

Main keywords	Translation in English
人工智能	artificial intelligence
权利归属	attribution of rights
著作权/版权	copyright
独创性	originality
可版权性	copyrightability
法律主体	legal entity

Table 2. Main keywords in timeline mapping (own editing)

In summary, combining the keyword co-occurrence network mapping and the keyword co-occurrence timeline mapping leads to the conclusion that China's research on the copyright of AIGC should focus on the copyrightability and attribution of rights; further analysis of the collected literature shows the research concerns of Chinese academics to be:

1. the copyrightability of AIGC, which is mainly discussed in relation to whether a piece of AIGC is a work due to the standard of China's Copyright Act. Scholars are trying to figure out whether AIGC has the value of protection;
2. the copyright attribution of AIGC, which mainly focuses on whether AI is equal to the concept of 'human being' in civil law or copyright law. The different answers lead to different solutions regarding copyright attribution.

3. Results

3.1. Copyrightability of AIGC

First, the main views that deny the originality of AIGC are:

Template theory: Scholars holding template theory believe that the work is the product of the author's spirit and consciousness, and that content generated by AI

is the result of applying a certain algorithm or template (Wang 2017; Lan 2020). AI is the imitation and upgrading of external human behaviour, and its generator does not undergo the process of injecting the author's thoughts into the work's expression. AIGC is a mere arrangement and combination of the elements of expression, and does not contain expression of human thoughts and emotions, i.e., it does not have the spirit and consciousness, and therefore it has even less of the intellectual creativity required to form a work (Wu 2020).

Selection space method theory: Scholars holding this point of view believe that the originality of a work should be judged from the perspective of 'selection space', i.e., objectively from the author's personalised choice of the breadth of expression for a specific category of work. The creations of AI are unique or limited, with a narrower breadth of expression, making it difficult to claim originality (Yuan 2020) .

Labour tool theory: Scholars who support this point of view mostly believe that AI and human beings have a subordinate relationship. Firstly, when humans are using AI to carry out creative activities, it is difficult for AI, as a tool, to fully convey human creativity, and this defect causes AIGC not to have originality (Miao 2020). Secondly, the technical principles embodied in AIGC are quite different from traditional intellectual labour that relies solely on natural human beings, in which human behaviour is manifested in the design and use of 'AI tools' rather than the creation of works (Wang 2023b).

Contribution measurement theory: Starting from the epistemology of the subject–object unity of 'human-centredness' established by Kantian philosophy, the contribution measurement theory holds that AI is in an auxiliary position in the process of generating AIGC, and that it cannot completely replace and negate the original contribution of human beings to the work. Therefore, human beings should be regarded as the owners of copyrights (Li 2018).

Creative intent theory: Scholars who agree with the creative intent theory believe that intentional labour is the standard for evaluating whether or not there is a productive activity of 'creation', and that only human beings can embody the creative intent in their works, which is also a manifestation of the subjectivity of human beings. The inherent working methods and principles of AI do not have the intent to create and do not have originality (Wang 2023a).

Incentive theory: Incentive theory holds that the legislative purpose of copyright law is to inspire human beings to realise the creation and dissemination of science, literature or art works (Liu 2020). From this point of view, AIGC either does not have the subject to be incentivised or does not have the object content that can be incentivised.

Reversal theory: Scholars holding reversal theory believe that due to the powerful evolutionary ability of AI, legal protection of AIGC will inevitably result in the crushing of human intelligence by AI, thus reversing human creativity (Zeng 2023). In other words, they believe that the recognition of AIGC's originality is a great threat to human creative ability.

Second, the main arguments that support the originality of AIGC are:

Objective theory of originality: This doctrine emphasises that 'human creation' is not a necessary condition for a creation to be protected by copyright law (He and

Zheng 2020; Zheng and Zhang 2021). That means originality can be non-human-origin, and copyright law should focus on the value of the work itself (Sun 2019), as long as some kind of result in form meets the standard of a work that is similar to the results produced by human beings (Yi 2017; Huang and Huang 2019).

Electronic labour theory: This doctrine holds that AIGC is essentially content generated through the 'labour' of AI, which is in line with Locke's basic theory of 'labour creates property' (Feng 2019). Therefore, the labour of AI is kind of original 'electronic labour', so AIGC should be protected.

Neuron theory: This theory believes that the originality of AIGC comes from AI's well-developed and unpredictable neurons. When enough neurons form a chaotic network, AI acquires a unique creative gene and creative ability (Huang 2020).

Free will theory: Scholars supporting this theory (W. Liu 2021) believe that AI has the same free will as human beings in the field of creation, which is reflected in the diversity of results and the infinity of creative options.

Human-machine cooperation theory: Scholars who agree with this view believe that AIGC is the intellectual works produced by human beings working together with AI. The 'contribution measurement theory' should be discarded to recognise the intellectual contribution of AI in the process of intellectual work production (Wu, Zhang and Zhang 2018).

3.2. Copyright attribution of AIGC

In the discussion of the copyright attribution of AIGC, the core argument is whether AI has the status of a civil subject in civil law. That is, 'who' has the right to claim AIGC's copyright. The 'black-box' nature of the operation of AI has led to the diversity of AIGC's right subjects, and the complexity of user behaviour on internet platforms has further magnified the negative impact of the absence of AIGC's right subjects. Intellectual property rights are in fact privileges that inhibit freedom; if we examine the system from an instrumentalist perspective, then privileges must be accompanied by obligations on the part of the privilege holders (Drahos 2017). In this perspective, only by clarifying the subject of rights in AIGC, i.e., who is entitled to copyright or claims copyright in AIGC, can the role played by AI in the creation of AIGC be clarified. Thus, the discussion on rights attribution can be carried out. In terms of the subjects involved in AIGC, they can be divided into human beings (natural persons, legal persons and unincorporated organisations) and AI. Human beings can undoubtedly be the authors of works in accordance with the law. Therefore, the discussion of the subject of the rights of AIGC actually involves the dispute over whether AI has the status of legal subject or legal personality. At present, Chinese academics have not yet reached a unanimous view on this issue, and there are currently three main doctrines as follows:

Legal personality theory: This doctrine believes that AI should be given the same complete civil legal subject status or legal personality as human beings, so that AI can enjoy and assume the rights and obligations corresponding to its behaviour. This doctrine is mainly based on the following reasons: First, AI has already pos-

sessed intelligence, creativity and autonomy similar to or beyond humans, especially the automatic decision-making characteristics of deep learning machines, which increase the unpredictability breaking the causal relationship between humans and the outputs (Zhou 2019). Second, granting AI legal subject status or legal personality is conducive to the protection of AI's legitimate rights and interests. It will accelerate the development of AI (Fan 2022). Third, granting AI legal subject status or legal personality is conducive to solving the problem of tort liability that it may cause (Guo 2018).

Limited personality theory: This doctrine believes that there are differences in rationality and consciousness between AI and human beings. The autonomy, interactivity and deep learning ability of AI determines that it is neither an object nor a human being but an objective existence between human beings and objects. The limited personality theory advocates that AI should be given the status of a limited legal subject or legal personality (Zhang 2019). This doctrine is mainly based on the following reasons: First, AI does not have full civil behavioural capacity and cannot independently participate in civil legal relations (Yuan 2023). Second, the application of AI is always under human control. AI can bear only limited legal responsibility for its consequences, which determines that AI can have only the mimicry of legal personality (Zhang and Yang 2018). Third, granting AI a limited legal subject status or legal personality is in line with the legitimacy of the mimetic subject (Zhang 2022).

Object theory: This doctrine believes that the legal attributes of AI are clear: AI is a technological tool created and utilised by human beings (natural persons, legal persons or unincorporated organisations), and even if it develops to the stage of strong AI, the intelligence of AI is different from that of human beings, and it does not have the ability to comprehend human morality and law (Zhu 2022). AI is merely a tool, and a tool cannot have subject status or legal personality. This doctrine is mainly based on the following reasons: First, AI lacks self-consciousness, free will and a sense of moral responsibility. It cannot exist as an independent individual (Cheng 2022). Second, granting AI subject status or legal personality will lead to the loss of the spirit of the law, and threaten the dignity and safety of human beings (L. Liu 2021). Third, the 'personality' shown by AI is only an appearance of some specific purposeful behaviour, not the subjective capacity itself, as with human beings (Chen and Zhang 2018). Fourth, AI is unable to assume responsibility directly, and granting AI subject status will mean facing not only huge legislative technical challenges but also the possibility that it may harm the current law's relevant institutional arrangements on meaning and tort liability (Wu, Zhang and Zhang 2018). Fifth, AI cannot exercise power, so there is no need to recognise the subject status of robots in the law. Just setting up a kind of work in law, entitled 'robot works', can achieve the purpose of protecting the creators and owners of robots (Wu, Zhang and Zhang 2018).

4. Discussion and future recommendations

Firstly, the discussion of the copyrightability of AIGC in existing studies often neglects the limitations and definition of the concept of AI. It results in there being

no chance to deal with AI through hierarchical or phased treatment. Whether the twelve main doctrines on the copyrightability of AIGC or the three doctrines on the attribution of AIGC, the concept bases of their arguments are not the same. There is a lack of a unified basis for academic dialogue. The discussion on the copyrightability of AIGC is in fact a dispute between author-centrism and work-centrism, and the different choices have given rise to different doctrines. Currently, the prevailing view in Chinese copyright academia is 'author-centrism', which is based on Kant's classical philosophy and the doctrine of personality rights. Author-centrism respects the free will of authors: under the guidance of this idea, the author shall have the right to exercise full control over the work (Lin 2021). The work exerted by the author on the work should be in line with the requirements of the Copyright Law on the originality of the work. The premise implicit in the tools of labour theory, measurement of contribution theory, creative intent theory, incentive theory, electronic labour theory, neuron theory, free will theory and human–machine cooperation theory is author-centrism. That means scholars attempt to figure out whether AI meets the 'author standard' of the Copyright Law for originality, so as to admit or deny the originality of AIGC. On the other hand, template theory, election space method theory, reversal theory and objective theory of originality are based on work-centrism. Scholars try to judge AIGC 'objectively', that is, to examine the originality of the generated content itself. There is less or even no consideration of whether AI is able to become the 'author'. If the originality of the generated content itself is comparable to the originality of the creation, which originated from human beings, then this content can be called 'work' and protected by law. There is still no settled answer on whether copyright law in the age of AI should adopt author-centrism or work-centrism. It is difficult to circumvent the discussion of AIGC's 'authorship' when discussing AIGC's copyrightability. Whether or not to adhere to the current author-centrism in the era of AI? How to adhere to author-centrism in interpreting the existing law? How to adopt work-centrism in the future so as to make appropriate legislative arrangements? These questions need to be answered by future researchers. The different answers to these issues will lead to different standards of originality. In addition, when specifically judging whether a certain AIGC has or does not have originality in the sense of copyright law, it is still necessary to further analyse different types of AIGC. Therefore, the research will be more hierarchical and targeted, which requires further research combining the current practices of the AI industry with the existing laws.

Secondly, for the AI civil subject status or author status of the three doctrines to be reasonable, the logic of the argument also has merit: Legal personality theory portrays a future picture of AI development and actively accepts AI as a new type of civil subject in human beings' law. However, it ignores the reality of the current development of AI, that is, that current AI does not have the independent personality or property to assume responsibility. Limited personality theory and object theory are based on the existing legal provisions. To interpret the perspective of the theory of AI into the 'limited personality' or 'object' category, this point of view has its reality and practicability. However, most of the scholars who hold these doctrines have a negative attitude towards the development of AI, i.e., they deny that AI may

have human-like intelligence in the future. Furthermore, AI ethics lack a reinforcement mechanism (Hagendorff 2020). Object theory ignores AI morality, which may ultimately lead to a failure of AI governance. With the current legislation seemingly focused on privacy and data protection (Lane 2022), coupled with the impossibility of full transparency for many machine learning (ML) and deep learning (DL) systems (Pizzi, Romanoff and Engelhardt 2020), there is a need to place ethical norms (or human rights principles) in the AI law rules. The emergence of ethical principles will contribute to the development of policy and regulatory frameworks (Rességuier and Rodrigues 2022). The White Paper on Artificial Intelligence published by the European Commission in February 2020, and the framework of ethical aspects of AI, robotics and related technologies proposed by the European Parliament in April 2020 were both designed with reference to the ethics guidelines for trustworthy AI. Therefore, we cannot deny the potential of ethics in determining the behaviour of AI. It should be noted that the above three doctrines correspond to different types of AI. Indeed, the core issue is whether the ethical nature of AI can be regulated by a single principle. When AI is developed in different stages, from the perspective of the concept, its connotation and extension will change. From the perspective of AI itself, it will bring different ethical problems due to different levels of intelligence. The attempt to apply the same standards or rules to regulate AI actually ignores the objective differences between different types of AI.

Finally, when Chinese scholars study topics related to AI and copyright, the breadth of comparative research on different countries is still lacking. Currently, Chinese scholars' research involves literature on the United States, the United Kingdom and the Netherlands, but less on Germany, Australia, Italy, Denmark and Canada, all of which have rich practical and theoretical accumulations in the field of AI. According to the *2022 Global Artificial Intelligence Innovation Index Report* released by the China Institute of Scientific and Technological Information, Germany, Canada and Australia are in the second tier of AI innovation evaluation, and Denmark is in the third tier, ranking 4th, 6th, 12th and 13th around the world, respectively. Therefore, it is necessary to pay more attention to the AI practices, theories and legislative trends in the above countries so as to provide material for comparative research on the copyright of AIGC.

5. Limitations

It is important to note that some relevant studies may not have been found in the collection of the literature, which is one of the drawbacks of a systematic literature review.

In this study, in order to ensure the accuracy of the views, only the core set of literature was used in CNKI's database, so some relevant studies were not included in this study because they did not meet the search criteria.

The main references are titled in Chinese. To facilitate readers' understanding, the author has translated the relevant titles into English and appended the original sources to the documents.

References

Chen, Peng, and Li Zhang. 'A Critique of the Theory of Robot "Personality" and the Legal Regulation of Artificial Intelligence Objects'. *Academia* 33, no. 12 (2018): 53–75. https://doi.org/10.3969/j.issn.1002-1698.2018.12.006

Cheng, Chengping. 'On the Autonomy of Artificial Intelligence'. *Journal of Shanghai Jiao Tong University (Philosophy and Social Science Edition)* 30, no. 1 (2022): 43–51. https://doi.org/10.13806/j.cnki.issn1008-7095.2022.01.005

Drahos, Peter. 'Intellectual Property: In Favour of Instrumentality and Against Exclusivity'. In *Philosophy of Intellectual Property Law*, 276–315. Beijing: Commercial Press, 2017.

Fan, Jinxue. 'Artificial Intelligence Legal Subject Theory: Present and Future'. *Political Law Series* 29, no. 3 (2022): 3–17. https://doi.org/10.3969/j.issn.1002-6274.2022.03.003

Feng, Gang. 'A Preliminary Exploration of the Legal Protection Path of AI-Generated Content'. *China Publishing* 42, no. 1 (2019): 5–10. https://doi.org/10.3969/j.issn.1002-4166.2019.01.003

Gervais, Daniel J. 'The Machine as Author'. *Iowa Law Review* 105, no. 5 (2020): 2053–2106. https://scholarship.law.vanderbilt.edu/faculty-publications/1164

Guo, Shaofei. '"Electronic Person" Legal Subject Theory'. *Oriental Jurisprudence* 11, no. 3 (2018): 38–49. https://doi.org/10.19404/j.cnki.dffx.2018.03.005

Hagendorff, Thilo. 'The Ethics of AI Ethics: An Evaluation of Guidelines'. *Minds & Machines* 30, no. 1 (2020): 99–120. https://doi.org/10.1007/s11023-020-09517-8

He, Xuxi, and Yuanmin Zheng. 'Review of the Protection Path of Artificial Intelligence Generated Objects under the Outcome Perspective'. *Technology and Law* 30, no. 3 (2020): 14–21. https://doi.org/10.19685/j.cnki.cn11-2922/n.2020.03.003

Héder, Mihály. 'A Criticism of AI Ethics Guidelines'. *Információs Társadalom* XX, no. 4 (2020): 57–73. https://dx.doi.org/10.22503/inftars.XX.2020.4.5

Héder, Mihály. 'AI and the Resurrection of Technological Determinism'. *Információs Társadalom* XXI, no. 2 (2021): 119–130. https://dx.doi.org/10.22503/inftars.XXI.2021.2.8

Huang, Jie, and Hui Huang. 'Reasonableness of Artificial Intelligence Generated Objects Being Considered as Works for Protection'. *Jiangxi Social Science* 39, no. 2 (2019): 33–42. https://kns.cnki.net/kcms2/article/abstract?v=MuRVhOLgpmvPt3abL_HqaiBNT_ZvZBryRT9N0Rk5BF6hXpBRqDQg089_lp9vZJkfWAquLB9-WYJis8Fh0auMjXe0zW_GeNQuy8xb5Z2PDmqPLqpdyL2pa0265fm2KRoR77I38SKhbSk

Huang, Shanshan. 'On the Impact and Response of Artificial Intelligence on the Copyright System'. *Journal of Chongqing University (Social Science Edition)* 26, no. 1 (2020): 159–169. https://doi.org/10.11835/j.issn.1008-5831.fx.2019.07.005

Katz, Daniel Martin, Michael James Bommarito, Shang Gao, and Pablo Arredondo. 'GPT-4 Passes the Bar Exam'. *Philosophical Transactions of the Royal Society A* 382 (2024). http://dx.doi.org/10.2139/ssrn.4389233

Lan, Chunjie. '"From Generated Content to Underlying Data – New Directions for Legal Protection of Artificial Intelligence'. *Technology and Law* 30, no. 3 (2020): 22–31. https://doi.org/10.19685/j.cnki.cn11-2922/n.2020.03.004

Lane, Lottie. 'Clarifying Human Rights Standards Through Artificial Intelligence Initiatives'. *International and Comparative Law Quarterly* 71, no. 4 (2022): 915–44. https://doi.org/10.1017/S0020589322000380

Li, Jie, and Chaomei Chen. 'General Description of the Citespace'. In *Citespace: Text Mining and Visualization in Scientific Literature (Third Edition)*, 1–34. Beijing: Capital University of Economics and Business Press, 2022.

Li, Xiaoyu. 'Ruminations on the Copyrightability and Rights Allocation of Artificial Intelligence Generated Objects'. *Electronic Intellectual Property* 22, no. 6 (2018): 31–43. https://doi.org/10.3969/j.issn.1004-9517.2018.06.004

Lin, Xiuqin. 'Reshaping the Fair Use System in Copyright Law in the AI Era'. *Legal Studies* 43, no. 6 (2021): 170–185. www.nssd.cn/html/1/156/159/index.html?lngId=FXYJ2021006010

Liu, Lianjun. 'Jurisprudential Reflection on the Legal Subject Theory of Artificial Intelligence'. *Modern Law* 43, no. 4 (2021): 73–88. https://doi.org/10.3969/j.issn.1001-2397.2021.04.07

Liu, Wei. 'Challenges and Responses to Copyright Law in the Age of Artificial Intelligence'. *Journal of Shanghai Jiao Tong University (Philosophy and Social Science Edition)* 29, no. 2 (2021): 40–49. https://doi.org/10.13806/j.cnki.issn1008-7095.2021.02.006

Liu, Yinliang. 'On the Copyright Law Status of Artificial Intelligence Works'. *Politics and Law* 39, no. 3 (2020): 2–13. https://doi.org/10.15984/j.cnki.1005-9512.2020.03.001

Miao, Chenglin. 'The Denial of the Originality of Artificial Intelligence Generatives in the Perspective of Marx's Labour Theory'. *Editorial Friend* 40, no. 5 (2020): 87–95. https://doi.org/10.13786/j.cnki.cn14-1066/g2.2020.5.014

Pizzi, Michael, Mila Romanoff, and Tim Engelhardt. 'AI for Humanitarian Action: Human Rights and Ethics'. *International Review of the Red Cross* 102, no. 913 (2020): 145–80. https://doi.org/10.1017/S1816383121000011

Rességuier, Anaïs, and Rowena Rodrigues. 'AI Ethics Should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics'. *Big Data & Society* 7, no. 2 (2020): 1–5. https://doi.org/10.1177/2053951720942541

Sun, Song. 'Work Characterisation and Institutional Response to AI-Created Content'. *Science-Technology & Publication* 27, no. 4 (2019): 106–110. https://doi.org/10.16510/j.cnki.kjycb.2019.04.034

Ursitti, Filippo. 'The Techne as Producer of Outdated Humans'. *Információs Társadalom* XXII, no.2 (2022): 117–127. https://dx.doi.org/10.22503/inftars.XXII.2022.2.7

Wang, Guozhu. 'Humanistic Logic in the Determination of Copyrightability of Artificial Intelligence Generated Objects'. *Journal of East China Normal University (Philosophy and Social Science Edition)* 55, no. 1 (2023a): 133–142. https://doi.org/10.16382/j.cnki.1000-5579.2023.01.013

Wang, Guozhu. 'Reinterpretation of the Translation Right System under the Artificial Intelligence Machine Translation Scenario'. *Journal of Politics and Law* 30, no. 3 (2023b): 79–89.
https://doi.org/10.3969/j.issn.1002-6274.2023.03.007

Wang, Qian. 'On the Characterisation of AI-Generated Content in Copyright Law'. *Legal Science (Journal of Northwestern University of Politics and Law)* 35, no. 5 (2017): 148–155.
https://doi.org/10.16290/j.cnki.1674-5205.2017.05.014

Wu, Handong, Ping Zhang, and Xiaojin Zhang. 'The Challenge of Artificial Intelligence to the Legal Protection of Intellectual Property Rights'. *China Law Review* 20, no. 2 (2018): 1–24.
https://www.nssd.cn/html/1/156/159/index.html?lngId=675159213

Wu, Wudong. 'Divergence between AI-Generated Content and Traditional Copyright System." *Shandong Social Science* 7 (2020): 36–42.
https://doi.org/10.14112/j.cnki.37-1053/c.2020.07.007

Yi, Jiming. 'Are Artificial Intelligence Creations Works?'. *Legal Science (Journal of Northwestern University of Politics and Law)* 35, no. 5 (2017): 137–147.
https://doi.org/10.16290/j.cnki.1674-5205.2017.05.013

Yuan, Feng. 'On the Judgement of Originality of Works under the Environment of New Technology – With the Perspective of "Choice Space Method"'. *Journal of Huazhong University of Science and Technology (Social Science Edition)* 34, no. 5 (2020): 94–102.
https://doi.org/10.19648/j.cnki.jhustss1980.2020.05.11

Yuan, Zeng. 'Research on the Liability Capacity of Generative Artificial Intelligence'. *Oriental Jurisprudence* 16, no. 3 (2023): 18–33.
https://doi.org/10.19404/j.cnki.dffx.20230505.002

Zeng, Bailing. 'Copyright Without Authorship: The ChatGPT's Rhetorical Question About People'. *News and Writing* 40, no. 9 (2023): 69–79.
https://kns.cnki.net/kcms2/article/abstract?v=nouGVBS_tgftrb_mdiqwygF5vMNNVZiVhgB1Cq3OGrTiHtMWys-HsVubB03UfSyG6itcSgNSoJ408KYW58I2nja_kxc3tKlsvCH6e0d5w1zNJBa-bMsmBrcDii1h5bQQTFLJRhGB6f4

Zhang, Lei, and Qingwang Yang. 'On the Mimetic Legal Personality of Artificial Intelligence'. *Journal of Hunan University of Science and Technology (Social Science Edition)* 21, no. 6 (2018): 91–97.
https://doi.org/10.13582/j.cnki.1672-7835.2018.06.013

Zhang, Li. 'Exploring the Rationale for Legalisation of Blockchain and AI Organizations'. *Oriental Law Journal* 15, no. 5 (2022): 45–59.
https://doi.org/10.19404/j.cnki.dffx.20220824.003

Zhang, Zhijian. 'On the Electronic Legal Person Status of Artificial Intelligence'. *Modern Law* 41, no. 5 (2019): 75–88.
https://doi.org/10.3969/j.issn.1001-2397.2019.05.06

Zheng, Peng, and Xiaoping Zhang. 'On the Dilution of the Natural Person Source of Originality of AI Creative Objects'. *Journal of Dalian University of Technology (Social Science Edition)* 42, no. 6 (2021): 106–113.
https://doi.org/10.19525/j.issn1008-407x.2021.06.011

Zhou, Xiang. 'Advocacy of the "Rights Subject Theory" of Intelligent Robots'. *Jurisprudence* 42, no. 10 (2019): 3–17.

https://kns.cnki.net/kcms2/article/abstract?v=nouGVBS_tgf1T9xKFGX_WXkz-YpWCpbDyLm7fsBA4U7xAk3de4PMMObQTY6w9Dx0IykEWV6qAS1G_iLsoPLveHWW6Rg_EJhvFmRJNeAmEsT6n_UZsr0XtoeHA0v9P1QS92ZqxPYjgoY

Zhu, Chengbin. 'Research on Strong Artificial Intelligence as Special Objects'. *Journal of Beijing Institute of Technology (Social Science Edition)* 22, no. 6 (2022): 113–122. https://doi.org/doi:10.12120/bjutskxb202206113

A Study of Willingness to Adopt Smart Aging Services: Evidence from Anhui Province, China

Exploring the adoption intention of smart senior care services can help improve the actual adoption rate. Taking a questionnaire survey of 1600 households in Anhui Province as an example, structural equation modeling was used to study the adoption intention and influencing factors of smart senior care services. The results show that subjective norms, perceived usefulness and perceived ease of use are the key factors influencing the adoption intention of smart senior care services. Further, age has an impact on perceived usefulness and adoption intention, education has an impact on perceived ease of use and adoption intention, and monthly household income has an impact on perceived usefulness and adoption intention. This paper broadens the existing technology acceptance model (TAM) theory and provides some basis for the development of economics. The findings will be beneficial to the government in formulating more precise policies on elderly services to properly address the challenges of aging.

Keywords: *smart aging services, willingness to adopt, structural equation modeling, China, elderly*

Declaration of competing interest

The authors declare that they have no known competing financial or other interests that could have compromised the work reported in this paper.

Acknowledgment

This study was supported by the General Project of Philosophy and Social Science Research in Colleges and Universities of Jiangsu Province[No.2023SJYB0423], Nanjing Institute of Technology Research Fund Project "Adoption of Smart Elderly Service and Its Policy Incentive Mechanism under the Threshold of Aging" [No.YKJ202343].

Author information

Xing'an Yao, Nanjing University, School of Social and Behavioral Sciences; Nanjing Institute of Technology, School of Marxism, https://orcid.org/0000-0001-9601-4455

Mengjun Zhu (Corresponding author), UCSI University, Wellbeing Research Centre, Faculty of Social Sciences and Liberal Arts, https://orcid.org/0000-0002-6129-1628

How to cite this article:

Yao, Xing'an, Mengjun Zhu. "A Study of Willingness to Adopt Smart Aging Services: Evidence from Anhui Province, China". *Információs Társadalom* XXIV, no. 2 (2024): 143–163.

https://dx.doi.org/10.22503/inftars.XXIV.2024.2.9

1. Introduction

With the rapid development and continuous progress of society, population aging has gradually become an important problem that needs to be solved for all human beings. By 2020, China's elderly population aged 60 and above will reach 255 million, accounting for 17.8% of the total population, and this proportion is still increasing; the degree of aging will exceed that of most developed countries by the middle of the 21st century (National Health Commission 2017). In the new era, promoting the comprehensive and healthy development of home care services is an important measure to solve the problem of constructing the current urban and rural elderly service system in China (Yi 2021).

The rapid development of information science and technology in recent years has led to the flourishing of "Internet+" home care services. The traditional community home care service has defects such as dysfunction of supply and demand, low efficiency of the service, etc. Smart senior care service can solve not only these problems but the aging problem, too. In 2017, the Ministry of Industry and Information Technology and the Ministry of Civil Affairs issued the "Action Plan for the Development of Intelligent Healthy Aging Industry (2017–2020)," proposing to build out 500 intelligent healthy aging demonstration communities in five years, greatly increasing the speed of intelligent aging development. The introduction of Internet technology into traditional home care services has enabled the integration of information technology, artificial intelligence, and other high technologies with home care services to provide personalized and efficient intelligent elderly care services (Xu and Zhang 2021). The Life Trust of the United Kingdom first introduced the concept of smart aging, which is a new intelligent aging model that can help the elderly enjoy a high-quality and efficient home aging life without the constraints of time and space. Since then, different scholars have improved the definition of the concept; based on the existing research, this study considers that intelligent senior care service is a kind of senior care service that takes the home life of the elderly as the basic starting point, and provides an efficient, fast, low-cost and intelligent senior care service with the help of an external sensor information system and an internet data structure platform. It is also divided into seven types of service: meal assistance, cleaning assistance, walking assistance, bathing assistance, medical assistance, shopping assistance, and remote care (Liao and Chen 2019; Van der Kloet 2019; Goundrey-Smith 2019).

By the end of 2020, the number of elderly people aged 60 and above in the province exceeded 11.7 million, accounting for 18.41% of the total population of the province, and the degree of aging is at the forefront of the country (Anhui Provincial Bureau of Statistics 2021). However, the overall economic development level of the province is not high, the intelligent senior care service industry is still in the period of development and exploration, the stable market demand has not yet formed, the operation regulation and development structure is still not perfect, and the gap between supply of and demand for home senior care service is prominent. Exploring the adoption behavior of smart senior care services and promoting the adoption rate will help solve not only the aging problem but also the imbalance between sup-

ply and demand. However, so far, the adoption rate of the elderly is still not high; so how can the service penetration and adoption rate be improved? The factors influencing adoption intention can help to improve the adoption rate, so early clarification of the adoption intention and influencing factors has become an important challenge to be solved. The findings of this paper can help strengthen the policy relevance and effectiveness of government subsidy policies for the promotion of smart senior care services and achieve the development goal of smart living.

2. Literature review

Many scholars have studied the willingness and constraints of the elderly to use intelligent products on different levels, which can be roughly divided into the following three categories: First, the influence of technology itself is a factor. Older people's concern about healthcare services that use wireless sensors depends on the level of system independence (Steele et al. 2009). Seniors are more satisfied with medical consultations of a service nature than with physical disease treatment alone (Chae et al. 2001). Evaluation of the robot's aging and acceptance by the elderly can clarify its ease of use, and satisfaction with use is directly related to ease of use (Körtner et al. 2014). Second is the influence of psychological and physical factors. It has been found that weaker health levels and higher awareness of novelty among the elderly increase their acceptance of and satisfaction with high-tech elderly products (Lu, Zhou and Wang 2009). Some scholars have also used empirical research to further verify the important role of social-psychological characteristics in determining elderly use of information technology social network systems (Godfrey and Johnson 2009). Third is the influence of economic- and education-level factors. Older adults have different attitudes toward using the Internet, and individual income levels directly determine the degree of willingness to shop online (Eastman and Iyer 2004). Older adults' age and education level affects their adoption of information systems, but continued use depends on differences in user beliefs and perceptions before and after use (Chakraborty, Vishik and Rao 2013). Besides, other scholars have explored issues such as the elderly's post-use perceptions of the new technologies. For instance, Internet of Things (IoT)-based elderly care services have advantages such as multidirectional sensing and powerful transmission (Xi, Ren and Zhai 2014); smart elderly services have advantages such as efficiency, speed, and convenience compared to traditional elderly services (Gu, Wu and Cao 2017); using smart elderly technology can eliminate the problem of spatial and temporal barriers to senior care services and improve service continuity (Zhu 2016); and the higher the health level, the lower the willingness to demand smart senior care, while the stronger the cognitive ability, the higher the willingness to demand it (Yu and Sun 2017; Davis, Bagozzi and Warshaw 1989).

There have been quite a lot of studies, but they mostly focus on the external level such as policy system and resource endowment, and there are limited studies that examine the problem from the internal level in relation to the elderly psychological cognitive perspective, and the research methods are still focused on traditional

models such as Logistic and Probit, which have obvious shortcomings in clarifying the internal relationship between factors and elements and specific influence paths (Jian, Dai and Dai 2016). In contrast, structural equation models not only depict the influence relationships among latent variables but also sort out the path relationships among latent variables (Dufhues et al. 2021). Besides, previous studies have not tested the robustness of the model, which cannot further improve the scientific and accuracy of the conclusions (Huang and Liu 2020).

3. Theoretical foundation and research hypothesis

3.1. Theoretical foundation

Davis first proposed the technology acceptance model in the 1990s, which is a model to explain and predict the degree and behavior of users' acceptance of information technology (Davis, Bagozzi and Warshaw 1989). This model clarifies that behavioral intention determines end-user behavior, while perceived usefulness (PU) and perceived ease of use (PEU) jointly influence behavioral intention. Subsequently, many scholars refined this model and proposed TAM2 and TAM3 models (Venkatesh and Davis 2000; Venkatesh and Bala 2008), in which TAM3 further demonstrates that behavioral intention is influenced by PU, PEU, and community influence, and directly affects subject behavior. PU and PEU are influenced by individual differences, convenience, system characteristics, and community influence. This model has been widely used in the field of information intelligence technology at the level of technology behavior acceptance, which better shows what factors influence the adoption behavior of information intelligence technology, but there is limited research related to its application to the adoption intention in the field of smart elderly services (Figure 1).

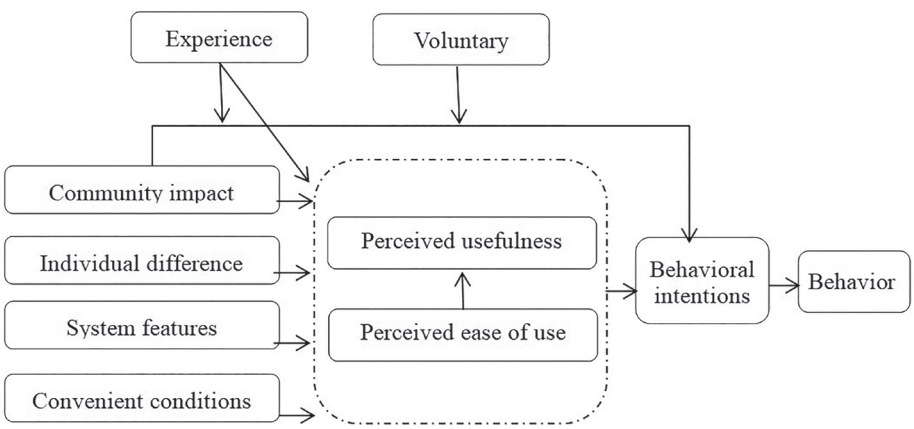

Figure 1. Technology acceptance model (TAM3) (Venkatesh and Bala 2008)

3.2. Research hypothesis

The external validity associated with the adoption of smart senior care services is called PU. Using adoption technology theory to explore the issue of factors influencing social networking sites, it has been shown that PU and PEU influence the elderly's willingness to adopt (Ji et al. 2010). The following hypotheses are proposed because the actual adoption effect is of greater concern when accepting smart senior care services, and seniors will choose to use them only if they perceive that the adoption brings higher effects than before.

H1: Perceived usefulness has a significant positive effect on the willingness to adopt smart senior care services.

The degree of mastery needed to operate smart senior care services easily is called PEU. Older adults' sense of self-enablement indirectly influences adoption intentions by affecting PU and PEU of social networking sites (Braun 2013). Older people are less able to accept new technologies, and the less effort it takes to use smart devices, the more willing they are to adopt them. The following hypothesis is made.

H2: Perceived ease of use positively influences both adoption intention and perceived usefulness.

The external environmental factors that constrain the adoption of smart senior care services are called subjective norms. Because of the risk of uncertainty in the use of smart senior care services, most elderly people do not have a clear opinion before deciding to adopt them and seek to consult their children, friends, and neighbors. If people around them have a positive attitude toward using the services, it will significantly increase the elderly people's willingness to adopt. The following hypothesis is made.

H3: Subjective norms not only positively influence adoption intentions but also positively contribute to perceived usefulness and perceived ease of use.

The help of smart senior care service promoters in adopting the service is called external facilitation. Studies have shown that external social support forces and internal technical capabilities have an indirect effect on PU and PEU (Sintonen and Immonen 2013). Identifying good external conditions can drive up the level of willingness to adopt. The following hypothesis is made.

H4: Convenience not only affects perceived usefulness but also affects perceived ease of use.

The degree of improvement in life achieved by adopting smart senior care services is called results display. The use of smart senior care services can promote the ability of senior care services (Yu and Sun 2017; Tian 2015). Results display includes direct results display, which indicates a direct improvement in the physical and mental health of the elderly, and indirect results display, which indicates an indirect improvement in physical and mental health by changing the life satisfaction of the elderly. The following hypothesis is made.

H5: Results display positively promotes both perceived usefulness and perceived ease of use, as well as adoption intentions.

The expectation theory is expressed by the formula: Expectation × Valence = Excitement power, which indicates that the product of expectation and valence deter-

mines the magnitude of human motivation, and the greater the degree of grasping the goal, the greater the original power motivation will be stimulated and the motivation will be significantly increased. That is, the greater the benefits of adopting smart senior care services for the elderly, and the closer it is to the user's expectation, the more it will increase the adopter's willingness to use. The following hypothesis is proposed.

H6: Performance expectations significantly affect perceived usefulness, perceived ease of use, and willingness to adopt.

Based on the above research hypotheses and the specific TAM, the following theoretical analysis framework (Figure 2) is proposed concerning the actual problem.

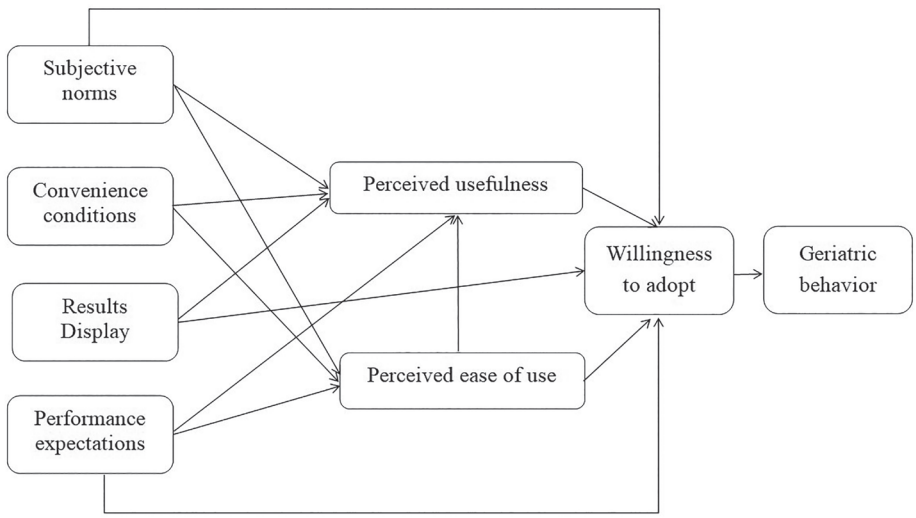

Figure 2. Model of factors influencing the willingness to adopt smart senior care services (own editing)

4. Materials and research methods

4.1. Questionnaire design and data sources

Combining the technology acceptance model and existing studies, seven latent variables were set: willingness to adopt, PEU, PU, presentation of results, performance expectations, convenience, and subjective norms. The indicator structures of willingness to adopt, PU, and PEU were set by drawing on the research scales of Ji and Braun (Ji et al. 2010; Braun 2013); the indicator structure of subjective norms was set by drawing on the research of Qunlin Zhang (Zhang and Atani 2019); the indicator structures of convenience and outcome were set by drawing on the research of Sintonen and Jie Tian et al.'s demonstration indicator structure (Sintonen and Immone 2013; Tian 2015); and the indicator structure of performance expectations was set with the help of Lei Wang et al.'s study (Wang et al. 2020).

The data in this paper were obtained from a survey conducted by three universities, including Shanghai University and Nanjing Agricultural University, on the precise supply of smart senior care services, and the research team conducted household surveys from February to October 2019 among elderly people aged 60 years and above in the target communities in three waves. Using the typical sampling and random sampling survey methods, the province was divided into six cities from north to south, including Huaibei and Suzhou, according to the socioeconomic development and taking into account the geographical and spatial distribution characteristics. Each city correspondingly selected 2~3 districts (counties), and each district (county) selected 3~4 representative communities with a wide coverage of traditional elderly services, and randomly selected elderly aged 60 and above in the community with medium distance sampling for the survey. A total of 1600 questionnaires were distributed. After excluding the questionnaires that were not successfully collected, a total of 1556 valid questionnaires were obtained, with an effective rate of 97%. There were 829 males and 727 females in the sample group, and the proportions of the two were 53.3% and 46.7%, respectively, which were roughly equal. The majority of the elderly were aged 65–80 years old, accounting for 70.6% of the total number, and the majority of the elderly were educated in primary and junior high school, while the proportion of the elderly in high school and above was only 9.3%. Only 685 elderly people adopted the smart elderly service; the adoption rate was thus only 44%.

4.2. Variable description

As shown in Table 1, this paper establishes a structural equation model (SEM) for the adoption of smart elderly services by the elderly, in which willingness to adopt, PU, and PEU are endogenous latent variables, and performance expectations, subjective norms, outcome presentation, and convenience are exogenous latent variables. The observed latent variables were measured using a 5-point Likert scale, with levels 1~5 indicating "totally disagree," "disagree," "generally agree," "agree," and "strongly agree." The specific index settings and variables are shown in Table 1, because the smart elderly service includes seven types of service: walking aid, bathing aid, medical aid, meal aid, etc. If any one or more of these services are used, it is considered as using this service.

4.3. Research methodology

SEM, as a multivariate statistical analysis method, is often used to deal with complex relationships between individual latent variables that cannot be directly observed (Zhang, Zhang and Wang 2018; He, Bai and Zhu 2019). It is widely used in questionnaire data processing, and the biggest advantage is that it provides observable treatment of latent variables that are difficult to observe clearly (Bentler and Chou 1987).

Specifically, it includes two models, structural equation and measurement equation, and the specific model equation is as follows:

$$\eta = B\eta + \Gamma\xi + \varsigma \quad (1)$$

$$X = \Lambda_X \xi + \delta \quad Y = \Lambda_Y \eta + \varepsilon \quad (2)$$

Equation (1) is the structural equation model describing the relationship between latent variables; η is the endogenous latent variable, ξ is the exogenous latent variable, ς is the residual term, and B and Γ are the path coefficient. Equation (2) is a measurement equation model describing the relationship between latent and observed variables; X is the observed variable of the exogenous latent variable ξ, Y is the observed variable of the endogenous latent variable η, Λ_X and Λ_Y denote the matrix of observed variables to latent variables ξ and η with factor loadings,and δ and ε denote the error terms of the exogenous and endogenous variables, respectively.

Variable		Variable description	Mean	Standard deviation
Willingness to adopt (WA)	WA_1	I would consider using smart senior care services when conditions permit.	3.72	0.891
	WA_2	I will keep an eye on the development of smart senior care services.	3.43	0.664
	WA_3	I would recommend the use of smart senior care services to my friends and family.	3.60	1.032
Perceived usefulness (PU)	PU_1	Smart senior care service can reduce the cost of senior care service.	2.47	0.455
	PU_2	Smart senior care service can improve the efficiency of senior care service.	3.11	0.504
	PU_3	Smart senior care service can promote further improvement of the social security system.	3.66	1.242
	PU_4	Smart senior care service can improve the quality of life.	3.75	0.900
Perceived ease of use (PEU)	PEU_1	I think smart senior care technology is easy to master.	3.20	0.451
	PEU_2	I can easily master the use of smart devices through simple training.	3.63	1.106

	PEU_3	Through instruction, I can clearly understand the principles of using smart devices.	2.77	1.290
Subjective norms (SN)	SN_1	Whether or not to use smart senior care services is influenced by children.	2.88	0.479
	SN_2	Whether to use smart senior care service will be influenced by friends and relatives.	3.06	0.874
	SN_3	Whether to use smart senior care service will be influenced by government staff.	3.41	0.635
Convenience conditions (CC)	CC_1	The government has corresponding subsidies for using smart senior care services.	2.70	1.344
	CC_2	The government and community have specialized personnel to provide door-to-door guidance services.	3.22	0.772
	CC_3	I have the financial conditions to adopt smart senior care services.	2.83	1.046
Results display (RD)	RD_1	I can visually find the benefits of smart senior care services.	3.31	0.765
	RD_2	I have heard from friends and relatives that smart senior care services work well.	3.45	1.309
	RD_3	Through media reports, I know that smart senior care service is a very scientific way of providing senior care service.	2.94	0.608
Performance expectations (PE)	PE_1	I hope the cost required for smart senior care service can be reduced appropriately.	3.07	1.299
	PE_2	I hope it can improve my physical health.	3.16	1.057
	PE_3	I hope to improve my mental health.	2.67	0.452

Table 1. Descriptive statistics of measurement indicators and question items (own editing)

5. Results

5.1. Reliability and validity tests

To ensure the reliability and validity of the questionnaire information, the latent variables and the validity of the questionnaire need to be tested. In this study, Cronbach's alpha value was used to test the internal consistency of the measurement factors, and a Cronbach's alpha value greater than 0.7 indicates good validity of the measurement factors; greater than 0.8 indicates very good reliability (Lu and Guo 2019). The seven latent variables were analyzed using SPSS22.0 software, and the results are shown in Table 2. From the results, it can be seen that Cronbach's α values are all above 0.7, which indicates that the questionnaire has high reliability. Meanwhile, the overall validity of the questionnaire was analyzed, generally using the Kaiser-Meyer-Olkin (KMO) test and Bartlett's spherical test. The results showed that the KMO test values for each latent variable ranged from 0.610 to 0.715 and Bartlett's spherical test values were placed at the 1% level of significance, making them suitable for factor analysis.

5.2. Overall model fitness test

The overall model was evaluated using AMOS 22.0 software and the model parameters were estimated using the great likelihood method to identify the goodness-of-fit of the model through the absolute fitness index, the value-added fitness index, and the parsimonious fitness index. Table 3 shows that the original model fit index is good, and the new path trajectory is obtained after the original model is revised several times and the least significant three paths are removed. Also, the new revised results were obtained by adding the new path of RD → PEU (Table 3). The modified model fitness index has improved significantly compared with the previous one, and the model itself has improved significantly in terms of fit. Although the Adjusted Goodness of Fit Index (AGFI) and Normed Fit Index (NFI) values are not too high, they are significantly greater than 0.8, and most of the fit index values are within the normal range. Thus, the overall model fit is good (Zhang and Lian 2019).

Variables	Measurement question items	Cronbach's alpha coefficient	Commonality	KMO	Bartlett's sphericity test (significance)
Willingness to adopt (WA)	WA1	0.782	0.765	0.711	289.311 (P=0.000)
	WA2		0.749		
	WA3		0.800		
Perceived usefulness (PU)	PU1	0.725	0.717	0.607	246.780 (P=0.000)
	PU2		0.705		
	PU3		0.736		
	PU4		0.683		

Perceived ease of use (PEU)	PEU1	0.805	0.746	0.746	312.078 (P=0.000)
	PEU2		0.772		
	PEU3		0.813		
Subjective norms (SN)	SN1	0.763	0.745	0.648	270.655 (P=0.000)
	SN2		0.759		
	SN3		0.782		
Convenience conditions (CC)	CC1	0.824	0.819	0.627	304.560 (P=0.000)
	CC2		0.746		
	CC3		0.755		
Results display (RD)	RD1	0.720	0.688	0.702	312.568 (P=0.000)
	RD2		0.705		
	RD3		0.719		
Performance expectations (PE)	PE1	0.716	0.730	0.653	268.041 (P=0.000)
	PE2		0.722		
	PE3		0.654		

Table 2. Reliability and validity tests (own editing)

Fitting index	Specific index	Reference value	Initial model fitted values	Modified model fit values	Test results
Value-added suitability index	CMIN/DF	1~3	1.667	1.258	Ideal
	GFI	>0.90	0.945	0.927	Ideal
	AGFI	>0.90	0.856	0.871	Acceptable
	RMSEA	<0.08	0.058	0.046	Ideal
	NFI	>0.90	0.865	0.893	Acceptable
	IFI	>0.90	0.912	0.955	Ideal
	TLI	>0.90	0.923	0.947	Ideal
Minimalist suitability index	CFI	>0.90	0.906	0.940	Ideal
	PNFI	>0.50	0.610	0.711	Ideal
	PCFI	>0.50	0.540	0.607	Ideal
	PGFI	>0.50	0.537	0.558	Ideal

Table 3. SEM model fitness index results (own editing)

5.3. Analysis of SEM results

The new test results were obtained after the revised model was evaluated again (Figure 3 and Table 4), and it is clear from the standardized coefficients that PU, PEU, and SN can produce a 1% significance test on adoption intention. In addition, the

ranking of each influence is: PEU (0.512)>PU (0.406)>SN (0.278), and we also find that the effects of RD, SN, and CC on PEU are significant at the 1% level, PE and PEU are significant at the 5% level, PE, RD, etc. PU is significant at the 1% level, and SN and PU are significant at the 5% level. As many as 10 of the 13 different path trajectories set up above have been validated and are explained as follows.

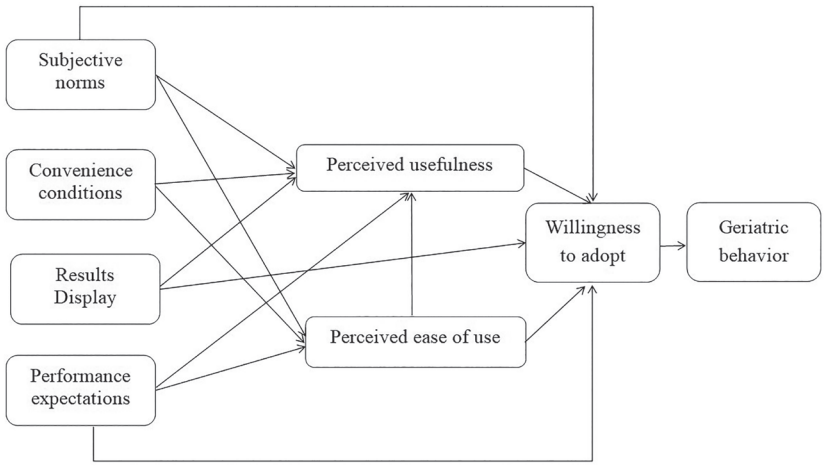

Figure 3. Adopting the Smart Aging Service Willingness Model Pathway Map (own editing)

Paths	Standardized path coefficient	Standard error	Is it significant ?
WA ← PU	0.336***	0.229	Significant
WA ← PEU	0.430***	0.288	Significant
WA ← SN	0.511***	0.188	Significant
PEU ← SN	0.279***	0.167	Significant
PEU ← PE	0.472**	0.230	Significant
PEU ← RD	0.340***	0.156	Significant
PU ← CC	0.110***	0.086	Significant
PU ← RD	0.455***	0.182	Significant
PU ← PEU	0.564**	0.285	Significant
PU ← PE	0.104***	0.054	Significant
PU ← SN	0.436**	0.146	Significant

Note: *, **, *** denote $p<0.1$, $p<0.05$, and $p<0.01$, respectively.

Table 4. Estimation results of the modified SEM (own editing)

(1) PU has a positive effect on the elderly's willingness to adopt smart senior care services, which is significant at the 1% level. The standardized coefficient is 0.336, and hypothesis H1 is verified. In recent years, smart elderly care has been continuously paid attention to by the government, and the documents "Guidance on Actively Promoting 'Internet+' Action" and "Action Plan for the Development of Smart Health Elderly Industry (2017–2020)" pointed out the direction for the future of smart elderly care and improved its development standardization and feasibility. More elderly users perceive the great advantages of smart senior care services in improving their quality of life, promoting the efficiency of senior care services, and reducing the burden of senior care on family members, which enhances the sense of identity of smart senior care services. Thus, the level of PU is improved and the willingness to adopt the services is increased.

(2) PEU has a significant positive effect on both PU and the willingness of the elderly to adopt smart senior care services, with standardized coefficients of 0.564 and 0.430, respectively, and hypothesis H2 is verified. The recognition of smart senior care services is high and maintains high coverage ratios of adoption (0.610), understanding (0.563), and mastery (0.745). Emerging smart senior care is highly professional and technical, which limits the degree of PEU improvement and plays a restraining role in further promotion in the later stage. Increasing the guidance of community senior care service personnel in households and simplifying the operation steps are the key measures to solve this dilemma.

(3) Subjective norms play a facilitating role in both the willingness of elderly people to adopt smart elderly services and PEU, which are significant at a 1% level. And it played a 5% positive effect on PU with standardized coefficients of 0.511, 0.279, and 0.436, respectively, and hypothesis H3 is verified. Meanwhile, the coefficients of each sub-option of subjective norms are SN1 (0. 66), SN2 (0. 68), and SN3 (0. 63), indicating that subjective norms have a stronger influence on the willingness of the elderly, and the degree of influence of each sub-variable is the same. The supportive behaviors of community elderly service organization personnel and government department cadres for smart elderly service can significantly increase the willingness to adopt smart elderly service.

(4) Convenience plays a positive contribution related to the PU of smart senior care services and is significant at the 1% level with a standardized coefficient of 0.110, and hypothesis H4 partially holds. The availability of good adoption conditions, the availability of sufficient government subsidies, and the ability to ensure that technical personnel enter the home to provide guidance are key factors limiting the adoption of smart senior care services and are also important factors affecting the PU.

(5) Both PEU and PU were positively promoted by the results display, which remained significant at the 1% level with standardized coefficients of 0.340 and 0.455, respectively, and hypothesis H5 partially holds. For the elderly, the demonstration of expected usage effects can significantly affect PEU and PU, but at the same time, the study verifies that the demonstration of results does not directly affect the willingness to adopt smart elderly services, which is inconsistent with the above hypothesis.

(6) Performance expectation positively affects PEU at the 5% level and PU at the 1% level. The standardized coefficients of the former and the latter are 0.472 and 0.104, respectively, according to which hypothesis H6 is partially valid. Combined with the validation results, it appears that the elderly have higher expectations, have a positive attitude toward this service, and psychologically perceive that the use will bring beneficial effects, although there is no way to know exactly how valuable the service is, due to the strong endogenous forces that motivate the elderly to have high expectations that the smart elderly service will be a great help in their lives, and, in terms of behavioral perception, to firmly believe that the complex operational technology becomes easier to use.

5.4. Robustness testing

In this paper, the robustness test is conducted by adding new variables to the original structural equation model by introducing a new covariate "individual characteristics of the elderly"(ICE) and treating it as a control variable. The new covariate includes three observable variables: age, education level, and monthly household income. After the introduction of the control variables, the significance and regression coefficients of the before and after models were compared to test whether the latter model reached the level of robustness (Table 5).

From the regression results of the structural model in Table 5, the effects of PU and PEU on the willingness to adopt are still significant at the 1% level when the new ICE covariate is added. This indicates that the direction and the significance level of the regression coefficients of the structural equation model are consistent with the original equation model after adding the ICE covariate. The addition of the ICE covariate has a strong and significant effect on adoption intention, and the coefficient increases by 1.3 percentage points compared with the original model. Thus, the model fits better overall.

Paths	Standardized path coefficient	Standard error	Is it significant ?
WA ← PU	0.339***	0.227	Significant
WA ← PEU	0.437***	0.290	Significant
WA ← SN	0.520***	0.188	Significant
PEU ← SN	0.283***	0.171	Significant
PEU ← PE	0.494**	0.254	Significant
PEU ← RD	0.340***	0.156	Significant
PU ← CC	0.113***	0.092	Significant
PU ← RD	0.472***	0.198	Significant
PU ← PEU	0.600**	0.331	Significant
PU ← PE	0.108***	0.058	Significant

PU ← SN	0.447**	0.152	Significant
WA ← ICE	0.117**	0.005	Significant
PU ← ICE	0.287*	0.203	Significant
PEU ← ICE	0.459**	0.256	Significant
Measurement models			
ICE ← Age	-0.176	0.124	——
ICE ← Education level	0.303	0.083	——
ICE ← Monthly household income	0.410	0.277	——

Note: *, **, *** denote p<0.1, p<0.05, and p<0.01, respectively

Table 5. Robustness test results after adding control variables (own editing)

To investigate more clearly the effect of the ICE control variables on various perceptual factors and willingness, this paper conducted one-way ANOVAs on three demographic characteristics: age, education level, and monthly household income, and the comparison of ANOVAs and cross-sectional results between groups showed the following: (1) Age has a significant effect on PU F=3.156, P=0.008) and WA (F=3.457, P=0.061). PU is reduced by the limited innovation awareness of older seniors, their low awareness and limited understanding of products with high technological content. Such elderly people are limited by the idea of insufficient usefulness and thus reluctant to actively adopt smart elderly services. However, the effect of age on PEU is not significant. (2) Education level has a significant effect on PEU (F=3.340, P=0.013) and WA (F=3.531, P=0.022). Older adults with higher education levels are more receptive to new technologies and more skilled in using and operating smart senior care services, thus increasing PEU. Moreover, educational experiences broaden the horizons of older adults, making them more likely to improve their quality of life through new technologies, and thus increasing their willingness to adopt services. However, there was no significant effect of education level on PU. (3) Monthly household income has a significant effect on PU (F=2.780, P=0.045) and WA (F=2.158, P=0.033). PU is enhanced by the fact that households with high income levels have fewer binding financial budgets and a greater demand for healthy elderly care services, and believe that smart elderly care services can better serve the elderly. At the same time, with the increase of income, elderly people's needs show diversified and complex characteristics, and the willingness to adopt services to better meet their demand level is increased. However, monthly household income does not significantly affect PEU.

6. Discussion

This study explored the adoption intention and influencing factors of smart elderly services in Anhui Province, and found that PU can significantly increase the adoption intention, which is consistent with the results of Liu Wei's study (Liu 2015).

However, the adoption intention degree is higher in this paper, indicating that the degree of demand for smart elderly services exceeds that of social network services, and PU is the primary prerequisite for measuring the adoption of new technologies (Xia and Zhang 2020). PEU has an impact on both adoption intention and PU, and PEU may not directly affect usefulness but is mediated by psychological effects to increase PU (Wang 2020), so increasing product ease of use will have multiple beneficial effects. SN significantly affect adoption intentions and PEU, and the neighborhood effect helps to dispel older adults' internal concerns and significantly increase PEU. The latter part of hypothesis 4 does not hold because convenience does not have a positive effect on PEU, probably because the first consideration of elderly users in adopting a smart senior care service is not how easy it is to use but how useful it is. Even if a service is easy to learn and master, if it does not have the desired effect, it will not ultimately increase adoption intentions. The reason why the results showed no increase in adoption may be that most of the elderly are unfamiliar with smart senior care services and direct adoption in a short period implies an increased risk of uncertainty. From a rational point of view, seniors are not likely to adopt the service directly; only when the PU and PEU of the smart senior care service are perceived to increase will it be possible to further increase the adoption intention. While Li et al. (2019) found an increase in farmers' willingness to adopt formula fertilizer application technology, the reason for the inconsistency with the findings of this paper may be because smart pension services, as a novelty, still need time to test their effectiveness and therefore reduce the willingness to adopt. In addition, an important implication of the results for us is that, when promoting and advertising the smart senior care service, we should pay much attention to whether the seniors are sure that they have felt the actual effects of the service, This is because only after the seniors' actual psychological feelings are improved can the adoption intention be finally increased. However, other studies (Bai and Zhu 2018; Zhang, Yuan and Zhou 2019) lacked measurement of service effectiveness in terms of demand-side psychological perceptions. Similarly, PE do not contribute effectively to adoption intentions. We believe that this may be because, despite the existence of hope values, the elderly are risk-averse and mostly hold a wait-and-see attitude. Fear hinders their use of the service, i.e., the increase in PU and PEU does not necessarily lead to adoption of the service, so it is more important to increase the actual adoption rate. However, another study (Niu, Zhang and Huang 2020) showed that PE can positively influence academic social network adoption intentions, which differs from this paper's study, probably due to the bias of the results caused by the different ages of the study group.

The robustness results show that age does not have a significant effect on PEU, and we argue that no matter how old, the elderly make usefulness their preferred goal, so there is a strong need to design products and services that are useful. The effect of education level on the PU of traditional Internet technologies has been studied previously (Chen 2020), but this paper shows that there is no significant effect of education level on PU, suggesting that the designers of new smart senior care services have taken into account the needs of people with different educational backgrounds. Monthly household income fails to affect PEU, so we can exclude the

interference of income factor when designing humanized services, and invest our efforts into other factor dimensions.

Although age, education level, and monthly household income have significant effects on adoption intentions, it is still not known what the adoption intentions and paths are for different types of elderly people, so group heterogeneity analysis of each of these three factors is a future research direction. Based on the TAM3 theory, this paper constructs a model of factors influencing the adoption intention of smart elderly services, which we believe not only expands the classical TAM theory to a certain extent but also deepens the research space for similar problems. For example, we can use the same framework to explore the technology adoption behavior of farmers by analogizing them to the elderly. This study not only deepens the social security research results but also extends the management theory and has some reference value for economics, especially agricultural economics.

Although this paper examines the relevant issues and makes some meaningful findings, there are some limitations. First, the research area of this paper is concentrated in Anhui Province, which is limited by the scope and number of studies, and more time is needed to test the generalizability of the findings. In the next step, the scope of the study can be broadened and more questionnaires can be used to further improve the persuasiveness. Second, many factors influence the elderly's willingness to adopt smart senior care services, so are there any important influencing factors missing? Are there any internal interactions among these factors? These can be expanded in depth in future studies. Despite the limitations, this study provides a strong basis for the government to conduct policy promotion and improve subsidy policies and early establishment of a demand-driven service system.

7. Conclusions

In this paper, after model validation, we found that the willingness of elderly people to adopt smart senior care services is influenced by various factors, and the study showed that SN, PU, and PEU are the key factors that directly affect the willingness of elderly people to adopt smart senior care services. That said, the magnitude of the influence of different factors is not consistent; in order of influence, they are PEU >PU > SN. Further, SN and PEU not only directly influence the willingness of elderly people to adopt smart senior care services but also act as mediators to positively promote WA by influencing PU. PE and RD have different degrees of positive effect on PU and PEU, and CC also increases the degree of PU. Age has a significant effect on PU and WA, while education level has a significant effect on PEU and WA, and monthly household income has a significant effect on PU and WA. The findings of the study are beneficial to the future development of more suitable policies for the promotion of smart elderly services, and the expanded TAM theory can be applied to agricultural economic research. Exploring the adoption intentions of different types of older adult based on a heterogeneity perspective is our next research direction.

Based on the findings of the study, we were able to draw several policy insights: First, improve the role of government policy orientation and enhance the cognitive

level of older adults. The role of smart elderly services in improving the overall quality and efficiency of elderly services is conveyed to enhance the PU of the elderly. Second, pay attention to the radiation effect of the demonstration group and play the leading role of cadres. Give full play to the exemplary leading role of cadres of government agencies and full-time senior care service personnel in the community to help the elderly solve technical problems and enhance the enthusiasm of elderly users to adopt the service. Third, improve the cost-effectiveness of smart senior care services and simplify the operation procedures. Establish a responsibility mechanism shared by individuals, government, and society to gradually eliminate the "digital divide" in the use of intelligent products by the elderly. Fourth, improve the standardization of intelligent senior care organizations and build a reasonable evaluation system. Establish a systematic evaluation model to effectively evaluate the implementation effect of intelligent senior care services in a precise and quantitative manner.

References

Anhui Provincial Bureau of Statistics data. "Anhui's 'aging' degree ranks 9th in the country." *Anhui Business Daily* (February 24, 2021). http://ah.sina.com.cn/news/2021-02-24/detail-ikftpnny9369558.shtml

Bai, Mei, and QingHua Zhu. "Impact Factors of Smart Care Needs and Volunteer Service Willingness for the Aged——A Case of Jianghan District in Wuhan." *Journal of Modern Information* 38, no. 12 (2018): 3–8. https://doi.org/CNKI:SUN:XDQB.0.2018-12-00

Bentler, P., and Chih-Ping Chou. "Practical issues in structural modeling." *Sociological Methods & Research* 16, no. 1 (1987): 78–117. https://doi.org/10.1177/0049124187016001004

Braun, Michael. "Obstacles to social networking website use among older adults." *Computers in Human Behavior* 29, no. 3 (2013): 673–680. https://doi.org/10.1016/j.chb.2012.12.004

Central People's Government of the People's Republic of China. "Notice of three ministries on the issuance of the Action Plan for the Development of the Intelligent Health and Aging Industry (2017–2020)." (2017). https://www.gov.cn/gongbao/content/2017/content_5222955.htm

Chae, Moon, Joo Heon Lee, Seung Hee Ho, Hee Ja Kim, Ki Hong Jun, and Jong Uk Won. "Patient satisfaction with telemedicine in home health services for the elderly." *International Journal of Medical Informatics* 61, no. 2–3 (2001): 167–173. https://doi.org/10.1016/s1386-5056(01)00139-3

Chakraborty, Rajarshi, Claire Vishik, and Raghav Rao. "Privacy preserving actions of older adults on social media: Exploring the behavior of opting out of information sharing." *Decision Support Systems* 55, no. 4 (2013): 948–956. https://doi.org/10.1016/j.dss.2013.01.004

Chen, Xin. "A study of the impact and differences of Internet use on social isolation of older adults." *Contemporary Economic Management* 42, no. 9 (2020): 53–59. https://doi.org/10.13253/j.cnki.ddjjgl.2020.09.008

Davis, Fred, Richard Bagozzi, and Paul Warshaw. "User acceptance of computer technology: A comparison of two theoretical models." *Management Science* 35, no. 8 (1989): 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Dufhues, Thomas, Judith Möllers, Diana Traikova, Gertrud Buchenrieder, and David Runschke. "Why villagers stay put—A structural equation model on staying intentions." *Journal of Rural Studies* 81, no. 6 (2021): 345–357. https://doi.org/10.1016/j.jrurstud.2020.10.054

Eastman, Jacqueline, and Rajesh Iyer. "The elderly's uses and attitudes towards the internet." *Journal of Consumer Marketing* 21, no. 3 (2004): 208–220. https://doi.org/10.1108/07363760410534759

Godfrey, Mary, and Owen Johnson. "Digital circles of support: Meeting the information needs of older people." *Computers in Human Behavior* 25, no. 3 (2009): 633–642. https://doi.org/10.1016/j.chb.2008.08.016

Goundrey-Smith, Stephen. "Technologies that transform: Digital solutions for optimising medicines use in the NHS." *BMJ Health & Care Informatics* 26, no. 1 (2019): 1–4. https://doi.org/10.1136/bmjhci-2019-100016

Gu, ShengZu, HuaJun Wu, and DongMei Cao. "Strategic Thinking on Rationalizing Aged Care Service System in China." *Population Research* 41, no. 1, (2017): 3–14. https://doi.CNKI:SUN:RKYZ.0.2017-01-001

He, ZhengYu, Mei Bai, and QingHua Zhu. "A quantitative study of China's pension policies from 2013–2017." *Journal of Information Resource Management* 9, no. 1 (2019): 21–29.

Huang, JingRu, and YongMo Liu. "Effect of Media Attention on the Cost of Corporate Debt Financing——Based on the Mediator of Accounting Conservatism." *Review of Investment Studies* 39, no. 2 (2020): 113–133. https://doi.CNKI:SUN:TZYJ.0.2020-02-008

Ji, Yong, Junho Choi, Jee Lee, Kwang Han, Jinwoo Kim, and In-Kwon Lee. "Older adults in an aging society and social computing: A research agenda." *International Journal of Human-Computer Interaction* 26, no. 11–12 (2010): 1122–1146. https://doi.org/10.1080/10447318.2010.516728

Jian, XiaoZhu, BuYun Dai, and HaiQi Dai. "Theoretical construction and simulation analysis of logistic weighting model." *Journal of Psychology* 48, no. 12 (2016): 1625–1630. https://doi.10.3724/SP.J.1041.2016.01625

Körtner, Tobias, Alexandra Schmid, Daliah Batko-Klein, and Christoph Gisinger. "Meeting requirements of older users? Robot prototype trials in a home-like environment." In *Universal Access in Human-Computer Interaction: Aging and Assistive Environments*, 660–671. Heraklion, Crete, Greece: 8th International Conference, UAHCI 2014, held as part of HCI International, 2014. https://doi.org/10.1007/978-3-319-07446-7_63

Li, ZiLin, Yi Han, Xi Guo, YeFeng Jiang, and WenLu Lin. "Analysis of Influencing Factors on Farmer's Willingness to Adopt Soil Testing and Formula Fertilization Technology Based on SEM." *Resources and Environment in the Yangtze Basin* 28, no. 9 (2019): 2119–2129. https://doi:CNKI:SUN:CJLY.0.2019-09-011

Liao, Chuhui, and Juan Chen. "Factors influencing smart aged care services acceptance of elderly in nursing homes: A perspective of perceived quality." In *Proceedings of the 5th Annual International Conference on Social Science and Contemporary Humanity Development (SSCHD 2019)*. Wuhan, Hubei, China: SSCHD 2019, 2019. https://doi.org/10.2991/sschd-19.2019.114

Liu, Wei. "SNS Acceptance of Older Adults Based the Extended TTF and UTAUT Model." *Soft Science* 29, no. 3 (2015): 120–124. https://doi:10.13956/j.ss.1001-8409.2015.03.26

Lu, JieHua, and Ran Guo. "Compression or expansion of morbidity? A study on the long-term change trends of health among the elderly: 1998–2014." *Population and Development* 25, no. 6 (2019): 76–86.

Lu, Yaobin, Tao Zhou, and Bin Wang. "Exploring Chinese users' acceptance of instant messaging using the theory of planned behavior, the technology acceptance model, and the flow theory." *Computers in Human Behavior* 25, no. 1 (2009): 29–39. https://doi.org/10.1016/j.chb.2008.06.002

National Health Commission data (China). "China's elderly population aged 60 and above to reach 255 million by 2020." *Thirteenth Five-Year Plan for Healthy Ageing* (2017). https://www.gov.cn/xinwen/2017-03/23/content_5180093.htm

Niu, YanXia, YaoKun Zhang, and Lei Huang. "The Factors Influencing the Use of Academic Social Network Services: An Empirical Study Using UTAUT Model." *Library* no. 4 (2020): 91–97.

Sintonen, Sanna, and Mika Immonen. "Telecare services for aging people: Assessment of critical factors influencing the adoption intention." *Computers in Human Behavior* 29, no. 4 (2013): 1307–1317. https://doi.org/10.1016/j.chb.2013.01.037

Steele, Robert, Amanda Lo, Chris Secombe, and Yuk Wong. "Elderly persons' perception and acceptance of using wireless sensor networks to assist health-care." *International Journal of Medical Informatics* 78, no. 12 (2009): 788–801. https://doi.org/10.1016/j.ijmedinf.2009.08.001

Tian, Jie. "Feasibility Analysis and Policy Suggestions to Carry out Science and Technology Endowment." *Studies in Dialectics of Nature* 31, no. 2 (2015): 56–60. https://doi:CNKI:SUN: ZRBZ.0.2015-02-014

Van der Kloet, Hieke. "Aging and smart independent living in Dutch rural areas." In *26th Annual European Real Estate Society Conference*. Cergy-Pontoise, France: ERES Conference, 2019. https://doi.org/10.15396/eres2019_197

Venkatesh, Viswanath, and Hillol Bala. "Technology acceptance model 3 and a research agenda on interventions." *Decision Sciences* 39, no. 2 (2008): 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Venkatesh, Viswanath, and Fred Davis. "A theoretical extension of the technology acceptance model: Four longitudinal field studies." *Management Science* 46, no. 2 (2000): 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Wang, Lei, Jing Qu, ZhongChao Zhao, and LiLi Ding. "Organizational performance expectation gap and behavioral choices of heterogeneous institutional investors: A dual principal–agent perspective." *Management World* 36, no. 7 (2020): 132–153. https://doi.10.19744/j.cnki.11-1235/f.2020.0107

Wang, XiaoFei. "Influencing factors and paths of farmers' willingness to adopt soil testing and formulated fertilization technology." *Journal of Hunan Agricultural University (Social Science)* 21, no. 1 (2020): 1–7. https://doi.10.13331/j.cnki.jhau(ss).2020.01.001

Xi, Heng, Xing Ren, and ShaoGuo Zhai. "Smart pension: The elderly care service innovation with information technology." *Scientific Research on Aging* no. 7 (2014): 14–22. https://doi.10.3969/j.issn.2095-5898.2014.07.002

Xia, EnJun, and ZhenMing Zhang. "A Meta-analysis of Adoption of Preventive Innovation Technology and Its Implications for Crossing the Chasm: Taking Wearable Healthcare Devices as an Example." *Journal of Technology Economics* 39, no. 2 (2020): 134–143.

Xia, EnJun, and ZhenMing Zhang. "A Meta-analysis of Adoption of Preventive Innovation Technology and Its Implications for Crossing the Chasm: Taking Wearable Healthcare Devices as an Example." *Journal of Technology Economics* 39, no. 2 (2020): 163. https://doi:CNKI:SUN:JSJI.0.2020-02-017

Xu, Lan, and Yu Zhang. "Quality improvement of smart senior care service platform in China based on Grey Relational Analysis and Fuzzy-QFD." *Grey Systems: Theory and Application* [preprint] (2021). https://doi.org/10.1108/gs-05-2020-0068

Yi, Yu. "Healthy aging in urban China: Governing the aging population." *Geographical Journal* 187, no. 1 (2021): 28–38. https://doi.org/10.1111/geoj.12372

Yu, Xiao, and Yue Sun. "The Model of"Internet + Old-age Service": The Innovative Development of Old-age Service in New Era." *Population Journal* 39, no. 1 (2017): 58–66. https://doi:10.16405/j.cnki.1004-129X.2017.01.006

Zhang, HuaiYing, DanQi Yuan, and ZhongLi Zhou. "Entrepreneurship, members' own capacity, and cooperative performance." *Guizhou Social Science* no. 5 (2019): 123–129. https://doi.10.13713/j.cnki.cssci.2019.05.018

Zhang, Jiang Guang, Teng Zhang, and Fang Wang. "A study of key information ecology factors that influence smart-government development." *Modern Information* 38, no. 3 (2018): 3–9. https://doi.10.3969/j.issn.1008-0821.2018.03.001

Zhang, QunLin, and Isabel Atani. "Rural older males' marriage concepts and marriage strategies under the marriage squeeze." *Population and Development* 25, no. 4 (2019): 106–116.

Zhang, QunLin, and Isabel Atani. "Rural older males' marriage concepts and marriage strategies under the marriage squeeze." *Population and Development* 25, no. 4 (2019): 93.

Zhang, Yuan, and NanNan Lian. "Institutional Intelligent Care Service and the Satisfaction of the Elderly: Empirical Evidence from Baotou City." *Chinese Journal of Population Science* no. 2 (2019): 112–125.

Zhang, Yuan, and NanNan Lian. "Institutional Intelligent Care Service and the Satisfaction of the Elderly: Empirical Evidence from Baotou City." *Chinese Journal of Population Science* no. 2 (2019):128. https://doi:CNKI:SUN:ZKRK.0.2019-02-011

Zhu, HaiLong. "Smart Care: Innovation and thinking of Chinese elderly care paradigm." *Journal of Social Science of Hunan Normal University* 45, no. 3 (2016): 68–73.