KGRI
Keio University Global Research Institute

KGRI Working Papers

No.4

Joint Statement：“Toward a Wholesome Platform for Speech: Implementing Information Health” (ver 2.1)

Version 2.1

October 2024

Fujio Toriumi

Professor, Graduate School of Engineering, The University of Tokyo

Tatsuhiko Yamamoto

Professor, Law School, Keio University

Deputy Director, Keio University Global Research Institute

Keio University Global Research Institute

Joint Statement

"Toward a Wholesome Platform for Speech: Implementing Information Health" (version 2.1)

Joint lead authors:

Fujio Toriumi (Graduate School of Engineering, The University of Tokyo)

Tatsuhiko Yamamoto (Professor, Law School, Keio University; and Deputy Director, Keio University Global Research Institute)

Coauthors:

Teppei Koguchi (Professor, College of Information, Academic Institute, Shizuoka University)

Masatoshi Kokubo (Assistant Professor, Interfaculty Initiative in Information Studies, Graduate School of Interdisciplinary Information Studies, The University of Tokyo, and member of Keio University Global Research Institute)

Kuniyoshi Sakai (Professor, Graduate School of Arts and Sciences, The University of Tokyo)

Yuya Suzuki (Formerly of *BuzzFeed Japan*)

Ryotaro Soma (Legal Apprentice, 77th term)

Kazuhiro Taira (Professor, College of Arts and Sciences, J. F. Oberlin University)

Takashi Chiba (Dentsu Inc.)

Katsue Nagakura (Chief Researcher, Nikkei BP Intelligence Group, Research Unit)

Shuya Hayashi (Professor, Graduate School of Law, Nagoya University)

Hiroyuki Fujishiro (Professor, Faculty of Social Sciences, Hosei University)

Taro Magome (Dentsu Digital Inc.)

Asako Miura (Professor, Graduate School of Human Sciences, Osaka University)

Eijiro Mizutani (Associate Professor, Faculty of Sociology, Kansai University)

Shinichi Yamaguchi (Principal Researcher and Associate Professor, Center for Global Communications, International University of Japan)

Yuki Tonfi (Attorney/SmartNews, Inc., ZeLo, a Foreign Law Joint Enterprise)

Research Assistants

Sakiko Ohki (Faculty of Law, Graduate School of Law and Politics, The University of Tokyo), Karen Shimizu (Department of Law, Faculty of Law, Keio University), Yui Takeuchi (Media Studies, Department of Sociology, Faculty of Sociology, Kansai University), and Sosuke Nakano (Media Studies, Department of Sociology, Faculty of Sociology, Kansai University)

Summary

Over a year has passed since the release of the joint statement, “Toward a Wholesome Platform for Speech: Digital Diet Declaration (ver. 1.0).” This initiative aimed to highlight the pressing issues related to public discourse and engage society in discussions on what is necessary to foster “information health.” While version 1.0 successfully introduced the concept, it was somewhat limited in its definition of “information health” and left certain key areas unexplored. These included perspectives from neuroscience, education and literacy challenges, advertising-related issues, and the role of telecommunications carriers. Furthermore, it did not analyze the harms caused by “unhealthy information.” Additionally, the rapid development and proliferation of generative AI in recent months were entirely beyond the scope of version 1.0. Therefore, this updated statement (version 2.0) expands on the original by addressing these overlooked issues and exploring pathways to achieve “information health.”

Through this statement, we hope to deepen understanding, both in Japan and globally, of the various challenges facing public discourse spaces due to the excesses of the attention economy. We also aspire to facilitate discussion on how to overcome these challenges.

Addendum

On March 26, 2024, the project held a symposium titled “The Shadow of the Attention Economy and ‘Information Health’: Healthy Spaces for Public Discourse Created by Collective Intelligence.”

This event showcased research and initiatives related to information health and, through lectures and panel discussions, examined the challenges confronting public discourse spaces today, as well as the current state of information health. It also provided an opportunity for interdisciplinary and cross-disciplinary discussions among researchers from various fields and key stakeholders to promote efforts for implementing information health.

This updated statement (version 2.1) incorporates key insights from the symposium and includes Appendix III as a detailed research report.

*This symposium was co-sponsored by Keio University Global Research Institute and supported by the “2040 Independence & Self-Respect Project <Security> Platform and the ‘2040 Problem’: Formation of a New Order in the Network Space.”

Table of Contents

Preamble

The COVID-19 pandemic revealed that fake news[1] and infodemics[2] disrupt the information space and are significant obstacles to infection control measures, endangering human lives[3]. Simultaneously, digital platforms (DPs), the primary venue for these phenomena, have become indispensable to daily life, supporting everything from shopping to communicating with friends to news consumption, dramatically enhancing convenience. These two dimensions are intricately intertwined. As the volume of the information we receive has exploded, specific algorithms designed to attract our attention have selectively filtered this information, leading to several problems, including imbalances in information consumption. This imbalance has amplified the power of fake news to influence us, creating filter bubbles and echo chambers.

In any case, the option of what information to consume is a matter of individual freedom, and the autonomy to consume information should be respected. However, as seen during the COVID-19 pandemic, when the spread of fake news and infodemics hinders infection control and disrupts society, it becomes a collective problem. This applies to the pandemic and democracy itself, as seen in the chaos surrounding the 2020 US presidential election and the US Capitol riot.

A particularly pressing issue is that under the "attention economy" business model, many of us are forced to consume unbalanced information. Algorithms designed to maximize the economic interests of business operators compel us to consume specific content in a non-autonomous manner. Moreover, many of us are not fully aware that we inhabit this kind of information environment[4]. Thus, the most critical task is to recognize the structure of the information environment we live in and the "diet" of information we consume daily[5]. Additionally, it is essential to provide a way for those seeking to improve their unbalanced diet of information to access diverse and balanced content.

We must achieve a state of "information health to enjoy the benefits of information and communication technologies while avoiding the harms of selective information consumption and fostering a healthy discourse environment consistent with constitutional principles. This concept refers to the idea that individuals achieve and maintain their desired state of well-being within the information environment, which forms the foundation of a democratic society. This well-being includes acquiring a certain "immunity" to misinformation, such as fake news, through a balanced intake of diverse information sources. Achieving this goal requires efforts from DP operators, media outlets, telecommunications carriers, advertisers, and users themselves. Furthermore, governments must provide various forms of indirect support.

---

[1] Although "fake news" can be defined in various ways, this paper broadly uses the term to encompass misinformation and false rumors in addition to false information.

[2] See Ministry of Internal Affairs and Communications, "Dissemination of Erroneous Information and Fake News," in *White Paper 2020: Information and Communications in Japan*, Part 1, Chapter 2, Section 3 (1). <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/html/nd123110.html>

[3] According to a survey of 8,001 individuals in the United Kingdom and the United States conducted in September 2020 by researchers affiliated with Imperial College London, after having viewed erroneous information regarding COVID-19 vaccines, such as "vaccinations will alter your DNA," those who planned to get vaccinated dropped by 6.2% in the United Kingdom and 6.4% in the United States. See Sahil Loomba et al., "Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA," *Nature Human Behaviour* 5, (2021): 337–348. <https://www.nature.com/articles/s41562-021-01056-1>

[4] In October 2021, Twitter announced that a seven-country survey that included the United States, several European countries, and Japan showed its algorithms were four times more likely to present right-wing vs. left-wing political posts. See Ferenc Huszár et al., "Algorithmic amplification of politics on Twitter," *PNAS* 119(1), (2022): e2025334119. <https://www.pnas.org/content/119/1/e2025334119>

[5] Public relations company Edelman Japan, in its "2021 Edelman Trust Barometer" global survey, measured information hygiene using four criteria: Do those surveyed actively engage with the news, avoid information echo chambers and engage with differing perspectives, verify information, and avoid spreading unvetted information to others? An average of 26% had "good" information hygiene levels for three or more of these criteria across the 27 countries surveyed, contrary to 19% for Japan alone. Conversely, an average of 39% had "poor" information hygiene—meeting one or zero criteria—across the 27 countries versus 56% for Japan. <https://www.slideshare.net/EdelmanJapan/2021-246556173/24>

These issues constitute national challenges that threaten basic constitutional principles, such as individual dignity and democracy. However, imposing overly strict legal restraints on DP operators could be considered governmental overreach. A hard-law approach risks suppressing innovations that drive digital technology development and could lead to state censorship and ideological manipulation. Thus, at this stage, this approach would be neither appropriate nor effective.

This joint statement presents actionable recommendations for various stakeholders—users, platform operators, and the government—to establish an environment where all individuals who desire information health in public discourse can achieve and enjoy it.

I. Current Issues

(1) The Big Bang of the Information Space

Before society entered the digital era, the information space was relatively closed, functioning as a privileged domain. There were physical restrictions on air waves and page space, and the transmission of information was unilateral, slow, and minimal in volume. Additionally, the primary information deliverers were restricted to professional journalists affiliated with mass media organizations, along with a select few intellectuals upon whom these outlets focused.

However, the advent of the internet has triggered a "big bang" in the public information space, resulting in an entirely new, hitherto unseen dimension in which information is exchanged bidirectionally, in real time, and without restrictions. In the internet's early days, this big bang was expected to create an open, democratic world. Nevertheless, the current reality is starkly different: a flood of user-generated content, credible and unreliable material, posted under real and assumed names by numerous ordinary users. It is a world in which fake news of unknown origin exists along with fact-based journalism and is further amplified by non-human actors, such as bots. The rapid spread of generative artificial intelligence (AI) has also made it easier to create fake news that appears plausible and disseminate it on a massive scale, creating an "information tsunami." Consequently, the risks associated with manipulating information and public opinion in the information space are increasing.

Many DP operators mix professionally generated content with content created by individuals possessing no formal training or shared ethical codes of conduct. This content is ranked by proprietary algorithms. Thus, traditional mass media, including newspapers and magazines, are forced to compete fiercely in an attention economy, where a momentary impact on readers or users governs visibility.

Recommendation algorithms that use AI and other technologies to provide personalized information have the advantage of ensuring that users receive content tailored to their interests. Nonetheless, these same algorithms often lead users to consume only content recommended to them by the platform, creating "filter bubbles." In such cases, individuals can only see the information they want, as judged by the algorithm, encasing them in an "information cocoon." This phenomenon fosters "information malnutrition," where users are exposed only to one-sided information.

(2) The Attention Economy

We are ceaselessly bombarded with information from the internet and social media, even as we produce content ourselves, perpetuating the endless and repeated consumption and sharing of information. In this environment, we never starve for information.

In a society with such information overload, consumers' attention and time are scarce compared to the quantity of information supplied. Thus, the economic value of that time and attention has grown, and these limited resources have become commodities in the "attention economy"[6]. The deep integration of smartphones and other mobile devices into our daily lives has also made individuals increasingly beholden to the attention economy.

In psychology's dual-process theory, human thought is divided into two modes: System 1, which is intuitive and automatic, and System 2, which complements System 1 and is deliberative and reflective[7]. Within the attention economy, stimulating System 1 is considered critical, as reflexive responses[8] to

[6] See generally Tim Wu, *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. (Vintage, 2017).

[7] See Daniel Kahneman and Akiko Murai (trans.), *Thinking, Fast and Slow: How Is Your Intent Determined?* (Hayakawa Nonfiction, 2014).

[8] See Ryosuke Nishida, *Information Armed Politics* (Kadokawa, 2018).

stimuli generate economic value. These reflexive reactions are evaluated in terms of page views (PVs), time spent on a web page, and site stickiness, among others, and monetized. Consequently, content that elicits more stimuli, including fake news, is more likely to have higher visibility and subsequent financial benefits, such as advertising revenue, rather than fact-based, accurate information.

If left unchecked, this economic model risks diminishing System 2 thinking, which forms the basis of discussion and deliberation, endangering the democratic principles on which these are based.

(3) Mind-Hacking

Advancements in AI-powered profiling have made it possible to analyze and predict users' political beliefs, emotions, and psychological traits with extremely high accuracy. For instance, during the 2016 US presidential election, the British consulting firm Cambridge Analytica reportedly conducted "psychographic profiling." This technology facilitates political micro-targeting, making it easier to "hack" individuals and manipulate their emotions and thoughts—a practice called "mind-hacking"[9]. Such psychological interventions and manipulations, conducted without individuals' awareness, undermine personal self-determination, autonomy, and the democratic principles that rely on such abilities.

Moreover, DP operators increasingly employ "dark patterns"—manipulative designs—in user interface (UI)/user experience (UX) interactions that unfairly influence user decision-making for their economic gain[10].

In neuroscience, breakthroughs in neural decoding technology enable brain functional imaging to collect data on individual neural activity, which is analyzed using AI to decode individual sensory inputs and motor outputs. Meta, a major DP operator, is also participating in the development of this technology[11]. While these techniques, sometimes referred to as "mind-reading," hold promise for various applications, including the treatment of mental illness and neurorehabilitation, their misuse could violate mental privacy and lead to mind control and brainwashing. Should such technologies be implemented in the information ecosystem without proper regulation, they could seriously undermine individual autonomy, decision-making, and democratic governance. Thus, it is necessary to monitor the development and implementation of these technologies in relation to DPs.

(4) Filter Bubbles and Echo Chambers

Big data-driven profiling and targeting allow users to access information curated specifically for them, filtered out from the excessive information surrounding them. However, while this personalization may seem beneficial, research has indicated a tendency among people to become more radical in their thinking when they engage in discussion with other like-minded people—a phenomenon known as "group polarization"[12]. This tendency, coupled with the characteristics of the internet, gives rise to phenomena such as "echo chambers" and "filter bubbles."

Algorithms analyze user preferences and present information prioritizing these preferences. Consequently, users are only exposed to information considered relevant to their interests, effectively

[9] See Christopher Wylie and Hiroshi Makino (trans.), *Mindf*ck: Cambridge Analytica and the Plot to Break America* (Shinchosha, 2020).

[10] See Keio University Global Research Institute, "Thinking about the 'Dark Pattern' of Deceptive Service Design - Special Presentation 'Dark Pattern' Technology and Ethical Issues" (held on June 17, 2021). <https://www.kgri.keio.ac.jp/news-event/083668.html>

[11] Jean Remi King, "Using AI to Decode Speech from Brain Activity," *Meta* (blog), August 31, 2022, <https://ai.facebook.com/blog/ai-speech-brain-activity/>

[12] Refer to the works of Cass Sunstein. For example, Cass Sunstein *#Republic*, trans. Naomi Date (Keiso Shobo, 2018).

trapping them in filter bubbles. Within these bubbles are thoughts and opinions similar to those of the users, whereas opposing views are filtered out, making them unaware of their existence.

A similar process occurs on social media. Users predominantly follow others with similar interests, leading to the reinforcement of certain viewpoints, referred to as an "echo chamber"[13]. Repeated exposure to the same opinions can lead individuals to believe in their accuracy and validity, which has fueled the spread of conspiracy theories.

Observers have pointed out that group polarization will accelerate because of filter bubbles and echo chambers[14]. As individuals become more extreme in their views, they grow less accepting of different views and increasingly unwilling to engage in dialogue. These two phenomena can exacerbate divisions in society and jeopardize democracy.

(5) Fake News

The digital space today is inundated with fake news, often created to generate advertising revenue or erode trust in prominent individuals, political groups, or corporations. This content is spread and amplified with the help of bots and other automated tools. The creation and dissemination of fake news can also be understood as a phenomenon deeply rooted in human behavior, as exemplified by concepts such as the "social porn hypothesis" (the tendency for users within a specific community to share sensational information reflexively) and slacktivism (participation in low-effort social movements). The attention economy encourages the spread of fake news by emphasizing "stimulus = reflex," where engagement is prioritized over accuracy. For example, most algorithms for digital advertisements calculate costs based on how many users they have attracted (e.g., number of impressions, PVs, unique users), creating a system that incentivizes the continuous creation and dissemination of fake news to generate advertising revenue.

Additionally, fake news is frequently deployed as part of efforts by other countries to conduct "influencing maneuvers,"[15] jeopardizing national security and sovereignty. These problems will be exacerbated by the accelerating development of generative AI, which can create highly convincing falsehoods.

(6) Slander and Flaming

Slander is also a major issue in today's digital public discourse. Extreme statements and abusive language proliferate on social media and internet forums, which is connected to the reflexive stimulus-driven structure of the attention economy. Indeed, cases of suicide attributed to slander have occurred, and "words" on social networking sites have been turned into potentially lethal weapons.

Social problems associated with slander and flaming (i.e., the act of posting angry or insulting messages to social media) have emerged in various situations. Recent studies have shown that younger generations are more likely to encounter slander on the internet[16], raising concerns about its detrimental effects on their

[13] See Kazutoshi Sasahara, *The Science of Fake News: The Mechanism of Spreading Hoaxes, Conspiracy Theories, and Propaganda* (Dojin Publishing, 2021), Chapter 3.

[14] However, careful empirical research will be crucial. For reference, see Tatsuo Tanaka and Satoshi Hamaya, *The Internet Does Not Divide Society* (Kadokawa, 2019). For a recent important empirical study, see Daisuke Tsuji, ed., *The Network Society and Democracy*. (Yuhikaku, 2021).

[15] According to Bradshaw et al. (2021), as of 2020, 81 countries were engaged in forms of "computational propaganda" like fake news, an increase of 11 countries from 2019. See Samantha Bradshaw, Hannah Bailey, and Philip N. Howard, *Industrialized Disinformation. 2020 Global Inventory of Organized Social Media Manipulation* (University of Oxford, 2021). <https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/127/2021/01/CyberTroop-Report-2020-v.2.pdf>

[16] See Shinichi Yamaguchi, Tsukasa Tanihara, and Hidetaka Oshima, *Innovation Nippon 2022: Survey on the Current State of Slander and Defamation in Japan* (Glocom, 2023), <https://www.glocom.ac.jp/activities/project/8806>

mental development and psychological well-being. Furthermore, journalists experience slander at rates over four times higher than the general public,[17] sometimes hindering their ability to continue reporting duties effectively[18] and undermining freedom of the press. New challenges have arisen with the development of technology, including slander in the metaverse and slanderous attacks by bots.

(7) Deficiency in Ad Verification

The "big bang" expansion of spaces for public discourse has also led to an explosion in advertising opportunities for businesses. The internet, in particular, has made it possible to advertise in numerous public discourse spaces, even in the metaverse. In Japan, internet advertising expenditures continue to grow, totaling 3,091.2 billion yen in 2022 (a 114.3% year-over-year increase)[19].

With this growth, the importance of ad verification—the process of ensuring advertisement integrity and appropriateness—has increased. Key aspects include brand safety (whether advertisements appear next to inappropriate content), viewability (whether advertisements are seen by users), and ad fraud (identifying fraudulent views or clicks).

The prevailing method of internet advertising distribution constitutes bundling large amounts of internet content and delivering large-scale, batch advertising. From a business perspective, this is an efficient method, as individual pieces of content are not scrutinized prior to distribution. Within the framework of the attention economy, advertisers are incentivized to pay more to reach larger audiences, measured by metrics such as PVs. Therefore, stimulating and attention-grabbing content will likely earn more advertising revenue. This situation encourages fake news and other forms of harmful information that impede achieving information health.

If consumers become aware that advertisers are actively placing ads alongside fake news or sensational content, they may criticize advertisers for prioritizing advertising gains over information quality or encouraging the spread of inaccurate or detrimental information. This can lead to a loss of credibility and damage the brand image.

Some advertising businesses have attempted to mitigate these risks by using blocklists and safety lists to control where their ads appear, but these measures have been insufficient. The problem of ad fraud is also growing, in which automated programs masquerading as humans are used to inflate the number of ad clicks, damaging the credibility of the advertising model itself and casting doubt on the reliability of PV-based indicators.

(8) Use of Generative AI

The use of generative AI, as exemplified by ChatGPT, introduces a range of new challenges in digital spaces for public discourse.

First, the origins of the content generated by AI often lack transparency, as it is unclear what internet data have been used and processed for the output. Thus, there is a risk that the unverified content will be reused and further spread. Moreover, as users tend to avoid explicitly stating their use of generative AI, a chain reaction may occur in which internet public discourse spaces are formed where the authenticity of

[17] See Shinichi Yamaguchi, Tsukasa Tanihara, and Hidetaka Oshima, *Innovation Nippon 2022: The Reality of Slander against Journalists* (Glocom, 2023) <https://www.glocom.ac.jp/activities/project/8806>

[18] Refer to footnote 17.

[19] See Dentsu Inc., "2022 Advertising Expenditures in Japan: Detailed Analysis of Internet Advertising Media Expenditures." March 14, 2023. <https://www.dentsu.co.jp/news/release/2023/0314-010594.html>

information cannot be confirmed. Generative AI could, in turn, recycle this unverifiable content. From an information health perspective, the unchecked spread of generative AI may result in users being unaware of what kind of information they are consuming (i.e., whether it is created by humans or AI and, if the latter, what kind of AI created it). While generative AI's probabilistic outputs often seem plausible and convincing, their unconstrained use can lead to cognition distortion. Metaphorically, these "delicious poison apples" can slowly damage our informational health without us even realizing it.

Second, there is a danger that someone with malicious intent will use generative AI to create and spread vast quantities of highly realistic fake news, simulating human authorship. This capability could create an "information tsunami," powerfully manipulating public opinion for a specific agenda.

Third, while natural language processing in generative AI, such as ChatGPT, has advanced, its understanding of user intent and meaning remains superficial. Nevertheless, humans compensate for these shortcomings[20], fostering a misconception that generative AI provides active, nuanced insights, unlike traditional passive information searches. Furthermore, these systems' interactive and conversational nature creates a sense of immersion, potentially leading to dependency or addiction to generative AI. To preserve information health and improve the integrity of public discourse spaces, it is critical to monitor future developments in generative AI.

(9) Absence of a "Cure-All"

As outlined thus far, the problems plaguing today's public discourse spaces are rooted in complex, structural mechanisms. Accordingly, no single remedy exists to address these issues comprehensively.

Compounding this challenge is the risk that continued exposure to a disrupted information environment—caused by fake news and other factors—could increase our vulnerability to misinformation. The problem extends beyond merely being complicit in the spread of fake news and falling prey to conspiracy theories. Such disruptions can exacerbate social divisions, erode tolerance toward others, and hinder our ability to coexist and maintain community bonds. This poses a threat to democracy and freedom.

In the absence of a universal remedy, we can strive to promote information health by first cultivating an awareness of balance in our information consumption.

[20] See Kuniyoshi Sakai (ed.), *The Brain and AI: Approaches to Language and Thought* (Chuko Sensho, 2022).

II. Guiding Principles of the Joint Statement

(1) Achieving Information Health

Considering the significant changes to the distribution of information in public discourse spaces, it is essential to achieve information health, which refers to a state in which individuals attain a satisfactory level of health in the information environment[21], which helps support a democratic society. The attention economy, through mechanisms such as filter bubbles, has led to an unbalanced "diet" of information with harmful effects on individuals and society. Thus, creating an environment in which individuals are exposed to diverse discourses and opinions is crucial. This aligns with the goal of fostering a state in which each individual has a certain degree of "immunity"—critical-thinking skills—to fake news by consuming a wide range of information.

Conversely, society's understanding of the importance of physical health has already advanced to the point where we can directly identify the nutrition we need through System 2-thinking or instinctively avoid dangerous substances through System 1-thinking. However, in this age of information overload and saturation, there is limited societal recognition of the importance of information health. Unlike the general awareness of maintaining a balanced diet for physical well-being, few people strive to consume a well-rounded diet of information. The fundamental problem with the attention economy lies in its attempt to exploit this lack of awareness to drive people to consume as much information as possible for economic gain. In this respect, it is necessary to enable users to consume diverse information sources autonomously and independently by raising their awareness through literacy and education, encouraging responsible practices by businesses, and receiving appropriate but indirect support from the government. Furthermore, ensuring access to basic information is vital for maintaining a democratic society.

The concept of information health is also closely related to the constitutional right to receive information. The Supreme Court of Japan has acknowledged that individuals must have opportunities to engage with diverse opinions, knowledge, and information to develop their thoughts, personalities, and social lives. This freedom is necessary for the effective achievement of the principles denoted in Article 21, paragraph 1 of the Constitution, which guarantees the free communication and exchange of ideas and information in a democratic society.[22]

Concurrently, the right to consume information selectively, reflecting individual preference, is also protected under Article 13 of the Constitution, which guarantees the right to pursue happiness. This ensures that individuals are not forced by the government to consume a variety of information but can make autonomous choices. Importantly, the government must ensure access to diverse information, and those who seek to maintain information health must have the opportunity to do so. This includes transparency regarding the algorithmic logic used to recommend information and the opportunity to withdraw from filter bubbles and echo chambers.

Information health is also associated with Article 25, the right to reasonable standard of "wholesome and fulfilling living." Access to diverse information is essential for achieving this standard. Under the Constitution, the decision of what constitutes a "good life" is left to the individual, and here, too, the state cannot force people to live healthily. Instead, the state's role is to provide resources for those who seek to live healthy lives.

(2) Education and Literacy on Information Health

---

[21] The World Health Organization defines health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." <https://www.who.int/about/governance/constitution>

[22] Repeta v. Japan, 43 Minshu 89 (Supreme Ct., Grand Bench, March 8, 1989).

While information and media literacy are important, it is crucial to first understand the characteristics of today's information space. With less than 20% of people aware of terms such as "attention economy," "echo chamber," and "filter bubble"—which describe the modern information space[23]—it is difficult for them to even recognize issues with their information intake.

Gaining literacy in the information space is analogous to understanding the role of diet and nutrition in physical health. In this context, examining the evolution of societal recognition of a balanced diet and food education can provide valuable insights.

In Japan, the word "nutrition" originated in the Edo period and gained broader use thereafter. During the Meiji era, nutrition education was promoted mainly within the military, while during the Taisho era, the private sector emphasized the importance of menus designed for meeting nutritional needs[24]. However, during the postwar period and the era of rapid economic growth, nutritional education was not generally prioritized[25]. In response to lifestyle changes, nutritional imbalances, irregular eating habits due to more single-person households, obesity and lifestyle-related illnesses, and food safety issues, the *Basic Act on Food and Nutrition Education* was enacted in 2005. Since then, nutrition education has been actively promoted in Japan, contributing to widespread knowledge of nutrition. The proliferation of cooking-related media and access to diverse food information have also contributed to the spread of nutrition education[26]. Through this process, individuals can now make independent and informed decisions regarding maintaining a healthy diet.

In today's information ecosystem, similar to the dietary problems that emerged after the period of rapid economic growth, issues related to "overconsumption" and an imbalanced "diet" have emerged at the individual and societal levels. This suggests that literacy and education initiatives akin to past nutrition education efforts are necessary. Equitable access to such programs must also be guaranteed.

The accelerated development of generative AI must also be considered in the landscape of literacy and education. For democracy to thrive, individuals must deliberately examine the nature of information via System 2-thinking and engage in dialogue with others who have different opinions. Contrary to notions of cognitive enhancement through digitization[27], the use of generative AI will not cause the human brain to evolve. US linguist Noam Chomsky warns that the uncritical use of AI will "degrade our science and debase our ethics"[28]. The societal discussion of generative AI's impact on literacy and education has only just begun. While some actively encourage its use, others are cautious from an information literacy perspective. A key concern is the blurred boundary between external input and original creation when using generative AI. Unlike legitimate citation practices, AI-assisted work may diminish creative abilities without the user being aware. Additionally, it is ethically unacceptable to misrepresent one's abilities or fail to credit the original author's ideas in any work. While traditional plagiarism and piracy committed by humans are straightforward in assigning accountability, misuse of generative AI may reduce the psychological burden of guilt felt by the user. How should we deal with these potential problems in achieving information health? In the following, we offer some ideas.

(3) Content Categorization and the "Information Ingredient List"

[23] See Appendix II, "Basic Survey on the Information Environment," in this statement.

[24] See Ayako Ehara et al., *A History of Japanese Food* (Yoshikawa Kobunkan, 2009).

[25] See Nobuo Harada et al., *Understanding Society through Food Culture!* (Seikyusha, 2009).

[26] Refer to footnote 25.

[27] See Kuniyoshi Sakai, *Reading to Create Brains: Why Paper Books are Necessary for People* (Jitsugyo no Nihonsha, 2011).

[28] See generally, Noam Chomsky, "The False Promise of ChatGPT," *New York Times*, March 8, 2023. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>

To support information health, it is important to empower users to make proactive and autonomous choices when consuming content and information on DPs. Achieving this goal requires clear indicators so that users can determine which content to consume and which DPs to utilize.

① Labeling the "Ingredients" of Content and Information

First, users must be given metadata transparently describing the nature of the content with which they are about to engage. These metadata could include the following: who provided the information, when, and for what purpose; whether the content was AI- or human-generated; the reactions of others who came into contact with the information.

Metadata must be provided through UIs and UXs that facilitate immediate user comprehension. They can be distilled into straightforward evaluation indicators, similar to calorie counts or nutrition labels. Moreover, methods to present more detailed information for users seeking it and visualize metadata indicators must be considered.

This approach mirrors the *Food Labeling Act*'s requirement to label the calorie content and nutritional value of food. This information helps consumers decide whether to consume a given food. Similarly, in the digital age, labeling and categorizing content, as well as visualizing its details, will be instrumental to improve users' understanding of the information they consume.

② Displaying a Balance of Content

Second, DPs should be transparent regarding the balance in the content offered through its services. For example, news sites often use algorithms to provide personalized information to users. Awareness of the degree of balance in the content offered would allow users to make more informed and autonomous decisions about which DPs to utilize.

Television broadcasting operates under specific regulations, such as the principle of programming consistency outlined in the *Broadcasting Act*, to avoid bias in the creation of content[29]. These regulations were deemed necessary because of airwave scarcity and the significant social influence of broadcast media. Similarly, a few DPs today (which are effectively limited in the sense that there are only a certain number of options) wield comparable social influence as the primary channels for information distribution.

Accordingly, for some DPs[30], it is desirable to recommend specific content targeting user interests and provide features enabling users to have a balanced exposure to a wide range of information genres. Specifically, drawing from the programming consistency principle, the ratio of content provided (e.g., what balance of political, economic, entertainment, or sports news should be displayed on the home page) should be determined independently, and this information should be presented in a way that is easy to understand.

③ Specific Measures and Implementation

[29] Article 106, Paragraph 1 of the *Broadcasting Act* stipulates that "excluding those based on special business plans, the basic broadcaster must establish cultural programs or educational programs and news programs, and entertainment programs and must maintain mutual consistency between the broadcast programs." In 2008, during the relicensing of terrestrial television broadcasters, the Ministry of Internal Affairs and Communications added requirements based on the inspection standards of the *Radio Act*. These stipulate that the NHK's General channel and private broadcasters must allot 10% or more of their weekly programming to educational programs and 20% or more to cultural programs for maintaining mutual consistency between broadcast programs. Under the obligations to announce program categories publicly, which was established in a 2010 revision, every year for six months starting in April, the programs broadcast during the third week of alternating months by the broadcasters in question would be categorized as "cultural programs," "educational programs," "news programs," "entertainment programs," and "other programs." After the six-month period, the broadcasters would then be obliged to publicly disclose the breakdown of program categories as soon as possible.

[30] Here, "DPs" does not necessarily refer to all DPs. The DPs expected to be more balanced are those that are especially large and highly public in nature. The precise definition is left for future discussion. Notably, the EU's *Digital Service Act* designates large online platforms that reach over 10% of EU consumers as DPs, on which it imposes particularly strict liability.

Give the diverse categories of content and information, no single indicator can fully describe their nature. Thus, careful consideration should be given to which indicators to establish to balance simplicity and depth, ensuring they capture a multifaceted perspective.

Originator profile (OP) technology is a potential mechanism to ensure the authenticity and trustworthiness of content creators. OP is described as "a technology that makes it easy to distinguish high-quality articles and media that have been authenticated by a third party by providing verifiable information about web content creators and advertisers." If implemented, OP would display digitally signed information regarding the originator's authenticity and trustworthiness (e.g., corporate values, editorial guidelines, responsibility for news accuracy, and privacy policy) verified by an impartial third-party certification body. By accessing this information displayed by OP, users can objectively ascertain the trustworthiness of content creators and site operators[31]. This transparency could lead users to make informed, proactive, and autonomous decisions about their intake of information, contributing to the achievement of information health.

(4) Providing an "Information Checkup"

Users should be given regular opportunities to conduct an "information checkup" to audit their information health. Through these checkups, individuals can identify the type of information they have been exposed to, which could motivate them to modify their information intake behavior.

In the context of rising prevalence of the attention economy, these checkups are recommended for individuals who feel uneasy about potential biases in the information they consume and those who seek greater balance in their content. However, some may lack the motivation or awareness to undergo such checkups, especially if they have already developed low immunity to fake news due to prolonged exposure to unbalanced or misleading information. To address this, it is important to design incentives to encourage participation in information checkups, but these should not be forced by other parties, especially the government. Finally, the collection and management of personal data for conducting checkups must prioritize user privacy and apply robust security protocols.

The information checkup could include the following specific functions:

① Assess the diversity of information sources users consume to detect potential bias and evaluate the reliability of these sources. The results should be presented as objective data.

② Calculate the degree to which users are affected by filter bubbles and echo chambers. For example, it is possible to visualize deviations in the information users encounter on social media compared to that on social media as a whole to help users recognize biases, deepening their understanding of their information space.

③ Simulate virtual information preferences. Specifically, users are provided a hypothetical information environment where they can explore extremely biased information and experience the results, allowing them to reflect on their information health. For example, interactive simulations can provide information only related to "COVID-19 conspiracy theories," "games," or "sweets." The kind of information provided in these simulated information spaces, such as what kind of videos and news are recommended on corresponding sites and how search results change, can clarify how a biased information diet impacts perception and cognition.

④ Create an avatar that reflects information health. Unlike physical health, information health is not easily perceptible, and even poor information health does not manifest as does physical pain. Therefore, avatars—virtual alter egos of users—can be designed to convey a person's information health status

[31] For more information on OP technology, see the Originator Profile Collaborative Innovation Partnership's website at <https://originator-profile.org/ja-JP/>.

visually. These avatars could undergo physical changes based on a user's information consumption patterns, fostering greater awareness. For example, an avatar might appear overweight if the user consumes only sensational content or sickly if their information is unbalanced. Conversely, a balanced intake of information would result in a considerably healthy-looking avatar.

(5) Providing a Digital Diet Plan

When a user detects an issue with their information health, it would be beneficial to offer tools enabling them to adjust a DP's degree of personalization independently. If the user, through an information checkup, becomes aware of problem of how they consume information and desires to transform that behavior, features that support this transformation should be provided. For severe cases, users may proactively consult specialists to manage their information intake, akin to dietary management in health care.

However, as with physical diet plans, not all digital diet plans are suitable or beneficial. Caution is necessary to avoid manipulative tactics that exploit deceptive phrases such as "think for yourself" and "your perspective is biased" to drive users toward even greater unhealthiness. Additionally, measures must guard against "ideological rectification" and "brainwashing" disguised as digital diet plans, with the government prohibited from serving as practitioners in such systems. To ensure trustworthiness, eligibility to implement digital diet plans could be certified, similar to how health guidance is provided by licensed professionals.

The digital diet plan is intended to achieve information health, not restrict information consumption. It is distinct from a "digital detox," in which a person disconnects from engaging with information altogether.

(6) Exploring a Replacement for the Attention Economy

Public discourse spaces today are dominated by the attention economy, an economic model that ensnares DP operators, users, traditional mass media, and advertisers. The various issues discussed thus far can largely be considered inevitable results of this system.

To address these issues fundamentally, an alternative economic structure must be sought. While this radical endeavor would undoubtedly face considerable difficulties in theory and practice, it is unavoidable as long as the root cause of the problem lies in the economic structure.

The exact nature of the alternative economic structure is not clear at present and must be explored through cross-disciplinary debate that includes economic perspectives. Initial efforts should focus on investigating the underlying subsystems that sustain the attention economy. Specifically, content should be assessed based on quality rather than simplistic indicators such as PVs, with compensation aligned with these assessments.

In the medium term, just as markets hold food manufacturers accountable for consumer health, they may criticize—and, in some cases, reject—companies (such as DPs) that disregard users' information health in pursuit of profit. This shift could be promoted through literacy campaigns, media coverage, and the integration of information health concepts into frameworks such as the Sustainable Development Goals and environmental, social, and governance criteria.

III. Basic Principles for Users

(1) Understand the Current Information Environment

Users should familiarize themselves with the attention economy model and recognize their embeddedness within it. This is the first step toward addressing associated challenges.

(2) Be Aware of Information Health

Users should treat their information health with the same importance as their physical health. Cultivating this awareness offers the following benefits.

① Building Immunity to Fake News

An overly selective information diet diminishes immunity and increases susceptibility to fake news and other misinformation. By being aware of their information health and consuming a well-balanced diet of information that includes opposing opinions, users can acquire "immunity" to fake news.

② Enhancing Mental Health

Consuming only specific types of information can lead to rigid thinking and an inability to accept others' ideas. Further, documents referred to as the "Facebook Files" point out the potential for curated and exaggerated content, such as unrealistic body images, to exacerbate feelings of inadequacy and harm mental health, particularly among younger users[32]. Awareness of information health can mitigate these risks.

Selective and fragmentary information diets can fuel harmful behaviors, such as slander. These behaviors damage victims' mental health and can harm the perpetrator's well-being and life. By achieving information health, individuals can avoid becoming perpetrators or victims of slander.

③ Serendipity and Opportunities for Choice (Self-Actualization)

Living within a filter bubble reduces the likelihood of serendipitous encounters with diverse perspectives, potentially limiting opportunities for other life choices. Increasing awareness of information health and proactively consuming different kinds of information help individuals escape these bubbles, opening themselves to new opportunities.

④ Strengthening the Health of Democracy

Democracy is built on the premise of citizens engaging in "dialogue" based on System 2-thinking (deliberation) and independently and autonomously engaging in political participation. Achieving information health supports these democratic ideals by facilitating exposure to diverse views and opinions and reducing informational biases.

If democracy deteriorates further, the freedom we currently enjoy could be jeopardized. For example, a lack of checks over authority could erode our basic human rights. Thus, promoting information health is essential for personal well-being and the preservation of democratic society in the medium and long term.

[32] Regardless of Facebook's awareness of these mental health effects, they were clarified when Frances Haugen disclosed internal documents that the company had neglected taking countermeasures <https://www.wsj.com/articles/the-facebook-files-11631713039>. When barriers to the use of depression-related words are lowered, people are more likely to be labeled as depressed, and when that label becomes part of their identity, feelings of self-denial are heightened. For more on this view, refer to Greg Lukianoff et al. *The Coddling of the American Mind*, trans. Yukiko Nishikawa (Soshisha, 2022), 210. Furthermore, according to the US Department of Health and Human Services, as the brain is developing in early adolescence and is particularly susceptible to social pressure, peer opinions, and comparisons among peers, there is a significant risk that frequent use of social networking sites that facilitate interaction among people with common interests and values negatively impact young people's mental health.
<https://www.nikkei.com/article/DGXZQOGN23DAA0T20C23A5000000/>

IV. Basic Principles for Operators

1. Basic Principles for DP Operators

(1) Status of Independent Initiatives by DP Operators[33]

To date, various DPs have taken steps to contribute to information health. In Japan, notable efforts include Yahoo! News' implementation of models for ranking constructive comments[34] and diversifying comments[35] in the comments section, which aim to create a platform supportive of inclusive discourse. Moreover, the trade association Safer Internet Association is actively addressing challenges faced by DPs.

The following section presents some essential ground rules to encourage further initiatives by DP operators.

(2) Safeguarding Cognitive Liberty

DP operators must respect their users' cognitive liberty[36] and neurorights[37], avoiding the use of AI or other technologies to manipulate individuals' thoughts in ways contrary to their interests—a process known as "mind-hacking," which should be prohibited.

The concept of "cognitive liberty" originated in the field of neurolaw and refers to the protection of the integrity of the central nervous system, especially the brain. Associated debates have included concerns about measuring neural activity through magnetic resonance imaging and electroencephalography, as well as electromagnetic interventions in the nervous system. Nevertheless, when the term is expanded to incorporate cognitive processes in the broader sense, it encompasses external environmental elements, such as digital spaces and information architecture, which can influence individuals' cognition. Given the brain's plasticity, ongoing exposure to mind-hacking can cause changes to the nervous system. DP operators must refrain from using advances in cognitive neuroscience to infringe upon users' cognitive liberty.[38]

Additionally, the European Union (EU) has announced the total ban of "the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behavior."[39]

---

[33] Numerous DPs exist, and a uniform approach should not be applied to all of them. DPs, as discussed here, are envisioned as considerably large online platforms with aspects of social infrastructure, as described by the DSA (see footnote 30).

[34] See Yoshimune Tabuchi, et al, "Introduction of a Ranking Model for Constructive Comments in Yahoo! News," in Proceedings of the Twenty-fifth Annual Meeting of the Association for Natural Language Processing, (March 2019), pp. 1371 and following <https://www.anlp.jp/proceedings/annual_meeting/2019/pdf_dir/P7-33.pdf>.

[35] Yahoo Japan Corporation press release, "Yahoo Launches 'Comment Diversification Model' Using Proprietary AI that Makes it Easier for More Diverse Opinions to Appear Higher in the Comments Section," April 18, 2023.<https://about.yahoo.co.jp/pr/release/2023/04/18a/>

[36] Masatoshi Kokubo, "Introduction to the Study of 'Cognitive Liberty": Neuroscience and Constitutional Studies," Journal of Law and Politics, 125 (2020), pp. 375 et seq

[37] See Yu Mizuno, "The New Social Contract [or What Replaces It]: 'Neurorights' and the Last Secret of the Inner Mind," WIRED, 48 (2023), p. 149.

[38] Some studies suggest that the novelty of information may be a factor in why people are attracted to fake news. Conversely, the risk of interference with an individual's cognitive processes by using neuroscience findings on human cognition and behavior should also be monitored. Grignolio, A., Morelli, M. and Tamietto, M. (2022), Why is fake news so fascinating to the brain? Eur J Neurosci, 56: 5967-5971. <https://doi.org/10.1111/ejn.15844>

[39] European Commission, "Proposal for a Regulation of the European Parliament and of the Council: Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts," Article 5(a) <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=75788>

(3) Filter Bubble Measures

DP operators must develop methods for delivering content that exposes users to a well-balanced range of information.

While the delivery of personalized information does have advantages—users can efficiently access the desired content and operators have higher click and engagement rates[40]—it also carries risk. By presenting users with only information deemed relevant to their preferences, operators may inadvertently isolate them from information they do not actively seek, reinforcing filter bubbles.

Accordingly, operators should implement features that provide users with the option to access a broader range of content, such as by randomly providing them with non-personalized information. Furthermore, DP operators should equip users with tools to manage their information health proactively by adjusting personalization levels of recommendation algorithms, except in extraordinary cases. As a prerequisite, DP operators should also ensure users can view and understand the degree of personalization influencing their current information displays at any time.

Such measures align with global regulatory trends, including the EU's *Digital Services Act*. Article 29 requires considerably large online platforms to provide recommender systems that are not based on profiling, provide the ability to select and change preferred options for recommender systems, and facilitate access to these functions.

(4) Fake News Countermeasures

To mitigate the spread and impact of fake news, DP operators must adopt effective strategies. Possible countermeasures include the following:

① Defining the Scope of Fake News

A critical step for dealing with fake news is establishing a clear definition for it. Fake news includes "disinformation" and "misinformation," which are defined as false or misleading information transmitted with and without malicious intent, respectively. As both can have a significant negative impact on society, countermeasures should target both types. Given the vast amount of fake news that exists, prioritization is key. Efforts should focus on combating fake news that has the greatest social repercussions.

② Technological and Architecture-Based Solutions

Building technological solutions into platforms is important for curbing the spread of fake news. Effective approaches have already been implemented by several companies (e.g., Twitter displaying links to articles in retweets that encourage users to view the site in question[41]), which will hopefully spread to other companies. Further, given the constant technological advancement, particularly fake text, images, and videos created by generative AI, countermeasures should be continuously updated to keep pace.

③ Publication of Policies Concerning Media Handling

DP operators must be transparent and accountable in disclosing their policies on how they manage media content. This includes clarifying the media's response to the disclosure of the policy. Operators

[40] According to data revealed by Facebook in 2013, the total number of posts a user could see from friends and followers averages 1,500 per day, but the algorithm narrows it down and displays an average of 300 posts. This means that only 20% of all posts are displayed. See, Backstrom, L., "News Feed FYI: A Window Into News Feed," Facebook, 2013 <https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed>

[41] Twitter Japan, September 25, 2020, <https://twitter.com/TwitterJP/status/1309289971690348545>

should collaborate with third parties, such as non-profit organizations and academic researchers, when formulating these policies to ensure their objectivity and fairness.[42]

④ Rules for Removal of Harmful Content

In cases involving severe disinformation or repetitive harmful posts by the same account (including bots), removal measures may be necessary, such as deleting the content or placing restrictions on the account in question. Nevertheless, these measures should balance the need to address the issue with the protection of freedom of expression. For example, less-restrictive measures should first be attempted, such as architecture-based countermeasures (see ②) and direct warnings to users.

⑤ Reducing Incentives for Creating Fake News

To discourage the commercial creation of fake news, targeted measures should ensure that reliable information is accompanied by appropriate advertising while minimizing revenue for websites that propagate harmful or false information.

⑥ Ensuring Transparency and Accountability

Given their substantial market influence, DP operators should clarify the standards and processes they use to counter fake news. Transparency regarding the measures taken and their effectiveness is essential.

⑦ Responding to Regional Characteristics

As global companies, DP operators must tailor their strategies to align with the language, culture, and geopolitical contexts of the specific areas they serve. This includes region-specific adaptations and the localized disclosure of information.

⑧ Cooperation with Fact-Checking Organizations

Fact-checking is essential in the fight against fake news. Collaboration with neutral, third-party fact-checking organizations is critical for addressing fake news while maintaining objectivity. The following are some possible countermeasures:[43]

a) Creating Mechanisms to Promote Awareness of Facts

DP operators should facilitate easy access to fact-checking results, for example, by integrating a prominent mechanism to present information fact-checked by a third part alongside disputed content or display verified information directly to those who view fake news.

b) Supporting Fact-Checking Organizations

Sustained activity by fact-checking organizations will help improve the information environment. Moreover, given the rapid pace of technological advancements, these organizations must have access to the latest tools and methodologies. Thus, DP operators should support fact-checking organizations,

[42]The Journalism Trust Initiative (JTI), a project led by Reporters Sans Frontières aiming to create a healthy information space, has created a framework for determining and disclosing the trustworthiness of media outlets through self-assessments and external audits by certification organizations. The self-assessments consider such factors as disclosing ownership-related information, including board members and revenue sources, and editorial guidelines, editorial structure, and accountability processes. These evaluation standards were published in 2019 as a non-binding workshop agreement (CEN Workshop Agreement [CWA]) by the European Committee for Standardization (CEN). See CWA No. 17493 <https://www.jti-app.com/footer/cwa>. In Japan, the Institute for the Next Generation of Journalism and Media at Waseda University has assessed the disinformation risk for 33 major news media in Japan using the Global Disinformation Index, which employs JTI's assessment criteria <https://www.disinformationindex.org/files/gdi_japan-japanese-mmr-report-online.pdf>.

[43] The issue of political neutrality of fact-checking organizations is a globally identified issue. Through supervision of fact-checking organizations by third-party organizations, sufficient fairness and transparency should be ensured in the fact-checking itself, and DP operators should decide who to collaborate with based on these results. In connection with this, the International Fact-Checking Network (IFCN) has identified five principles in its Fact-Checking Code of Conduct, which fact-checking organizations are encouraged to follow: ① Nonpartisanship and Fairness; ② Transparency of Sources; ③ Transparency of Funding and Organization; ④ Transparency of Methodology; ⑤ Open and Honest Corrections Policy.

enabling them to sustain and enhance their operations.

(5) Obligation to Switch Algorithms in Exceptional Situations

DP operators should adopt alternative display algorithms in the following exceptional situations to protect the public interest.

① Elections Mode

During key democratic events such as public office elections and referendums on constitutional amendments, there are particular demands for users' information health to enable informed and fair decision-making by voters. DP operators must prioritize content that supports appropriate voting actions, such as materials on party and candidate policies and major electoral issues.

The *Public Offices Election Act* has introduced various regulations related to election campaigning, including restrictions on the use of campaign sound trucks and banning door-to-door canvassing. These rules are based on the understanding that election campaigns "should be considered not as opportunities in which candidates having diverse arguments can freely compete with each other under minimum necessary constraints but as opportunities in which candidates conduct campaigns in accordance with rules that are set to ensure the fairness of the election."[44] Moreover, certain broadcasters are mandated to transmit political campaign materials so that eligible voters can freely and equally obtain information regarding political parties and candidates.

DP operators should also temporarily move away from the attention economy model during elections and referendums to adopt algorithms and UI/UX that emphasize fairness. Specific actions include the following: the broad, fair dissemination of information on political parties and candidates, similar to political broadcasts; prioritization of content from highly reliable media outlets; implementation of architecture to control the rapid spread of misinformation; regulation of micro-targeted political advertising; establishing specific guidelines for election periods, with compliance monitored.

② Disaster Mode

In the event of a disaster, DP operators must swiftly adjust their display algorithms to prioritize accurate and actionable information, particularly in the case of evacuations. Specifically, they must display information from credible sources, including the government, local public entities, and news organizations, and communicate evacuation directives, orders, and details of evacuation sites. They must also prevent the rapid spread of malicious disinformation that attempts to take advantage of the confusion, as well as panic-inducing information that could complicate evacuation efforts.

③ Pandemic Mode

In public health emergencies, such as the outbreak of an infectious disease, DP operators must activate "pandemic mode" to deliver accurate information promptly and encourage appropriate behavior. As with the disaster mode, preventing the spread of false or unverified information is critical. Information necessary for individual decision-making, such as vaccinations, must be delivered appropriately (informed consent) to drive System 2-thinking, as in the election mode.

④ Protection of Minors Mode

Minors face unique vulnerabilities in digital environments due to their limited decision-making capacity and susceptibility to harmful content. Accordingly, to protect minors while respecting their rights, DP operators should implement age-appropriate paternalistic algorithms that prioritize their information

[44] Supreme Court decision prohibiting door-to-door canvassing, Masami Ito's concurring opinion (Supreme Ct. Third Judgment, 35 Keishu 5, at 568, July 21, 1981).

health. Seamless switching between algorithms corresponding to age (e.g., junior vs. high school students) and settings set by the minors or their guardians should be facilitated.

⑤ Contingency Mode

In the era of "hybrid warfare," which refers to the combination of conventional military strategy with information warfare, platforms operated by DP operators can become battlegrounds, and special measures must be taken. For example, states that commit acts violating international law must be objectively identified, and users must be prevented from being unwittingly manipulated or complicit in disseminating fake news. Additionally, users in war zones should be proactively provided with information from trustworthy sources necessary to ensure their safety. DP operators should draft and preemptively publicize their crisis response guidelines. This can involve establishing a specialized internal department to manage information warfare that includes security experts.

(6) Ensuring Algorithm Transparency

DP operators should disclose the parameters and methodologies underpinning their algorithms that determine how content and other information are displayed. This will clarify DPs' values and priorities regarding information health, potentially incentivizing competition that could improve overall information health. For example, DPs found to be using malicious algorithms to manipulate users should be subject to public criticism and market consequences. Transparency will also allow the media and journalists to make informed decisions regarding the DPs with which to collaborate.

DP operators should also regularly draft and publish reports on algorithm rules and operational status to guarantee transparency regarding how they display content and manage advertisements.

(7) Systems for Responsibility and Governance

① Editorial Autonomy

DP operators manage content and information through the collaboration of various departments, including those that develop algorithms, select articles and advertisements, and design UIs. To maintain editorial independence, operators should establish an internal system that ensures that the editorial team operates separately from management.

This structure creates a firewall between management and editorial functions, preventing undue interference from management or external stakeholders in content creation and distribution.

② Establishing an Ethics Committee

DP operators should form ethics committees comprising outside experts to ensure their practices respect human dignity and individual rights, particularly regarding their collection and analysis of personal data and any "nudges" made by their UI/UX design. These committees should also guarantee that practices from behavioral psychology, cognitive neuroscience, and similar fields are applied ethically and responsibly.

③ Establishing a Content Review Committee

DP operators should establish internal content review committees that include outside experts. These committees would oversee various aspects, such as when articles and content are displayed or removed, how sources are vetted (i.e., media screening), how advertisements are assessed, and how public relations content is presented.

Furthermore, if an operator who is also a content source files a complaint regarding issues such as the suspension of distribution or the outcome of a media screening, the content review committee should have the authority to adjudicate the matter. The committee must be given sufficient autonomy to ensure impartial decision-making.

④ Duty to Understand Actual Conditions

DP operators must accurately understand the issues arising from their platforms. They should regularly publish the results of any surveys they conduct, as long as doing so does not infringe on business confidentiality. Where possible, detailed survey results, including raw data, should be shared with news organizations and researchers.

Moreover, DP operators should strive to quickly identify and address any risks they may face, for example, by establishing an internal reporting system.

2. Basic Principles for the Media

(1) Principles for All Media

Traditionally, mass media, such as newspapers and broadcasters, have played a significant role in providing information to support democratic societies. However, with the rise of social media and other platforms, the sources of information have diversified, and algorithms and AI influence how information is presented and shared. This shift has made achieving information health more difficult. All media actors should acknowledge their responsibility in shaping the information ecosystem and disclose their policies regarding how they select and share information in an easy-to-understand manner.

(2) Principles for Mass Media

As powerful influencers, mass media organizations must be aware of their responsibility in shaping the information landscape that underpins democratic societies. As leaders in the media ecosystem, they must set an example by fostering trust and credibility, which serves to support a healthy democracy[45]. To preserve this trust, mass media outlets must consistently deliver high-quality, fact-based reporting and adhere to rigorous editorial practices (e.g., a peer review system). They should also maintain strong internal discipline, focus on the integrity of their coverage, and prioritize substance over sensationalism. Achieving these goals requires efforts to enhance the ethics and professionalism of journalists and editors.

① Publishing Policies

Considering their significant influence, mass media organizations must transparently disclose their policies on information selection and dissemination, including whether they engage in original reporting or rely on tools such as generative AI when preparing news. To earn public trust, media organizations should collaborate with third parties, such as nonprofits and researchers, to evaluate the content of publications, ensuring objectivity and fairness in their approach.

② Distancing from the Attention Economy

The attention economy prioritizes eye-catching or sensational content designed to maximize viewership and PVs. This focus on entertainment value and highly stimulating material can trigger System 1 responses, leading to a decline in the quality of news and broadcast content and the overall credibility of the mass media.

[45] In its decision in the Hakata Railway Station case (Supreme Ct. decision, 23 Penal Code 11, at 1490, November 26, 1969,), the Supreme Court stated its position that “news reports by the press provide important information for people to make decisions regarding their involvement in national politics in a democratic society.”

Given their public role, mass media must strive to distance themselves from the pressure of the attention economy, particularly when covering matters of significant public interest.

However, public media, which operate with stable, non-advertising revenue streams, are less influenced by the attention economy and tend to produce higher-quality content. To support this media structure, public media must be integrated into the routines of daily life, ensuring their content remains visible and accessible (i.e., a prominence mechanism)[46].

③ Preventing Feedback Loops

While the spread of fake news is often associated with social media, it is increasingly evident that mass media coverage of social media trends can amplify the dissemination of fake news, creating a feedback loop[47]. News programs and online platforms frequently highlight trending topics on social media, which requires heightened caution to avoid perpetuating misinformation[48].

Given their enduring influence, mass media have a key role in preventing the spread of fake news. Beyond refraining from publishing misinformation, they should actively work to correct it through fact-checking and other means.

3. Basic Principles for Telecommunication Carriers

(1) Principles of the Open Internet

The expansion of the Internet and the resulting "big bang" of public discourse spaces have effected numerous challenges, including the global polarization of knowledge and societal division, which can no longer be ignored. Given this situation, it is essential to revisit the ideals of the early Internet as a pathway to a democratic and open world and reassess the principles of the "Open Internet." A critical priority is minimizing incentives and mechanisms that allow telecommunications carriers and DP operators to promote or suppress information selectively for commercial purposes. This would ensure that all Internet users can freely access content and information without bias. These principles support Internet neutrality[49] and are crucial to information health.

(2) Zero-rating

Telecommunications carriers, particularly in mobile communications, have adopted pay-as-you-go or capped flat-rate pricing models. However, some carriers offer "zero-rating" services, which exempt certain content from data usage limits, and these are offered worldwide. In Japan, this type of service is referred to

[46] Ofcom (the UK's regulator for broadcasting and communications) has made recommendations to certain platforms to make public media content more prominent.

[47] Claire Wardle, from the US nonprofit First Draft News, illustrates the fake news dissemination process in her "Trumpet of Amplification" graphic, positioning professional media as the last step. See Claire Wardle, "5 Lessons for Reporting in an Age of Disinformation," First Draft, 2018. <https://firstdraftnews.org/articles/5-lessons-for-reporting-in-an-age-of-disinformation/>

[48] One example is the false rumor of toilet paper hoarding during the COVID-19 crisis. See Fujio Toriumi, "The Essence of the Problem of False Rumors Spread through SNS," *NII Today* 89, September 2020. <https://www.nii.ac.jp/today/89/4.html>. Regarding the impact of online news and portal sites, see Yoko Nakamura, "The Fake News Problem Goes beyond the Scope of Self-Responsibility: Interview with Hiroyuki Fujishiro, Professor of Sociology at Hosei University," *Toyo Keizai Online,* October 2021. <https://toyokeizai.net/articles/-/575453>.

[49] There is ongoing debate regarding the definition and scope of network neutrality, but there is a consensus that the standard should cover the following four principles: 1) Users should be able to use the Internet flexibly and access and use content and applications freely; 2) users should be able to provide content and applications freely to other users; 3) users should be able to connect to and use the Internet freely using devices that meet technical standards; 4) users should be able to use communications and platform services in a fair and equitable manner at reasonable prices. See Ministry of Internal Affairs and Communications, *Guidelines Concerning Application of the Telecommunications Business Law to the Provision of Zero-rating Services* (March 2020), 2.

as “count-free services”; nevertheless, users who sign up will receive preferential telecommunications rates only when they access eligible content and information.

While zero-rating provides financial benefits for users, it can influence competition among content providers. Carriers may prioritize content that attracts more users, enabling these providers to gain a competitive edge. Consequently, lured by the financial benefits of zero-rating, users are potentially limiting their means of accessing diverse information. Moreover, if zero-rated content aligns with the attention economy’s focus on high-engagement material, telecommunications carriers are inadvertently contributing to its negative consequences.

Therefore, this practice raises concerns about whether zero-rating violates the principle of network neutrality (i.e., the Open Internet). Many countries in Europe, in particular, have taken a considerably tough stance on zero-rating. For example, in September 2020, the European Court of Justice ruled (in a prior ruling) that zero-rating services, which exempt certain apps from data usage or apply discriminatory speed limits, are prohibited under Article 3 of the EU Network Neutrality Regulation[50]. The Court clarified that zero-rating must be evaluated on a case-by-case basis but held that practices outside the three traffic management exceptions in Article 3(3)[51], whether based on contract or mere business practice, infringe on the principle of non-discriminatory treatment and the user’s right to free Internet access. Thus, even seemingly convenient services can infringe on users’ rights, requiring ongoing vigilance.

(3) Blurring the Boundaries between Communications and Broadcasting

Broadcasting has traditionally been defined as a medium that “directly and instantaneously transmits information to a large, unspecified audience across the country simultaneously, exerting a strong social influence (special social impact) compared to other information and communication media”[52]. This “special social impact” has been cited as a key justification for the distinct legal structure governing broadcasting[53]. However, the immediacy and simultaneous nature that once distinguished broadcasting from other forms of communication are no longer unique because of the emergence of broadband Internet and subsequent content distribution services (“communications open to the general public”). Consequently, the degree of “social impact” has been reemphasized as a criterion for distinguishing between broadcasting and communications. In this regard, the Ministry of Internal Affairs and Communications’ report *Study Group Report on a Comprehensive Legal Framework for Communications and Broadcasting* identified the following characteristics for assessing social impact: ① type of content (e.g., video, audio, data), ② quality of the service (e.g., screen resolution), ③ ease of access via terminals, ④ audience size, and ⑤ distinction between paid and free services, which must be identifiable by appearance.

Nevertheless, the report did not thoroughly explore how these indicators can effectively function as measures of social impact or whether the concept of social impact remains a valid basis for considering broadcasting an independent category. As analysis increasingly relies on externally measurable quantitative indicators for determining the social impact of broadcasting, the boundary between traditional broadcasting and Internet-based media (e.g., online video content) becomes increasingly blurred. For instance, according to a survey, for younger audiences, the Internet has already surpassed television in importance[54]. Furthermore,

[50] Joined Cases C-807/18 and C-39/19 Telenor Magyarország Zrt. v Nemzeti Média- és Hírközlési Hatóság Elnöke.

[51] ① The action must be in accordance with laws and regulations and ② necessary for network security and ③ network congestion management.

[52] Ministry of Internal Affairs and Communications, *Study Group Report on a Comprehensive Legal Framework for Communications and Broadcasting* (December 2007), 17.

[53] Of course, this is not the only legal basis for broadcasting. In addition to the social impact theory, there are various other theories, such as the theory of spectrum scarcity, the theory that the airwaves are public property, and the theory of program uniformity.

[54] See, for example, Institute for Information and Communications Policy, Ministry of Internal Affairs and Communications, *FY 2018 Survey Report on Information and Communications Media Usage Time and Information Behavior* (September 2019)

the distribution of TV broadcast programs via the Internet—defined as an Internet service rather than broadcasting under the *Broadcasting Act*—is expected to grow further as a service similar to broadcasting. In practice, the social functions of broadcasting, Internet Protocol (IP) broadcasting, and "Internet TV" (telecommunications) are converging, leaving many viewers unable to distinguish between them. Few viewers, aside from experts, can accurately differentiate what constitutes broadcasting versus telecommunications in the various types of Internet video distribution. With the expected improvements in the quality of broadcast-like Internet transmissions, such as simultaneous broadcasts over the Internet, the distinctions between these media will likely blur even further. This makes it increasingly difficult to establish clear criteria for distinguishing between broadcasting and telecommunications. Moving forward, it will be key to reconsider the legal basis for media through discussions that are not bound by conventional definitions and align with contemporary realities.

4. Basic Principles for Advertisers

A major challenge for advertising businesses in promoting information health is minimizing mismatches between content and advertisements. For advertisers, this involves presenting creative elements in advertisements, such as catchphrases, design, color schemes, music, and dialogue, in an appropriate way for the user.

In traditional media, such as television, radio, newspapers, and magazines, advertisers often select platforms based on the compatibility of their advertisements with adjacent content, ensuring a certain degree of affinity. Moreover, broadcasters and publishers independently conduct reviews of advertisements, with media operators bearing responsibility for the creative expression and making ethical judgments. These practices help maintain the medium's distinct characteristics—such as the quality of programs and publications, cohesion of creative expressions and involvement of notable personalities, and the separation of content and advertising, which fosters audience acceptance—and should be prioritized in the Internet and metaverse spaces, as well.

To achieve information health in advertising, it is necessary to develop a system that evaluates content based on quality rather than simplistic indicators such as PVs. Compensation models should also be aligned with such evaluations. For example, a system that certifies media outlets and publishers with transparent editorial policies and reporting methods could enhance the value of advertising spaces associated with certified content (refer to the previously mentioned OP technology for more information). Advertisers could target certified media and publishers to ensure a baseline of credibility, helping curb the spread of fake news and potentially strengthen their positive brand image by contributing to consumers' information health.

Additionally, new indicators such as content match rate could be created to evaluate the alignment between advertisements and the content they accompany. In the future, it may be possible to determine ad placement based on the characteristics of the hosting content and an analysis of the language used in the content. If advertisements aligned with content can be delivered to more relevant users at the precise time, it could help prevent adverse effects such as provocative content and damage to brand image in the attention economy.

Although these are illustrative examples, they emphasize the need to move beyond basic metrics such as PVs and time spent on a site. Instead, a system should be built to evaluate and reward content that contributes to information health, incorporating these advanced indicators into the advertising model and appealing to advertisers and users alike. By building a business model that utilizes these new indicators, it would also be possible to reevaluate media and advertising profit distribution.

<https://www.soumu.go.jp/main_content/000644168.pdf>. According to this report, the importance of TV, Internet, newspapers, and magazines as information sources was 88.3%, 77.5%, 59.5%, and 18.8%, respectively, for all age groups, while the importance of the Internet exceeded that of TV for those in their 20s and 30s.

To achieve information health, society as a whole, including users, must recognize the social value of responsible advertising. Advertising businesses should promote efforts in this direction, taking a backcasting approach to discuss and implement necessary actions for a better future information ecosystem.

V. Basic Principles for Government

(1) The Constitutional Obligation to Provide "Lateral" Assistance to Information Health Initiatives

The government has the constitutional responsibility to help people maintain "the minimum standards of healthy and fulfilling living" (Article 25) and ensure their "right to know" (Article 21). This includes fostering media diversity and facilitating the distribution of varied information. Consequently, one can reasonably presume that it has the constitutional duty to create a competitive environment that supports a plurality of media and provide "lateral" assistance to diverse initiatives by DP operators for enhancing information health.

During critical democratic periods, such as public office elections and constitutional referendums, citizens must access a well-balanced variety of information and opinions. In these moments, citizens must also develop immunity—critical-thinking abilities—to fake news. Unlike at other times, the government is called upon to work toward maintaining healthy, deliberative public discourse by, for example, encouraging DP operators to give special consideration to their users' information health.

(2) Prohibition of Direct and Excessive Government Intervention

The government should neither directly intervene in shaping the public's information health nor impose such health on them. It should also not impose its interpretation of what constitutes this health. Such direct intervention could invite "information health fascism" and distort democratic principles. This caution cannot be overstressed.

Similarly, excessive intervention concerning DP operators should not be permitted. These operators are private businesses entitled to the freedom of expression (Article 21) and occupational choice (Article 22). Overreach by the government could effectively "domesticate" DP operators, leading to censorship.

(3) Ensuring DP Operator Transparency and Accountability

As outlined earlier, DP operators must guarantee transparency in their operations and be accountable to the public (see IV.1.(6)). The government should design systems that effectively enforce these obligations.

Moreover, how DP operators consider users' information health could stimulate competition among DP providers (see II.(6)). To facilitate this, users need clear information on DPs' efforts to achieve information health. Legislative measures should be considered, including mandating transparency through information disclosure, akin to the *Act on Improving Transparency and Fairness of Digital Platforms* and other regulations[55]. Transparency in DP operations was regulated to a certain extent in June 2024 with the *Act on Promotion of Competition for Specified Smartphone Software* (commonly known as the *Smartphone Software Competition Promotion Act*), building on the achievements of the DP Transparency Act[55]. These mark important steps toward ensuring accountability[56].

(4) Maintaining Just and Fair Dealings and a Competitive Market Environment

To prevent public discourse from being dominated by the attention economy, which prioritizes stimulating content designed to elicit reflexive responses, content from newspapers and other trusted media must be

[55] e.g., the *Act on the Protection of Consumers Who Use Digital Platforms for Shopping*.

[56] The *Smartphone Software Competition Promotion Act* stipulates that, when displaying search results, a company shall not give priority to its own services over those of other competing companies without justifiable cause (Article 9), as well as the obligation to disclose data management systems (Article 10), to provide data portability tools (Article 11), and to disclose changes in operating software and browser specifications (Article 13).

substantially distributed on DPs in a way that enables users to engage with it meaningfully. To this end, traditional media must have a certain degree of bargaining power concerning DP operators. In cases where media organizations lack this bargaining strength, the government must create an environment for just and fair dealings between DP operators and the media.

Furthermore, factors other than price (e.g., media pluralism) must be considered when evaluating the adverse effects of oligopoly and monopolization of DPs, including social networking services. For example, in the United States, when reviewing mergers in traditional media (e.g., newspapers and broadcasters), the Bureau of Competition does not focus on increased surplus, and the price consumers do not pay for content post-merger. Rather, one of its objectives is to ensure media diversity, which guarantees that citizens have free access to diverse information and viewpoints, empowering them to form and share their opinions without undue influence from dominant perspectives. As the digital media market grows more concentrated, the undermining of media pluralism, restricted diversity of opinions, and the erosion of the free market of ideas should not be overlooked from competition law and policy perspectives. The increasing centralization of the social media market may also exacerbate self-censorship, which, in a broad sense, can be considered consumer harm and a violation of competition law. In this age of overwhelming information overload, the attention and time we can devote to the information supplied on the market are decreasing, while our information intake is becoming increasingly isolated and one-sided. The distortion of the free Internet space/market must not be allowed to weaken democracy. Thus, whether competition and consumer policies play a role in this process must be reexamined.

(5) Promoting ICT Literacy

Achieving information health requires users to develop information and communication technology (ICT) literacy. Therefore, the government must actively work to improve the ICT literacy of its citizens related to information health.

Among the different aspects of ICT literacy, AI literacy will be particularly important in the future. AI literacy includes various skills and knowledge, including recognition, understanding, application, analysis, evaluation, and creation[57]. This includes recognizing the existence and application of AI and understanding the basic principles and algorithms governing it. It also entails educating users and companies so that they can critically evaluate AI technology and its potential biases, risks, and ethical implications. Finally, AI literacy extends to the ability to engage actively with AI systems, including developing AI-based solutions and participating in discussions surrounding AI policy and regulation.

[57] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao, "Conceptualizing AI Literacy: An Exploratory Review," *Computers and Education: Artificial Intelligence* 2 (2021): 10041.

VI. Prospects for the Future

Recent violent incidents, such as the attack on the Capitol in the United States and the attack on the Three Powers Plaza in Brazil, highlight the role of echo chambers in undermining democracy. These events illustrate how chaos in public discourse spaces and the information environment, exacerbated by the excesses of the attention economy, can harm democratic systems. Furthermore, the continuous stimulation of System 1-thinking by the attention economy appears to be fundamentally changing human communication formats and the essence of human behavior. The emergence of generative AI has compounded this effect, making achieving information health increasingly critical and urgent.

However, the pursuit of information health is not without risks. Historically, the concept of "health" has been misused, as seen with totalitarianism and Nazi Germany. Therefore, this statement seeks to provide an inclusive definition of information health, and it is essential to remain vigilant of the dangers of totalitarianism and develop this concept with individual freedom as its foundation. Thus, specialists in various fields—including but not limited to law, economics, social psychology, psychiatry, education, history, linguistics, and neuroscience—must collaborate and engage in cross-disciplinary dialogue to ensure the appropriate realization of information health.

Many DP services, which are key to achieving information health, operate across national borders. Thus, the challenges of maintaining a healthy information ecosystem are not confined to those within Japan. Therefore, global countermeasures, including international cooperation, are essential for the worldwide implementation of information health.

This statement identifies several issues that must be addressed to advance the concept of information health:

① Developing a clear understanding of the "nutritional elements" of information and establishing standardized methods and criteria for labeling them.

② Identifying effective methods for providing meta-information.

③ Designing an economic model and advertising systems that align with information health.

④ Determining which of the many DP operators should be particularly responsible for fostering information health.

⑤ Clarifying the nature of human intelligence and creativity while assessing the use of generative AI.

Further, to achieve information health, the following specific measures are planned for future development and implementation.

Technological Measures

① Building and implementing systems to measure users' level of information health and present personalized results.

② Creating and implementing systems to re-experience a selective information diet virtually.

③ Designing and implementing avatars that allow people to recognize their information health levels.

④ Developing algorithms to ensure that diverse information is presented in a balanced manner.

⑤ Conducting research and development on methods for measuring human biological responses to information health.

Social and Institutional Measures

① Developing methodologies and teaching materials for information literacy that incorporate the aims of information health.

② Investigating the cultivation of social norms regarding information health based on comparisons with the history of nutrition education (i.e., dietary literacy).

③ Defining principles or guidelines for the appropriate use of generative AI in alignment with information health objectives.

④ Designing incentives to encourage the implementation of technologies that support information health.

⑤ Researching new broadcasting concepts and systems based on the aims of information health.

⑥ Promoting international collaboration with universities, research institutes, and international organizations, such as the World Health Organization, to advance information health initiatives.

Hybrid Technological and Institutional Measures

① Supporting the development and implementation of OPs.

② Designing and implementing systems for ad verification.

③ Developing and implementing recommendation systems that prioritize socially beneficial content.

④ Creating platforms that enable users to engage with information responsibly (e.g., a system that awards points to users who consume information with an awareness of information health, which could then be used to support local nonprofit and other organizations featured in other content).

⑤ Holding contests for developing technologies and institutions that contribute to information health.

In the future, discussions on these measures will be expanded through cooperation with diverse stakeholders, and efforts will focus on developing and implementing these specific measures.

Measure overview	Category	Relevant academic disciplines	Parties responsible for implementation
Building and implementing systems to measure users' level of information health and present personalized results	Technological	Computational social science, sociology, media theory, statistics	DPs
Creating and implementing systems to re-experience a selective information diet virtually	Technological	Computational social science, information science, information engineering, psychology	DPs
Designing and implementing avatars that allow people to recognize their information health levels	Technological	Information architecture, design studies, psychology	DPs (including metaverse operators)
Developing algorithms to ensure that diverse information is presented in a balanced manner	Technological	Information architecture, human computation, psychology, marketing	DPs, mass media
Conducting research and development on methods for measuring human biological responses to information health	Technological	Information architecture, neuroscience, law, philosophy	Research institutions, including universities

Developing methodologies and teaching materials for literacy education that incorporate the aims of information health	Social and institutional	Education, media literacy	Public facilities, schools, DPs
Investigating the cultivation of social norms regarding information health based on comparisons with the history of nutrition education (i.e., dietary literacy)	Social and institutional	Nutrition, history	Schools
Defining principles or guidelines for the appropriate use of generative AI in alignment with information health objectives	Social and institutional	Law, information science, human computation, psychology, philosophy	Mass media, DPs, schools, universities
Designing incentives to encourage the implementation of technologies that support information health	Social and institutional	Sociology, behavioral economics, psychology, law	Mass media, DPs
Researching new broadcasting concepts and systems based on the aims of information health	Social and institutional	Law, media theory, sociology, information architecture	Research institutions, including universities
Promoting international collaboration with universities, research institutes, and international organizations, such as the World Health Organization, to advance information health initiatives	Social and institutional	All disciplines	Research institutions, including universities
Supporting the development and implementation of OPs	Hybrid technological and institutional	Information science, information architecture, media theory, law	OP technical research associations, media
Designing and implementing systems for ad verification	Hybrid technological and institutional	Sociology, information science, information architecture, media theory, marketing	Advertising agencies, advertisers, DPs
Developing and implementing recommendation systems that prioritize socially beneficial content	Hybrid technological and institutional	Media studies, information science, information architecture, human computation, law, psychology, philosophy	Mass media, DPs
Creating platforms that enable users to engage with information responsibly	Hybrid technological and institutional	Media studies, information science, Information architecture, human computation, behavioral economics, psychology, law, philosophy	Mass media, DPs
Holding contests for developing technologies and institutions that contribute to information health	Hybrid technological and institutional	All disciplines	This project, DPs, mass media, government

Appendix I: Challenges Arising from "Information Ill-health"

1. Risks of Information Ill-health

Distortions in information health can have negative societal impacts. In addition to adverse effects on elections, infectious disease control, and disaster preparedness, the most immediate risk is the incitement to violence. Furthermore, there is the risk of information warfare during armed conflict. These trends are not confined by national borders, and Japan is increasingly subjected to their effects, even at the household level.

(1) Risk of Inciting Violence

United States: On January 6, 2021, thousands of supporters of former President Donald Trump stormed the US Capitol in Washington, D.C., with some forcibly breaking into the building. This unprecedented incident resulted in five deaths and more than 900 people facing criminal charges. Trump, who had lost as the incumbent in the 2020 US presidential election, baselessly claimed that the election was rigged. The slogan "Stop the Steal" was the driving force behind the insurrection.

Brazil: On January 8, 2023, thousands of supporters of former President Jair Bolsonaro and others stormed Brazil's National Congress, the presidential office building, and the Supreme Court in Brasilia. Bolsonaro, also referred to as "Brazil's Trump," alleged election fraud after losing the 2022 presidential election, spurring the attack.

Germany: On December 7, 2022, the Federal Prosecutor General announced the arrest of 25 people suspected of attempting a coup d'état[58]. Influenced by the far-right Reichsbürger (literally "Reich citizens") and the conspiracy theory group QAnon, the group, which reportedly included judges and former military personnel, claimed that "Germany is controlled by the 'deep state'" (i.e., a shadow government).

These incidents demonstrate how claims of rigged elections, even if unfounded, can become a "crisis of democracy" for supporters who believe these narratives.

A July 2022 report by Joan Donovan and her team at the Shorenstein Center at Harvard University analyzed court documents from defendants involved in the US Capitol attack[59]. According to the report, "support for Trump" and the "rigged election" were the primary motivations for 20.62% of defendants. However, underlying these direct motivations was a pervasive sense of fear among attack participants, fueled by the "Great Replacement" conspiracy theory, which claims the social weakening of the white race due to increased immigration.

Thus, in addition to the biased information environment, distortions in information health are amplified by anxiety and fear of social change, highlighting the link between information ill-health and incitement of violence.

(2) Risk of Information Warfare

The Russian invasion of Ukraine, which began in February 2022, has highlighted the significant threat to information health from information warfare. This conflict has been marked by the widespread dissemination of disinformation and propaganda related to the armed conflict.

To combat this threat, immediately after the invasion began, 89 organizations certified by the International Fact-Checking Network launched #UkraineFacts, a website to share the results of debunked disinformation[60].

---

[58] Der Generalbundesanwalt beim Bundesgerichtshof, "Festnahmen von 25 mutmaßlichen Mitgliedern und Unterstützern einer. terroristischen Vereinigung sowie Durchsuchungsmaßnahmen in elf Bundesländern bei insgesamt 52 Beschuldigten." December 7, 2022, via Internet Archive.
<https://web.archive.org/web/20221207082300/https://www.generalbundesanwalt.de/SharedDocs/Pressemitteilungen/DE/aktuelle/Pressemitteilung-vom-07-12-2022.html>

[59] Joan Donovan, Kaylee Fagan, and Frances Lee, *'President Trump is Calling Us to Fight:' What the Court Documents Reveal About the Motivations Behind January 6 and Networked Incitement*, Working paper, Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, Technology and Social Change Project (2022) <https://mediamanipulation.org/sites/default/files/2022-07/j6_motivations_working_paper.pdf>

[60] International Fact-checking Network Signatories, "#Ukraine Facts," https://ukrainefacts.org/ (accessed on December 9, 2024)

According to the site, as of February 24, 2023, one year after the invasion, there were 2,809 verified cases of false information.

The aim of disinformation in information warfare is to cause confusion, demoralize the opposing country, and influence international public opinion.

During the invasion of Ukraine, false information proliferated rapidly. For instance, claims were spread that Ukrainian President Volodymyr Zelensky had fled Ukraine. On March 16, 2022, three weeks after the invasion began, a deepfake video was released, falsely depicting Zelensky declaring surrender, along with cyber-attacks on Ukrainian TV stations and other sites.

2. From the "Between Fact & Fiction" Interview

The COVID-19 pandemic amplified the prominence of individuals disseminating extreme opinions and conspiracy theories in Japan. These also helped create rifts between those who believed in conspiracy theories and their families.

The *Yomiuri Shimbun*'s long-running "Between Fact & Fiction" investigative team has been covering these issues continuously since 2020. The following are some examples of conspiracy-driven claims and their societal impacts compiled by the research team in the book *Information Pandemic*[61].

(1) People Disseminating Conspiracy-based Claims

In 2021, demonstrations against COVID-19 vaccinations and infection control measures repeatedly occurred throughout Japan. These events were fueled by conspiracy theories, including claims that the COVID-19 pandemic was planned and orchestrated for the benefit of an elite minority, COVID-19 is an international fraud, and vaccines are human experiments. One demonstration held in Tokyo in July 2021 drew approximately 700 participants. Demonstration organizers used social media to propagate messages such as "there is no scientific evidence that COVID-19 exists" and "vaccines contain toxic substances" to gain support.

Claims of fraud in the 2020 US presidential election also spread on Japanese social media. From November to December 2020, demonstrations were held in Tokyo and Osaka, with participants shouting, "Stop stealing votes" and "The presidential election is a battle between good and evil." Hundreds of people participated in the Osaka demonstration.

Some groups became so radicalized that their activities escalated into criminal acts. For example, in March and April 2022, co-leaders of the Yamato Q group, which had organized nationwide anti-vaccination demonstrations, were convicted of breaking and entering vaccination sites in Tokyo.

When Russia invaded Ukraine in 2022, statements such as "Ukraine is controlled by neo-Nazis" and "President Putin is a warrior of light" spread on social networking sites. Many individuals interpreted the invasion as an attack on the deep state, and analysis of past postings revealed several cases in which these same people shared conspiracy theories about COVID-19 or the US presidential election. A notable trend emerged: individuals who believed in one conspiracy theory often adopted others, even when the theories appeared unrelated.

(2) Rifts in Families

The COVID-19 pandemic led to an increase in people adhering to conspiracy theories, which, in turn, caused rifts within families.

For instance, a company employee in western Japan recounted that his typically mild-mannered wife told him, "If I get the vaccine, the deep state will control me. I could die." When he attempted to explain the information about vaccines as provided by public authorities, she became angry and tearful, leading them to stop speaking to each other. Similarly, a woman divorced her husband, who became aggressive after he began believing in a similar conspiracy theory. In another case, a wife left her home before her 50th wedding anniversary, fearing her vaccinated husband carried "poison" in his body. Since May 2021, a man publicly

[61] Yomiuri Shimbun Osaka, Social Affairs Department, *Information Pandemic: The Identity of the Things that Mislead You* (Chuokoron Shinsha, 2022).

sharing his troubled relationship with his conspiracy-theory-believing mother has received over 100 emails from people with the same problem.

(3) Drivers of Conspiracy Beliefs

Many of those profiled in the *Information Pandemic* became devoted to conspiracy theories through their engagement on social networking sites. These platforms influence these beliefs through echo chambers and filter bubbles, which contribute to a selective information diet.

In the same book, religious scholar Ryutaro Tsuji and others point to frustration, anxiety, and isolation as factors driving people's belief in conspiracy theories. Tsuji explains that conspiracy theories allow individuals to reshape reality into a more comprehensible form by identifying an enemy and attributing blame. They can also serve as a source of connection for people with similar beliefs, alleviating feelings of loneliness. For those who believe, he also states that conspiracy theories "provide a kind of salvation."

Appendix II: Basic Survey on the Information Environment

1. Overview of the Survey

This appendix presents the results of a survey of individuals' understanding and attitudes toward the current information environment. The findings serve as foundational data for exploring the future state of information health[62].

The survey was conducted via a crowdsourced online questionnaire from November 2 to 3, 2022, and received 2197 responses. Table 1 shows the demographic characteristics of respondents.

Table 1: Demographic Characteristics of Respondents

Gender			Age			Marital status		
	Percentage	*n*		Percentage	*n*		Percentage	*n*
Male	38%	838	20s and under	19%	427	Unmarried	53%	1168
Female	60%	1319	30s	34%	753	Married	46%	1008
Other/prefer not to answer	2%	40	40s	29%	639	Other	1%	21
	100%	2197	50s	13%	292		100%	2197
			60s and over	4%	82			
				100%	2193			

2. Survey Findings

(1) Awareness of the Information Environment

Table 2: Perceived Bias of News Consumed (Political Tendencies)

	Percentage	*n*
Conservative	5.0%	110
Rather conversative	30.0%	660
Neutral	46.3%	1018
Rather liberal	17.0%	374
Liberal	1.6%	35
	100%	2197

Table 3: Perceived Bias of News Consumed (Content-oriented Trends)

	Percentage	*n*
Mostly serious news	6.6%	145
Leans toward mostly serious news	26.9%	592
Matches my preferences; no bias in particular	21.2%	465
Leans toward mostly entertainment news	38.4%	843
Mostly entertainment news	6.9%	152
	100.0%	2197

Tables 2 and 3 show respondents' subjective perceptions of bias in the news they consume, focusing on political tendencies and content-oriented trends as key axes of bias.

The results show that, although perceptions of political and content-oriented trends differed depending on the individual (conservative/liberal, entertaining/serious), approximately half of the respondents recognized a predominant political bias in their news consumption; approximately 80% did for content-oriented trends. These results suggest a subjective awareness that their information environments are biased.

[62] These results are based in part on a research presentation (Teppei Koguchi and Toshiya Jitsuzum, "Monetary Evaluation for Realization of Information Health: Estimation of WTP for Services") at the Fall 2022 (47th) Annual Conference of the Japan Society of Information and Communication Research.

Table 4: Number of Information Sources

	Percentage	*n*
Too many	8.6%	188
Slightly too many	37.9%	832
Just right	45.9%	1009
Slightly insufficient	7.2%	158
Not sufficient at all	0.5%	10

Table 5: Trust Levels in Personally Significant Information Sources

	Percentage	*n*
Can trust sufficiently	2.0%	44
Somewhat trust sufficiently	38.3%	841
Hard to say; on a case-by-case basis	54.4%	1195
Somewhat unable to trust	4.4%	97
Difficult to trust in most cases	0.9%	20
	100.0%	2197

Tables 4 and 5 show respondents' attitudes toward information sources (media). Approximately half of the respondents felt that the number of sources they encountered was excessive, reflecting a feeling of diversity in the current media landscape and the volume of information. Regarding trust in their most important information sources, approximately 40% of respondents expressed general trust, while more than half said that their trust depended on the context. This suggests a widespread perception of the unreliability of the information with which they are presented.

Table 6: Attitudes toward Policies Addressing Fake News

	Percentage	*n*
All fake news should be eradicated immediately	15.2%	335
Fake news should be kept out of the public eye	31.6%	694
To some extent, fake news is inevitable; it should be minimized more than it currently is	44.9%	987
Fake news is inevitable; thus, no action is necessary	7.0%	154
Fake news is rare; thus, no action is necessary	1.2%	27
	100.0%	2197

Table 6 presents respondents' opinions about responses to fake news. More than 90% said that fake news should be reduced to some extent from its current level, although the preferred degree of reduction varied, suggesting that most individuals acknowledge the seriousness of fake news. However, this survey question only asked whether fake news should be reduced and did not explore specific strategies for achieving this reduction.

(2) Understanding of Issues Related to the Information Environment

Table 7: Awareness of Echo Chambers

	Percentage	*n*
I am familiar with the concept	12.1%	266
I have heard the phrase, but I am not really familiar with the concept	14.5%	319
I have never heard of it	73.4%	1612
	100.0%	2197

Table 8: Awareness of Filter Bubbles

	Percentage	*n*
I am familiar with the concept	14.7%	324
I have heard the phrase, but I am not really familiar with the concept	16.2%	356

I have never heard of it	69.0%	1517
	100.0%	2197

Tables 7 and 8 present respondents' awareness of the concepts of echo chambers and filter bubbles. The results show that approximately 70% of respondents had never heard of either term, suggesting a generally low awareness of these concepts.

(3) Attitudes toward the Concept of "Healthy Information Consumption"

Table 9: Agreement with "Healthy Information Consumption"

	Percentage	*n*
I agree with it	19.4%	426
I somewhat agree with it	44.0%	966
I do not have any feelings about it	26.9%	592
I somewhat do not agree with it	7.0%	153
I do not agree with it	1.5%	33
I do not understand it	1.2%	27
	100.0%	2197

Table 10: Perceived Importance of "Healthy Information Consumption"

	Percentage	*n*
Very important	17.3%	379
Important	50.7%	1114
Neutral	26.6%	584
Not very important	5.0%	109
Not important at all	0.5%	11
	100.0%	2197

Tables 9 and 10 present respondents' agreement with and perceived importance of "healthy information consumption." The survey defined "healthy information consumption" as "consuming information that satisfies one's interests while being aware of the existence of filter bubbles, fake news, and echo chambers and the likelihood of encountering them."

The results indicate a certain understanding of the concept, as more than 60% of respondents expressed agreement with the concept and recognized its importance. However, approximately 10% of respondents disagreed with its principles and importance.

3. Summary

The survey results can be summarized as follows.

Many individuals perceive some bias regarding the current information environment. Whether this bias exists is another question, but from a subjective standpoint, individuals experience an information environment that appears biased. However, perceptions of the degree of bias also vary.

Individuals also feel a sense of information overload and have limited trust in the information they receive. When combined with the finding that most individuals desire some kind of response to fake news, this suggests a general awareness of the need to consume trustworthy and unbiased information amid extensive available content.

Given these circumstances, fostering information health remains a goal. Nevertheless, given the insufficient awareness of issues such as filter bubbles and possible resistance from a certain percentage of individuals to the concept of "healthy" consumption of information, the immediate priority is to promote the understanding of the challenges surrounding the information environment and encourage broader discussions of these issues[63].

[63] Additionally, a related survey, the "Survey on Awareness and Behavior regarding Information Intake," was conducted in October 2022 by the Dentsu Research Institute of the Dentsu Group, Inc. https://institute.dentsu.com/articles/2635/.

Appendix III: Research Reported at the Symposium "The Dark Shadows of the Attention Economy and Information Health: Creating Healthy Spaces for Public Discourse with Integrated Knowledge" (held on March 26, 2024)

Issues Surrounding Information Health

Shinichi Yamaguchi

(1) Issues Related to Understanding the Information Environment and Actions for Information Verification

An international comparison survey conducted in December 2023 in collaboration with *Yomiuri Shimbun*[64,65] revealed that Japanese respondents were less likely to engage in actions for information verification compared to those from other countries[66]. Additionally, the percentage of respondents who were familiar with terms and concepts related to the information environment, such as "attention economy," "filter bubble," "echo chamber," and "confirmation bias," tended to be low in Japan. This vulnerability to falsehoods, misinformation, and the information environment was further corroborated by a survey by the *Nikkei*[67], which found that Japanese respondents were significantly less aware of fact-checking methods than respondents in other Asian countries.

(2) Generative AI and the "Misinformation 2.0" Era

Advances in AI technology and the proliferation of generative AI have made deepfakes widely accessible, enabling anyone to create and disseminate them easily. This development marks the onset of a new phase that could be termed "Misinformation 2.0," characterized by an explosion of falsehoods and misinformation. In Japan, hoax images of disasters and fake videos of prime ministers have made headlines.

The spread of deepfakes has also facilitated public opinion manipulation. With cheap and accessible technology, individuals and groups can increasingly influence society and politics. For example, reports indicate that one organization is creating a business around generating numerous avatars, assigning them social networking accounts, and using AI to generate posts automatically on social media to shape public opinion. This technique has already been utilized in elections in several countries and has been confirmed to operate in the Japanese language. Between June 2022 and May 2023, at least 16 countries reportedly employed AI-based generative tools to distort information on political or social issues[68].

The amount of AI-generated information and content is poised to skyrocket and will soon surpass that of human-generated information, with a considerable portion comprising falsehoods and misinformation. This influx of false information will likely cause widespread confusion in varied situations, including disasters, political events, fraud schemes, and court cases, and there is also concern that legitimate images and videos may be erroneously identified as fake[69].

[64] Survey of 1,000 men and women in Japan, the United States, and South Korea.

[65] "[A Selective Information Diet, Warped Cognition] Information and Casually-placed Trust...'Information Health' Survey in Three Countries: Japan, the US, and Korea," *Yomiuri Shimbun,* March 26, 2024. <https://www.yomiuri.co.jp/national/20240325-OYT1T50187/>

[66] Among the nine information-verification actions presented as those usually taken after being presented with information of interest, 94.7% of respondents in Korea and 92.5% in the US took at least one of these actions "sometimes" or more frequently, compared to 77.3% in Japan.

[67] Kentaro Takeda, Masaharu Ban, and Tomoya Onishi, "Japan lags Asian Peers in Dealing with Fake News," *Nikkei Asia,* June 10, 2023. <https://asia.nikkei.com/Spotlight/Datawatch/Japan-lags-Asian-peers-in-dealing-with-fake-news>

[68] Allie Funk, Adrian Shahbaz, and Kian Vesteinsson, *The Repressive Power of Artificial Intelligence* (Freedom House, 2023). <https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence>

[69] This phenomenon is referred to as the "liar's dividend," in which the side that lies benefits.

Acceptability and Cost Burden of Information Health

Teppei Koguchi, Shizuoka University

This paper presents and discusses the results of a survey on the acceptability of the concept of information health among individuals, focusing on their perceptions and the potential cost burden associated with its adoption.

1. Perceptions of Issues Related to Information Consumption

The concept of information health is currently the subject of considerable debate and will likely continue to evolve. However, regardless of its societal importance or desirability, if individuals are unaware of the various issues surrounding information consumption, they will feel no motivation to engage with the idea of information health.

To illustrate this, we conducted a survey assessing awareness of echo chambers and filter bubbles[70]. The results showed that in Japan, the awareness rate—defined as the percentage of respondents who recognized the terms and understood the concepts—was just over 10%, at 12.7% for echo chambers and 11.6% for filter bubbles. This indicates a widespread lack of knowledge or consideration of these issues. Similar surveys conducted in 2022 and 2023 also showed an awareness rate of approximately 15%, albeit with a distinct sample distribution, suggesting no significant change over the last three years[71]. Similar surveys in the United States, United Kingdom, France, and Germany found an awareness rate ranging from 18.5% to 39.7%, showing a slightly higher awareness rate in these countries than in Japan, although there are differences by country[72].

These findings highlight the need for increased public understanding of issues related to information consumption and the measures that should be taken to foster the adoption of information health practices.

2. Individual Perception of Information Consumption

As mentioned in the previous section, awareness of the issues surrounding personal information consumption is not necessarily widespread. Even among those who are aware of these issues, individuals must consider these issues as personally relevant. Even if such problems exist in society, if individuals think, "I am fine," or "I can consume information without any problems," societal discussions on improving information consumption may not move forward.

To examine this perspective, we conducted a subjective survey asking, "How healthy do you think your information consumption is compared to that of those around you?"[73] Participants rated their information consumption health on a 100-point scale, with 50 representing the perceived average health of information consumption in their immediate environment. The average score among respondents in Japan was 56.4[74]. In surveys conducted in 2022 and 2023, the score was slightly higher at 58.8 for both years. These results suggest that, on average, respondents tended to believe that they had healthier information consumption than those around them. Notably, this survey reflects subjective perceptions rather than actual health. Moreover, while the average score was above 50, not every individual perceived themselves as healthier than their peers. If a cognitive bias leads individuals to overestimate the health of their information consumption consistently, addressing this bias is crucial.

Comparatively, similar surveys in the United States, United Kingdom, France, and Germany had average scores from 60.9 to 65.5, suggesting that individuals in these countries have a more favorable view of their information consumption health than respondents in Japan.

3. Potential Cost Burden for Individuals

Achieving information health requires a realistic consideration of the cost. While the benefits of information health are evident, if the costs outweigh the benefits, pursuing it may not be desirable. Costs in this context include costs related to discussion, policy development, public awareness, and many others.

In this regard, we analyzed individuals' willingness to bear the costs of achieving information health[75].

[70] The online survey was conducted in March 2024. The sample comprised 757 individuals, evenly distributed across genders and age groups.

[71] These results are consistent with Teppei Koguchi and Toshiya Jitsuzumi, "Analysis of the Preferences for the 'Information Health' Concept: Consideration of Impact on Willingness to Pay," *Proceedings of the International Telecommunications Society Asia-Pacific Conference 2023* (2023).

[72] This online survey was conducted in March 2024. The sample comprised 411 individuals from each country (410 samples for France), evenly distributed across genders and age groups.

[73] See footnote 64.

[74] See footnote 65.

[75] See footnote 64.

The first potential avenue is through policy actions, which would spread the cost across society. Furthermore, recent advancements suggest the possibility of technological responses to promote information health, such as Internet-based tools. If such technologies at the individual service level were to be developed, a "beneficiary pays" system could be implemented in which individuals would bear the costs.

The analysis focused on estimating individuals' willingness to pay (WTP) per month for hypothetical applications that could eliminate filter bubbles and fake news. Respondents indicated they would be willing to pay approximately 100–200 yen per month. Whether this individual WTP is considered high or low and to what extent it is commensurate with the cost of these technological solutions require careful discussion. In any case, the issue of cost burden will be an important point of contention in future discussions on information health.

Analysis of the Current State of the Media and Challenges to Achieving Information Health

Yuya Suzuki

This paper examines the current state of media organizations, highlighting their initiatives and the challenges they face in fostering information health in the media. As part of this analysis, three key problems observed in recent years are outlined below.

① Cost-cutting Issues

Japanese and international media companies are resorting to cost-cutting measures, including reducing personnel, consolidating branch offices, decreasing helicopter and camera operations, limiting taxi vouchers, and curbing business travel expenses. These cost-cutting strategies impact the media's ability to cover events in real time and on location. Consequently, the information landscape becomes dominated by easily reportable stories or non-time-sensitive content, leading to worse quality of news coverage.

② Problem of "News Content ≠ Reporting Content"

In the attention economy, decision-making driven by cost performance has resulted in a distortion in the perceived value of news content. For example, restaurant reviews, celebrity gossip, and events trending on social media are considered cost-effective, while investigative reporting and election coverage, which require considerable amounts of time and human resources, are deemed cost-ineffective. Consequently, media priorities have shifted toward producing high-cost-performance content.

Furthermore, financial incentives have become a major distorting factor in value determination. Traditionally, to preserve content integrity, media organizations have maintained a separation (i.e., a "firewall") between the advertising department, responsible for generating revenue, and the editorial department, responsible for content production. However, as profit margins worsen, measures to maximize advertising revenue are increasingly being promoted as a business decision. For instance, practices such as product placement, where products being advertised are incorporated into content naturally, have caused confusion among the public, leading to a loss of trust in the media.

③ Issue of PV Quotas

An increasing number of media companies are adopting PVs as a metric for evaluating content quality. From a management perspective, this has certain advantages, such as tracking goal achievement, assessing progress, and boosting overall site PVs by setting quotas. Conversely, from the perspective of front-line staff, PV quotas affect their evaluations, promotions, and bonuses, and the pressure to meet quotas may result in the mass production of easily consumable and attention-grabbing content to maximize PVs. Consequently, only cost-effective content is generated, while high-quality content creation is minimized.

Thus, the precarious profit environment of the media industry is seriously impacting the information space. In Japan, total advertising expenditure, the primary source of revenue for many media companies, was approximately 7.1 trillion yen in 2022, about the same level as in 2007 (https://www.dentsu.co.jp/news/release/2023/0224-010586.html). Nevertheless, the distribution has shifted: advertising expenditure among the four major traditional media outlets fell sharply, while that for Internet advertising grew approximately fivefold. Competition is becoming increasingly fierce as various players, large and small, vie for a large slice of the advertising pie.

In such a challenging profit environment, the attention economy—where advertising expenses are allocated based on PVs and impressions—affects web-based media reliant on digital advertising and traditional media outlets. This has led to a content production landscape increasingly driven to capture attention. Thus, content designed to provoke strong emotional reactions, such as fear and anger—often as fake news or conspiracy theories—has proliferated. The expansion of the attention economy is becoming unstoppable.

Simultaneously, this shift also accelerates the consolidation and decline of media companies. In particular, media companies that devote personnel, financial, and time resources to deliver quality information have been forced to scale back, shut down, or compromise content quality because they cannot make a profit. This situation fosters a climate where "honesty is a fool's errand."

Efforts to advance information health are necessary to address these systematic challenges. We propose

two key directions to guide such efforts.

First, stakeholders should discuss the feasibility of “nutritional labeling” of media content and outline specific steps toward its adoption. Nutritional labeling would allow media organizations to create professionally curated, well-edited, and reviewed content from multiple perspectives, considering its impact on audiences. Users can also navigate the digital space while being aware of the “nutritional value” of content and the risks of overindulging in junky material. Furthermore, for advertisers, the shift from buying ad space based on attention metrics such as PVs and impressions to “informational nutrients” will help improve brand reputation as responsible operators. Such a three-pronged approach requires comprehensive discussion to build momentum for implementation.

The second proposed direction involves introducing the concept of an information checkup as a mechanism to support information health.

One example of an information checkup is a system that visualizes an individual’s informational and lifestyle habits on a radar chart. This mechanism could contribute to the improvement of information health by assessing and displaying the quality and quantity of information users are exposed to on a daily basis. Technologically, this tool could be developed by media companies that already retain user-viewing data, catalyzing social momentum. For broader implementation, advancing OP technology and leveraging data held by platform operators would be essential.

As outlined, the various problems associated with the deteriorating profitability of media companies are escalating worldwide, requiring swift countermeasures. Promoting information health serves as a pathway for media companies to reclaim a sense of discipline that has been lost in recent years. It would also benefit users and advertisers who play a dual role within the media and is crucial for maintaining a well-functioning, balanced information society.

The spread of generative AI such as ChatGPT poses a risk of accelerating information pollution caused by disinformation and misinformation. This risk is especially critical in the election year 2024 when national elections will be held in many countries. Concerns about election interference via cognitive warfare using falsehoods and misinformation, now made more sophisticated through generative AI, indicate the need for better societal resilience in addition to international measures. A transformation from an AI-dominated information space to one featuring human-centered intelligence augmentation (IA) is necessary.

●Cognitive Warfare and AI

A 2020 report by the North Atlantic Treaty Organization (NATO) and Johns Hopkins University defines "cognitive warfare" as "the weaponization of public opinion, by an external entity, for the purpose of (1) influencing public and government policy and (2) destabilizing public institutions"[76].

A key example is the cyberattack on the US Democratic National Committee during the 2016 presidential election and the subsequent leak of internal documents. There were indications that the Main Intelligence Directorate of the General Staff of the Armed Forces of the Russian Federation was involved, while WikiLeaks and other platforms disclosed internal Democratic Party emails. This led to widespread doubts about the fairness of the party's candidate selection process and the resignation of the party's national chairperson.

Eight years later, 2024 is marked by national-level elections in over 50 countries and regions, including the US presidential election[77]. With the rapid advancement of generative AI, election interference through cognitive warfare using disinformation and misinformation looms as a global threat.

●Boundary between Real and Fake

The impact that generative AI has had on the spread of disinformation and misinformation is characterized by its ability to achieve larger scale, lower costs, higher sophistication, and faster dissemination.[78]

This technology blurs the lines between the truth and falsehood, complicating efforts to discern credible information. One consequence of this blurring is a situation referred to as the "liar's dividend"[79], wherein false or misleading information is accepted as real, and real information is misidentified as false. The diffusion of deepfakes and fake AI-generated videos about wars and elections is also evident.

For instance, videos about the Russian invasion of Ukraine began to spread on social media in 2022, including those featuring Ukrainian President Volodymyr Zelensky calling for surrender, Ukrainian Commander-in-Chief Valerii Zaluzhnyi urging a coup, and Russian President Vladimir Putin announcing a peace agreement.

Furthermore, during the presidential election in Taiwan in January 2024, a fake video circulated of the elected president, Lai Ching-te, selling virtual currency. During the US presidential elections in November of the same year, an automated phone call with a faked voice of President Joe Biden urging people not to vote and a fabricated video of Taylor Swift declaring Donald Trump's victory were also spread.

●Response to AI Risks in Cognitive Warfare and Misinformation

Efforts are also underway to address risks posed by generative AI in cognitive warfare and disinformation.

In Japan, the December 2022 revision of the National Security Strategy and three other documents indicated the goal to strengthen the country's ability to respond to cognitive warfare, including the proliferation of disinformation. Moreover, the Hiroshima AI Process Comprehensive Policy Framework for 2023 outlines measures to address AI risks, including the proliferation of disinformation.

In terms of Europe, in 2015, the EU established the East StratCom Task Force under the European

[76] Alonso Bernal, Cameron Carter, Ishpreet Singh, Kathy Cao, and Olivia Madreperla, *Cognitive Warfare: An Attack on Truth and Thought* (NATO and Johns Hopkins University, 2020). <https://innovationhub-act.org/wp-content/uploads/2023/12/Cognitive-Warfare.pdf>

[77] Jill Lawless, "Over 50 Countries Go to the Polls in 2024. The Year Will Test Even the Most Robust Democracies," *The Associated Pres s,* January 10, 2024. <https://apnews.com/article/global-elections-2024-preview-cb77b0940964c5c95a9affc8ebb6f0b7>

[78] Kazuhiro Taira, *Chat GPT vs. Humanity* (Bunshun Shinsho, 2023).

[79] Bobby Chesney, and Danielle Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," *California L aw Review* 107, no. 6 (2019): 1753–1820. <https://www.californialawreview.org/print/deep-fakes-a-looming-challenge-for-privacy-democra cy-and-national-security>

External Action Service to monitor and create a database of false information originating in Russia through the EUvsDisinfo project. In May 2024, it also passed the *Artificial Intelligence Act*, a comprehensive risk-based regulatory law that prohibits subliminal techniques and addresses AI risks. In the same month, the Council of Europe adopted the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy, and the Rule of Law, which establishes a framework for the protection of human and other rights.

France established the Vigilance and Protection Service against Foreign Digital Interference in 2021 under the Secretariat-General for National Defence and Security to monitor foreign information warfare, while Sweden established the Psychological Defence Agency in 2022. In Taiwan, the Cognitive Warfare Research Center was created in January 2024 by the Ministry of Justice.

The United Nations unanimously adopted the resolution "Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development" in March 2024.

In terms of IT industry initiatives, Microsoft, Google, Meta, Adobe, Amazon, and other tech leaders established the AI Elections Accord in February 2024 to combat election-related AI disinformation. It emphasizes seven principles, including preventive measures to reduce the risks associated with deceptive AI election content, detection of AI-generated election content, and appropriate responses to identified deceptive content.

- **Social Resilience and the Human-Centered Information Space**

Addressing the risks of cognitive warfare and generative AI requires a societal shift from an AI-centric, algorithm-created information space to a human-centric information space rooted in IA.

US journalist John Markoff described the evolution of computer development as a conflict between the AI-based approach of John McCarthy and others, in which computers replace humans, and the IA-based approach of Douglas Engelbart and others, in which computers are used to extend human capabilities[80].

In today's digital landscape, AI algorithms dominate the attention economy, leaving humans vulnerable to disinformation and misinformation. This environment leaves the media, advertisers, and users as passive participants in a system of platform-driven priorities. Stakeholders must recognize and address this current challenge. Achieving a resilient society will require a shift in perspective from AI to IA, fostering a human-centered, self-aware information space.

[80] John Markoff, *Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots*, trans. Noriko Takiguchi (Nikkei BP, 2016).