

Study Group 'AI governance and its Evaluation'
Report on the Session #10

1. Introduction

The Japan Deep Learning Association establishes study groups as a forum for deepening knowledge and discussing domestic and international policy trends related to artificial intelligence (hereafter AI) and Deep Learning (hereafter DL). This study group, "AI Governance and its Evaluation," defines "governance" as a system of management and evaluation by various actors, and launched a study group in July 2020 to investigate what forms of governance are possible and conduct a year-long study to help build trustworthy AI systems.

In the 10th session (February 25, 2021), Mr. Hitofumi Yanai, FactCheck Initiative Japan (FIJ), a non-profit organization, and Mr. Hideto Takada, Japan Advertisers Association (JAA) presented topics under the theme of "Monitoring" and "Advertising" for AI governance.

This report is a reconstruction of the topical presentations and the discussions of the study group participants.

2. Current Status and Issues of Fact-Checking in the Face of Infodemics

Mr. Yanai gave a presentation on the topic of "Current Status and Issues of Fact-Checking to Confront Infodemic."

Current Status of Infodemics

Since the outbreak of COVID-19 virus, "infodemic"¹ has increasingly drawn attention. Factors that can cause infodemic include psychological factors and changes in the information environment. Psychological factors are the cognitive, psychological, and behavioral characteristics that humans originally possess, such as confirmation bias and scapegoating². Changes in the information environment include the spread of platform services and social networking services that aggregate and organize news information, the rise of emerging media that allow anyone to transmit and spread information around the world, and the decline in the influence of conventional media. In addition, recent

¹ Infodemic refers to a rapid and far-reaching spread of false and inaccurate information in the midst of a massive flood of information, and its impact on society.

² Scapegoating is the practice of shifting the cause or responsibility for the occurrence, expansion, or worsening of a negative event to another target, rather than to the direct cause, or of blaming the target, with a certain collective spread.

infodemic have been characterized by the echo chamber phenomenon³ and the increasing spread of untruthful information across language barriers. Infodemic lead to harmful rumors and health hazards, promotion of discrimination and prejudice, and intensification of hatred and conflict.

Current Status of Fact-Checking

Fact-checking is a process that sharing accurate information with the public by checking whether information, news, and discourse spread in society are based on facts and publishing the results of verification of their veracity and correctness. There are various types of untruthful information and discourse that influence society, including not only misinformation but also intentional false information, partially inaccurate information, and information whose basis is not clear. The role of fact-checking is to make misinformation visible based on verification results. By visualizing misinformation, it is possible to alert people and reduce the impact of misinformation.

Fact-checking activities began in the U.S. in 1995, and since the late 2000s, major media outlets have been responsible for fact-checking. In addition, during the 2016 U.S. presidential election, fake news spread on the Internet and criticism focused on the large platforms responsible for the distribution of information, which led the platformers to cooperate with fact-checking organizations and display fact-checking results in response to the criticism.

Over the next four or five years, the number of fact-checking organizations grew rapidly and spread around the world. However, with the increase in the number of organizations, their investigative approaches and methods diversified, and some organizations emerged that were not certain that they were verifying based on fair and impartial information. The International Fact-Checking Network (IFCN), therefore, developed the following Code of Principles for Fact-Checking in 2016.

1. Nonpartisanship and Fairness
2. Transparency of Sources
3. Transparency of Funding and Organization
4. Transparency of Methodology
5. Open and Honest Corrections

The IFCN has 78 member media and organizations from around the world that have met the Code of Principles and other criteria, and these media and organizations are continuously conducting fact-checking activities in accordance with the policies and

³ An echo chamber refers to situations in which beliefs are amplified or reinforced by communication and repetition inside a closed system.

standards set by the IFCN.

In Japan, the Fact-Checking Initiative (FIJ), a non-profit organization established in 2017, promotes fact-checking activities. At the time of the FIJ's inception, fact-checking was not recognized by the public in Japan and there was a lack of fact-checkers. The FIJ was established with the aim of increasing the number of fact-checkers and networks to promote fact-checking across media, journalists, and technology companies. The FIJ is not an organization that conducts fact-checking activities, but rather supports fact-checking activities and promotes fact-checking principles and methods. FIJ aims at protecting society from mis/disinformation.

As one of the specific initiatives of the FIJ, a special website for fact-checking related to COVID-19 has been created and is available to the public. In addition, an international cooperation framework (project) was launched from January to early February 2020, and fact-checkers from around the world are verifying various misinformation about COVID-19, centralizing the fact-checking results in a data base, and sharing the information. In this project, fact-checking related to COVID-19 in Japan circulating overseas is being conducted with the cooperation of FIJ.

Another important role of the FIJ is to operate a fact-checking support system. This system uses machine learning to monitor questionable information and notifies media partners of the results. The functions of the system are divided into three major stages. First, the Fact-Checking Console (FCC) System, which uses machine learning, automatically detects questionable information. Secondly, the information is then registered into the ClaimMonitor Database after being checked by FIJ staff. Then, the registered information is automatically notified to media partner fact-checkers. The media partner fact-checkers investigate and create fact-checked articles⁴ based on the questionable information collected in ClaimMonitor Database, and register the summary information of the fact-checked results in a web application called FactCheck Navi⁵. The registered summary information is available to the public through the web site. Currently, there are more than 2,600 questionable statements registered in ClaimMonitor Database.

The FIJ is also interacting with organizations in South Korea. The Seoul National University (SNU) FactCheck Center displays the results of verified information on the web platform. Thirty media outlets, including newspapers, TV stations, news agencies, and online media, collaborating with SNU Fact Check Center conduct fact-checking, and disperse the validity of statements to the public via NAVER website. In Japan, the Sankei

⁴ As of January 1, 2021, the number of fact-checked articles by FIJ media partners ranged from 10 to 20 per month.

⁵ <https://navi.fij.info/>

Shimbun and Mainichi Shimbun have joined the FIJ as media partners, and there is momentum for fact-checking to spread.

Toward future expansion of fact-checking

At the root of the FIJ's goal of increasing society's immunity to misinformation, as mentioned above, is the premise that misinformation cannot be eliminated. As long as we are engaged in free social activities, it is common for misinformation to appear as part of our expressive activities, and legal regulation of misinformation could be a dangerous and arbitrary infringement of expression. Therefore, as a countermeasure against misinformation, it is necessary to expand fact-checking by a variety of private sector players. By visualizing misinformation and understanding the actual situation through fact-checking, it will be possible to take measures to prevent the spread of misinformation in cooperation with platformers and SNS providers, and to develop effective countermeasures through academic research. The spread of fact-checking will also help raise the level of media information literacy among the general public.

The future model of fact-checking will be one in which roles are shared and collaborated among individual bearers, the media, and professional organizations. Even students and the general public can monitor and supplement information that is sensibly questionable. On the other hand, the media and professional organizations should take on the role of distributing professional research and judgments of authenticity in articles for the general public, as this requires expertise. Therefore, it is essential that individual bearers, the media, and professional organizations share the roles and work together, and that the FIJ supports fact-checking activities through its projects to spread fact-checking.

3. Discussion points in the Question & Answer session ("Monitoring" for AI Governance)

Based on the topic presented by Mr. Yanai, the following question and answer session was held.

FIJ's Media Partners

- ✓ Among TV stations, Chukyo TV is a media partner, but no key stations are. In South Korea, the SNU FactCheck Center and a major TV stations are working together to distribute weekly fact-checking programs. According to public opinion polls in South Korea, a large percentage of people are aware of fact-checking through television. In Japan, too, the role of television is significant, and it is expected that TV stations will participate in fact-checking.
- ✓ Some companies are media partners of the FIJ even though they do not conduct

fact-checking.

Time required for fact-checking

- ✓ While some content can be verified in half to one day, cases that require expertise take one to two weeks. While fact-checking requires a certain amount of time, the media on the information dissemination side has a strong desire to disseminate information while it is still fresh, because the value of information becomes less meaningful if it is disseminated too late.

Target of fact-checking

- ✓ Fact-checking is the process of checking whether detailed "facts" (numbers, content of statements) were objectively such, but "truth" is not subject to fact-checking because of human subjectivity.
- ✓ Items that are clearly parodies will not be subject to fact-checking.
- ✓ Not only deep fakes, but also fabrication of purpose and manipulation of impressions by cutting out and editing parts of real videos (done in the classic way) are subject to fact-checking.

Responding to slander and defamation that spreads and continuously posts negative information about individuals

- ✓ Anything that constitutes defamation or violation of personal rights will be a legally regulated issue. Therefore, the early identification of victims and perpetrators of slander and defamation, and the establishment of a system that facilitates their vigilance and compensation, will serve as a deterrent.
- ✓ It is not desirable for misinformation itself to be directly regulated, but slander should be.

Fact-checking and AI

- ✓ The technical help of AI will reduce the burden of manual tasks performed by humans.
- ✓ The FCC System only detects information that Internet users say may be untrue about a certain news, not fact-checking it.
- ✓ The technology for transcribing spoken Japanese is still in its infancy. As this technology develops, more information will be available in text form, which will be useful for fact-checking.
- ✓ Fact-checking methods for deep fakes is an issue for the future.

Platformer's response after fact-checking

- ✓ Forced exclusion, such as account suspension, is a last resort on the part of the platformer. Forced exclusion can also lead to secondary problems such as conflicts.
- ✓ Rather than treating users equally and suspending accounts in an opaque manner, platformers should work to reduce the negative effects of untruthful information and discourse by providing a mechanism for users to recognize it for what it is and prevent it from spreading.
- ✓ Since it is humans who do the fact-checking, the results are not always 100% correct, and there is a great possibility that the results could be wrong. The results of fact-checking should ultimately be judged by individuals.

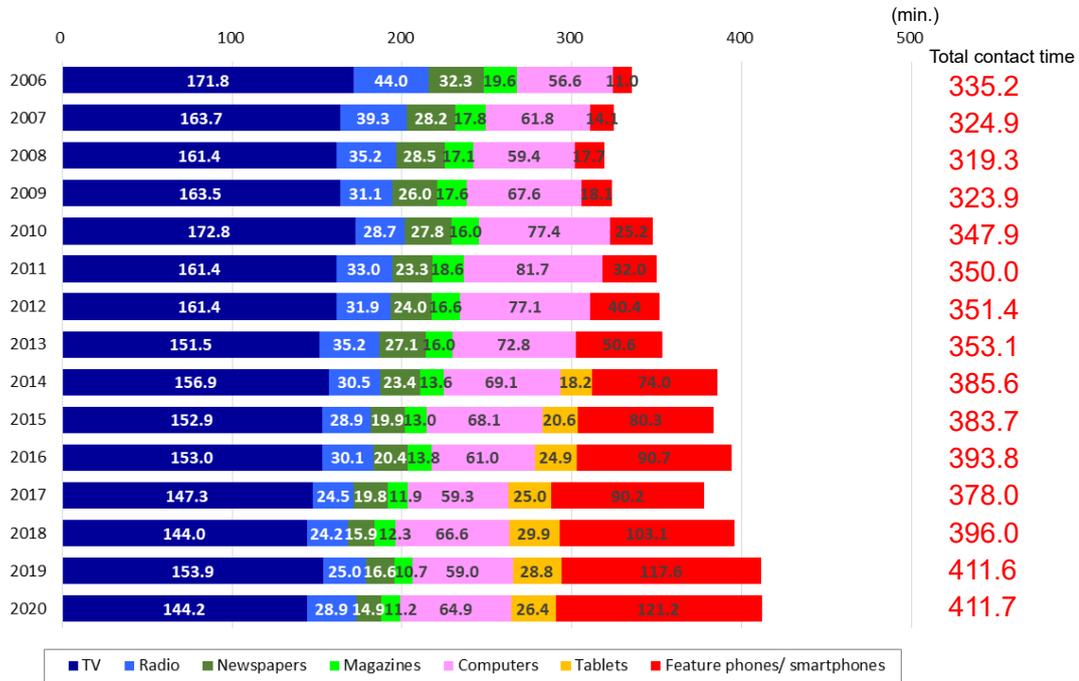
4. Efforts to Improve the Credibility of Digital Advertising

Next, Mr. Takada spoke on the topic of "Efforts to Improve the Credibility of Digital Advertising."

Changes in the advertising environment due to digital technology

According to the time-series analysis of consumers' total daily media contact time (see Figure 1), weekly average of media contact time has been steadily increasing, from 335 minutes in 2006 to 411 minutes in 2020. In terms of media, computers, tablet devices, and feature phones are getting longer, while TV, radio, newspapers, and magazines are steadily declining. This shows that the content of media contact is changing.

**Figure 1: Time series analysis of total daily media contact time (weekly average):
Tokyo area⁶**

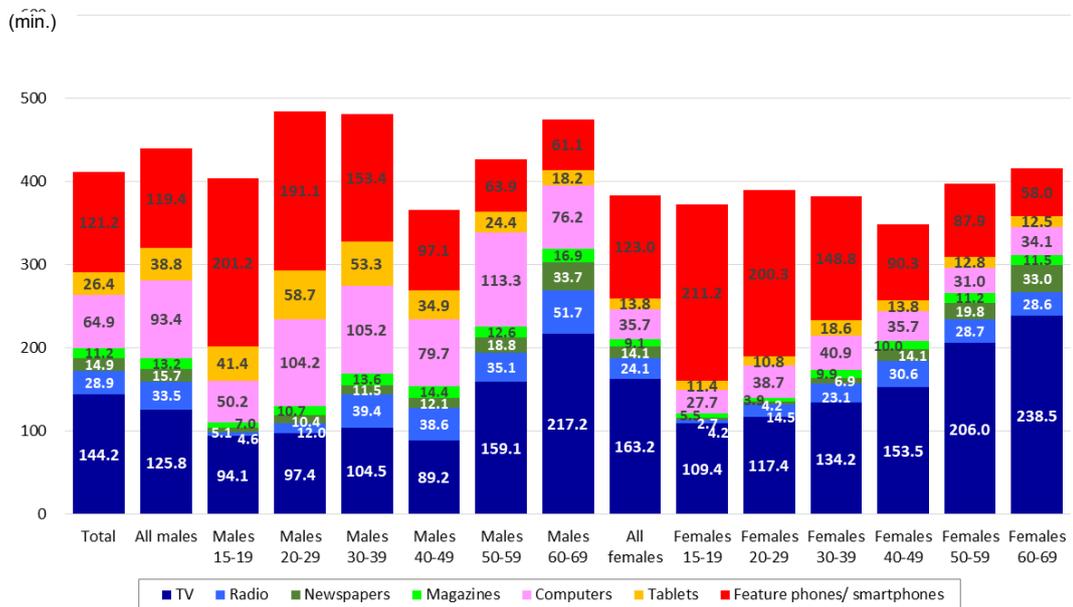


- Total contact time for all of those interacting with each type of media.
- The “feature phone/ smartphone” category was labeled “mobile phones” in the 2014 survey, and it included smartphones.
- Tablets were added in 2014.

According to the total daily media contact time by age and gender group (see Figure 2), men tend to spend more time on computers than women, regardless of generation, and the older generation tends to spend more time on TV. Younger generations tend to have longer contact time with feature phones/smart phones. This shows that we are no longer living in an age where advertising on TV alone is enough to sell products.

⁶ Excerpt from “Annual Media Consumption Report 2020” by The Hakuhodo DY Media Partners Institute of Media Environment

Figure 2: Total daily media contact time by age and gender (weekly average): Tokyo area⁷

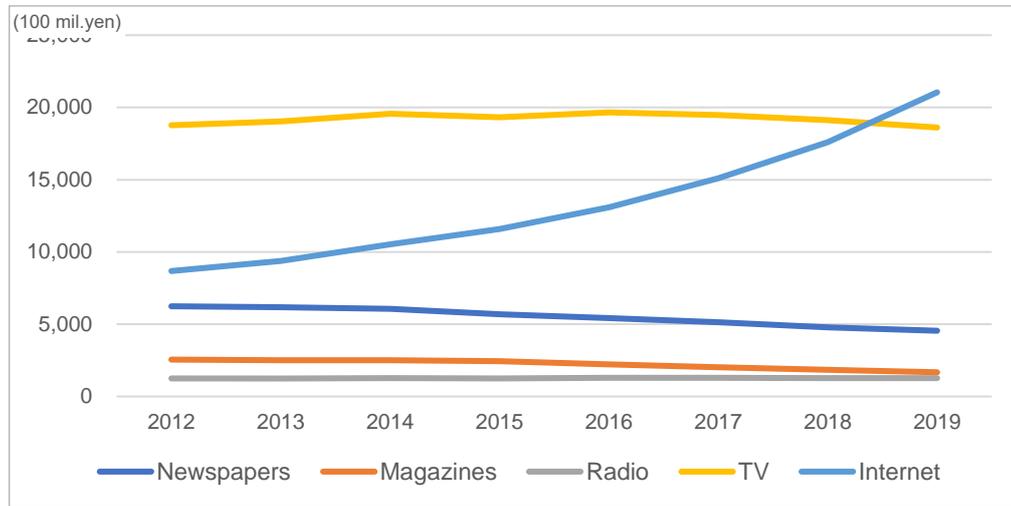


- Total contact time for all of those interacting with each type of media.
- The “feature phone/ smartphone” category was labeled “mobile phones” in the 2014 survey, and it included smartphones.
- Tablets were added in 2014.

Changes are also occurring in advertising expenditures (see Figure 3), with Internet advertising expenditures overtaking TV advertising expenditures to become the top domestic advertising expenditure in 2019. For media other than the Internet, advertising expenditures are declining or remaining flat, but Internet advertising expenditures are absorbing this decline.

⁷ Excerpt from “Annual Media Consumption Report 2020” by The Hakuho DY Media Partners Institute of Media Environment

Figure 3: Trends in Advertising Expenditures in Japan⁸



Communication methods are also changing. Before the development of the Internet, the mainstream communication method was for companies to send out their messages unilaterally by deciding their brand image and message in advance, selecting the media that would be viewed most by their target consumers, and launching advertisements. In contrast, today, even if a company unilaterally decides the brand image, consumers do not receive it as it is transmitted. The subject of communication has shifted to the consumer, and companies are required to have two-way communication with consumers. It is important for companies to take an approach that builds long-term relationships with consumers who freely come into contact with advertisements at their own timing. Companies are also expected to provide a consistent and genuine customer experience while communicating with consumers in both directions.

The following three points are basically the strategy for digital advertising, and the mutual effect can produce results that exceed expectations (see Figure 4).

➤ Owned Media

Promote understanding of the brand message among potential customers and fans of the company's products and services by disseminating information about the company. (e.g. Company's own website)

➤ Paid Media

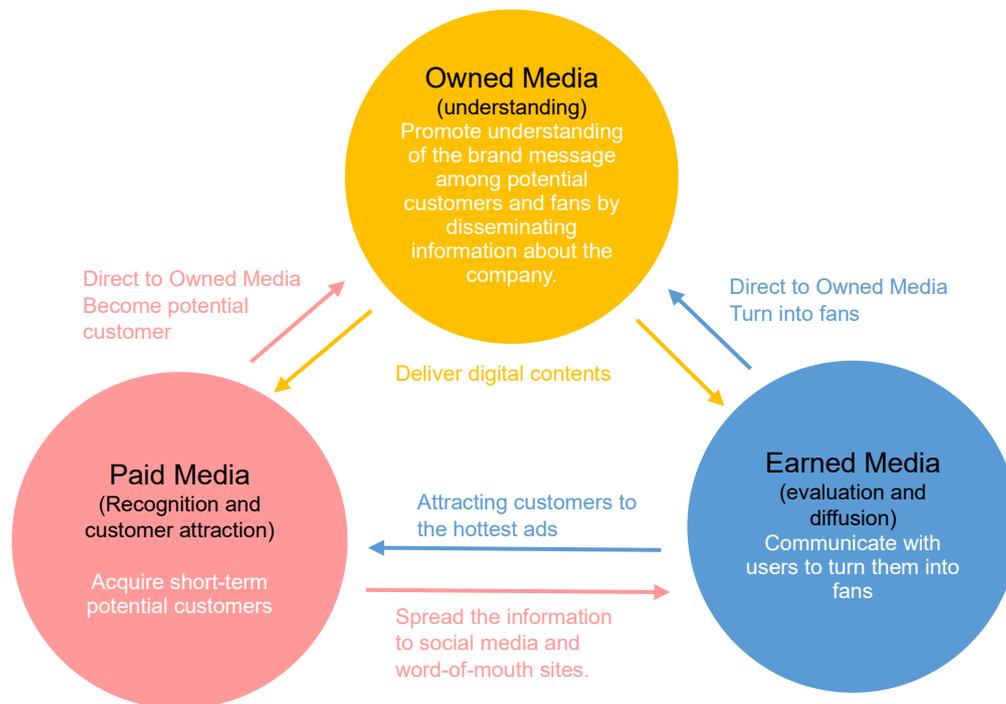
Acquire short-term potential customers and direct them to owned media. (e.g. Display Ads)

➤ Earned Media

⁸ Graph created from Dentsu Inc.'s "Advertising Expenditures in Japan."

Communicate with users to turn them into fans of company's products and services, and direct them to owned media. (e.g. SNS)

Figure 4: Types of Digital Advertising⁹



The following are the main types of digital advertising, and companies need to use them according to their objectives and communicate for branding purposes. Source: Sendenkaigi

- **Display Advertising**
Banners and other advertisements that mostly displayed on websites and applications.
- **Listing Advertising (Keyword Targeted Advertising)**
Text ads that are displayed in conjunction with keywords searched by users. Advertisers buy the words that are searched for, and when those words are searched, the ads are displayed. The order of display is determined by the price of the ad.
- **Native Advertising**
An advertisement that are integrated and camouflaged into the platform on which they appear in a “natural way.” For example, in a magazine tie-up ad, the editors create the ad themselves, so it can introduce products without destroying the world

⁹ Excerpts from this study group's public materials

view of the magazine.

- Video Advertising
Video advertisements on YouTube and other video channels
- Affiliate Advertising
Advertising in which the ad provider posts the ad on the affiliate's site and pays a fee based on the performance as rewards.
- Social Media Advertising
An advertisement with a mechanism for spreading information by using connections with friends on social media.

There are various methods of billing for digital advertising, as shown below.

- Cost-Per-Mille (CPM) Advertising
A fee is charged for impressions, i.e. the number of times that ad is displayed and viewed. (e.g. Display Advertising)
- Cost-Per-Click (CPC) Advertising
A fee is charged for each time the ad is clicked. (e.g. Listing Advertising)
- Cost-Per-Acquisition (CPA) Advertising
Fees are charged according to the number of final results generated, such as product purchases and requests for information materials.
- Cost-Per-Install (CPI) Advertising
Charges are incurred when a smartphone users click on an ad, go to an application download site, install and launch the specified application. (e.g. Social Media Advertising)

Quality Issues in Digital Advertising

As quality issues in digital advertising, it is necessary to take measures against the following three problems.

- Ad fraud
Digital ad fraud is a method of disguising and mass-producing impressions (number of times ads are served) and clicks by using automated programs called "Bot" in order to illegally increase revenues, and is also referred to as "ad fraud." In the case of Cost-Per-Mille (CPM) advertising and Cost-Per-Click (CPC) advertising, Bot and other programs roam the site and display advertisements. Even though the consumer does not view the ad, the advertiser is charged a fee based on the padded exposure and number of clicks. Another trick is to put another ad on the back of the displayed ad, so that when the ad on the front is displayed, the advertiser is charged for the invisible and hidden ad as well.

For companies that have already taken measures against ad fraud, ad fraud is seen in only 2.6% of ads shown on desktop displays and 1.9% of ads on mobile web displays, while for companies that have not taken measures, it is around 10%.

➤ Brand safety

Brand safety is a measure to prevent ads from appearing on inappropriate pages or content that may damage the image and reputation of brands, such as pornographic content or content related to anti-social activities. Just because advertisements can be displayed many times does not mean that they can be placed everywhere. That is, even if an advertisement is placed on a site that contains adult sites, violent or illegal drugs, illegal contents, etc., and earns the number of times it is displayed, there is a risk of brand damage.

According to one survey, 3.2% of ads on desktop displays and 7.6% on mobile web displays are brand-safety problematic.

➤ Viewability

Viewability is the percentage of impressions that were actually in the realm of the user's eyes. This is the issue of whether the advertisement is displayed in a visible location. Even if the ad is in an area that cannot be viewed by the user until the ad is scrolled to the bottom of the web page, the ad is counted as a single display and a fee is charged.

According to one survey, 58.0% of ads on desktop displays and 43.5% on mobile web displays are potentially problematic in terms of viewability.

One of the reasons behind the above-mentioned issues is that the flow and mechanism of placing advertisements differ between mass advertising¹⁰ and digital advertising. While mass advertising is basically a conventional system of securing an advertising space in advance and placing an ad in that frame, digital advertising is mainly a system of targeting ads to targeted people. The advantage of digital advertising is that advertisers can advertise to their target audiences without unnecessary expenses.

There are two types of digital advertising placement processes: pure advertising and operational advertising. In pure advertising, the advertiser and the Web media side set the price for the advertising space and placement period in advance, and in this case, the space and placement period are guaranteed, but the ad unit price tends to be high. Operational advertising is the main digital ad placement process today, and is a system in which advertisements are decided through auctions for ad spaces. In this case, the number of clicks is guaranteed, and the advertising unit price tends to be lower.

The digital ad placement process is changing in the following ways.

¹⁰ Advertisements in TV, newspaper, radio, and magazine media.

1. Advertisers place orders with advertising companies, the advertising companies select the most appropriate media, and the advertising flow is established.
2. Because of the large number of media on which advertisements are posted on the Internet, ad network companies that bundle media have emerged.
3. There are now many ad network companies, and an ad exchange has been established to purchase ad inventory from multiple media and networks using a bidding system.
4. In order to serve advertisements to targeted audiences, a system has been established to collect audience data (consumer attributes, behavioral history, etc.) held by each media and ad network provider and exchange it among businesses.
5. Systems have been established for media to improve the efficiency of ad spaces and maximize revenue, and for automatically purchasing ad spaces that match the conditions set by advertisers and advertising companies, such as posting surface, price, and target, and distributing the ads.

From the transition of the digital ad placement process described above, it can be seen that the number of players involved in the transaction is increasing and the route is becoming more complex. As a result, in digital advertising, a situation occurs where it is not clear where and how the ads are appearing, or which players are involved, leading to the issues mentioned above.

The advertiser's response to the aforementioned issues could include block lists of inappropriate sites, safe lists of trusted destinations, Private Market Place (PMP)¹¹, and ad verification¹² tools. The Japan Advertisers Association (JAA)¹³ has released the "Advertisers' Declaration on Digital Advertising Issues¹⁴," which outlines basic actions for partners¹⁵ to take in response to quality issues in digital advertising. In addition, Japan Advertisers Association (JAA), Japan Advertising Agencies Association (JAAA), and Japan Interactive Advertising Association (JIAA) will establish JICDAQ¹⁶ (Japan Institute for Certification of Digital Advertising Quality), a third party certification organization for the quality of digital advertising, in March 2021. This is a certification organization that has been decided to be launched with the aim of solving quality issues in the digital advertising market, thereby ensuring the healthy growth of the market. The

¹¹ Private Market Place (PMP) is an advertising transaction marketplace in which advertisers and media that can participate is limited.

¹² Ad verification is a process which verifies that digital ads are being placed on the appropriate sites and how much they are being viewed.

¹³ http://www.jaa.or.jp/about/about_jaa/

¹⁴ http://www.jaa.or.jp/wp-content/uploads/2019/11/JAA_proclamation.pdf

¹⁵ The partners mentioned by the JAA refer to "media," "platformers," "technology companies," and "agencies."

¹⁶ <https://www.jicdaq.or.jp/>

first step of the organization will be to eliminate ad fraud and ensure brand safety. This initiative not only certifies businesses, but also registers advertisers who agree with the philosophy, and recommends that registered advertisers be given priority in doing business with certified businesses.

5. Discussion points in Question & Answer session ("Advertising" for AI governance)

Based on the topic presented by Mr. Takada, the following question and answer session was held.

Status of Digital Advertising Issues Overseas

- ✓ In the U.K., one organization is in charge of verifying issues such as ad fraud and viewability, and JICDAQ is using this U.K. organization as a reference model.
- ✓ In the U.S., organizations have been established for each issue, such as ad fraud and viewability, and each of them is examining the issue professionally.
- ✓ In the U.K. and the U.S., verification organizations have been established due to advertisers' belief that they should self-regulate rather than have the government regulate the issue of digital advertising.

Dealing with advertising restrictions on content related to antisocial forces

- ✓ Since it is advertisers who use the PMP (Private Market Place), advertisers and advertising companies are cooperating with government agencies in an effort to exclude illegal content from overseas using block lists.
- ✓ Platformers are also becoming more careful, and the number of advertisements for content that is overtly related to antisocial forces is decreasing.

Status of application of ergonomic devices and technologies to serve a favorable advertising

- ✓ For example, advertisements that are displayed repeatedly may give a negative impression, and advertisers are conducting research and raising awareness in order to place advertisements that do not make consumers feel uncomfortable. However, it is difficult to apply technology in this area at present, as it depends on the subjective judgment of the individual as to what situations make them feel uncomfortable in the first place.

How to detect Ad fraud

- ✓ Ad fraud detection is based on a technology that uses a system in which a

specialized company patrols websites to identify fraudulent impressions. The system is expensive to use and not available to everyone. Even advertisers who have adopted this system do not check it in every transactions.

Advertiser-Media Relationship

- ✓ Media and ad network companies that do not place importance on ethics and credibility should be eliminated through the industry's self-cleansing process. It is hoped that advertisers will clarify the issues of digital advertising, and by sharing these issues, the media themselves will voluntarily change to place greater emphasis on ethics and credibility.
- ✓ From the advertiser's perspective, we believe that trustworthy digital advertising transactions are essential for further increasing the value of media, and that the party that ultimately sells the media should be responsible for the transparency of such transactions.

What is considered important as the credibility of the platform (environment in which information flows)

- ✓ It is important to ensure transparency, as a black box of a platform can lead to mistrust.
- ✓ It is expected that the information obtained by the platformer will be legitimately and appropriately used in marketing. However, if the platformer refuses to provide the information, it will be difficult to provide a consistent and genuine customer experience because it will not be able to deliver the information accurately to the consumers who require the information or the message.

We will continue to discuss AI governance in Japan and abroad through this study group.

Written by Yuki Kiyomi
Translated by Michiko Shimizu

<Outline of the 10th Session of the Study Group>

Date & Time: Thursday, February 25, 2021, 16:00-18:00 (Zoom)

Agenda:

- Topical presentations:
 - "Current Status and Issues of Fact-Checking to Confront Infodemic" provided by Mr. Hitofumi Yanai (FactCheck Initiative Japan (FIJ))
 - "Efforts to Improve the Credibility of Digital Advertising" provided by Mr. Hideto Takada (Japan Advertisers Association (JAA))
- Question and answer session / discussion