May 17, 2023

Opinion on the Use of News Content by Generative AI

The Japan Newspaper Publishers & Editors Association

Because of the rapid development of generative artificial intelligence (AI) technology, improved convenience in various aspects of society is expected. On the other hand, there are negative effects such as unauthorized use of copyrighted works and other content by AI, or inappropriate use of AI. Laws and social systems cannot yet catch up with the advances in AI technology. In addition, information disclosure by the companies developing AI is not sufficient. Challenges have arisen from the perspective of protecting the healthy discourse space that underpins democracy. We state here our opinion about concerns in the press-related field.

(1) Disruption of discourse space and social disturbance

Leading generative AI can generate answers as if conversing with a human. However, they often contain mistakes and inaccurate information and there is a danger of countless users believing the AI's false answers.

The ability of AI to generate a large number of articles in a short time can be exploited to spread disinformation, harmful information, or politically motivated information to mislead public opinion, largely on the internet discourse space. Also it is possible to mass-produce new articles by combining multiple articles from news companies and slightly changing the wording. We have already seen a phenomenon known as "trend blogging," through which a large number of articles of uncertain authorship and content appear on the internet and are highly ranked in search results. If unregulated development proceeds, there are concerns that AI may exacerbate this kind of behavior.

If such disruption of the discourse space proceeds, it will lead to social disturbance. It cannot be overlooked in terms of protecting democracy.

(2) Concerns about protection of personal information

Generative AI training uses vast amounts of information on the internet and is likely to include "sensitive personal information," as defined by Japan's Act on the Protection of

Personal Information. In principle, the acquisition itself of sensitive personal information is prohibited without the consent of the individual concerned, but there is a risk that AI may include such information in its answers.

If the companies developing and providing AI provide services using information on the Internet without permission and in an unregulated manner, without sufficient consideration on the ethics and responsibilities essential to deal with information, the legislation of personal information protection would be undermined.   We are seriously concerned about this situation.

At the end of March, the Italian government temporarily banned the use of ChatGPT, provided by the US company OpenAI, indicating suspicions that its collection and use of personal data violated personal information protection laws.

The current Japanese Act on the Protection of Personal Information is not sufficient to apply to cases of AI answers containing sensitive personal information.

(3) Problems of the current Copyright Act and of the process leading to law amendments

As concerns about the unauthorized and unregulated use of news content by AI are growing in other countries, it has led to multiple lawsuits against AI developers in the US and Europe. In March, the US Copyright Office indicated its intention to examine how to apply legal discipline to generative AI. In April, the News/Media Alliance (NMA), an industry association consisting of about 2,000 newspaper and magazine publishers in the US, stated "Generative AI (GAI) developers and deployers should not use publisher Intellectual Property (IP) without permission, and publishers should have the right to negotiate for fair compensation for use of their IP by these developers" and "Absent permission and specific licenses, GAI systems are not simply using publishers' content, they are stealing it."

According to experts, although the Directive on Copyright in the Digital Single Market of the European Union (EU) has provisions to limit the rights of copyright owners as in Japan, right holders can refuse (opt-out of) commercial use. In the UK, the limitation of copyright is restricted to research purposes only and not allowed for commercial use. Compared to these rules, Japan's Copyright Act is extremely broad in its coverage for AI machine learning.

In Japan, amendment of the Copyright Act in 2018 legalized in principle the collection and use of existing copyrighted works without permission in the process of developing

AI and other technologies (Article 30-4 of the Act). Restrictions on commercial use and opt-out as in the EU were not established. The reason for this is as follows: At the time of the law amendment, use for technological development was considered not to be people enjoying copyrighted works through their own senses. Therefore this use was not considered to be impairing the right holders' opportunity for compensation and, as long as it would not harm the interests of right holders, protection of right holders such as opt-out was deemed unnecessary.

However, at the time, the negative effects of advanced AI such as generative AI were not fully envisioned. The emphasis during the legislative process was on the amendment of the law so as to encourage innovation originating from Japan, specifically Japanese search engine development. As there were only a few references to AI, there was no indication from the government that AI could generate new expressive work and threaten right holders. So the right holders did not believe that the use of copyrighted works for technological development would become a problem, and for this reason, the law was amended to favor AI development without major debate in the Diet.

(4) Issues about copyrighted works of news organizations

News companies such as newspapers and news agencies have copyrights and other legal rights over many articles, photographs, images, and other content.   However, there is growing concern about unauthorized and unregulated use of such news content by AI.

Newspaper companies have long provided databases containing past newspaper pages and articles for a fee, and in recent years, they have also sold article data for AI development. If AI imports large quantities of articles, photos, images, and other data from news organizations' digital news sites without permission, commercial AI services for third parties developing it could conflict with the existing database sales market and "unreasonably prejudice the interests of the copyright owner" (Article 30-4 of the Act). Also, even if all or part of a particular article communicates nothing further than facts and hardly constitutes a copyright, many contents deserve legal protection because of the effort and cost invested by the newspaper companies. Even if an AI collects and uses only the factual part, it does not mean it will always be legal.

If a user operates a developed AI and generates new articles or images by inputting specific keywords or instructions, and the outputted content relies on and is highly similar to original articles or images, it would be a violation of the Copyright Act.

(5) Opaque operation status and insufficient information disclosure to right holders

The actual usage of news content by AI is a black box, so cannot be known. As countless AI services around the world can be operated online by ordinary users to generate articles and images, it is difficult to know exactly how each service collects and uses news content and how users use it. Even if inappropriate use is found, there are no established procedures or rules on how to remove the relevant data from the AI. So there must be rules requiring AI developers to disclose and notify, and to make partly transparent, about which news organizations' content has been used for training data.

Large-scale machine learning by AI and the use of developed AI are increasingly being conducted on the Internet. Since information is exchanged across national borders, it can be said to be necessary to harmonize the legal systems of different countries while developing legislation.

As indicated above, if unauthorized and unregulated use of news content by AI becomes a fait accompli, it could cause huge damage to the business of news organizations and make it difficult for them to continue providing quality news content. If quality news content, which underpins democracy, decreases, the public's right to know may be hampered.

Generative AI is not only a problem for the news industry, but a common and urgent issue for various creative industries that produce copyrighted works such as literature, comics, animation, videos, and music.

The Japanese government should hasten institutional measures to make generative AI harmonize with society from the perspective of the entire legal system, including the Copyright Act and the Act on the Protection of Personal Information.

At the House of Councillors' Accounts Committee on April 3, issues under the Copyright Act were pointed out as negative aspects of AI, and Japan's Prime Minister KISHIDA Fumio stated "I am aware of the suggestion that there are issues that have not yet been sorted out." On April 30, the G7 Digital and Technology Ministers' Meeting in Takasaki City adopted a ministerial declaration on AI which mentions the necessity to "continue promoting safety and trust as these technologies develop," indicating such issues as dealing with disinformation and the protection of intellectual property rights and personal data. We hope that the Summit Meeting in Hiroshima City on May 19-21 will also have a full discussion on how to achieve responsible AI.

Even if lawsuits develop over rights infringement by and the legality of generative AI, judicial precedent would take years to accumulate in Japan, Europe and the US. To ensure the rights of right holders are respected, we strongly hope that the Japanese government and others will examine appropriate measures before unauthorized and unregulated use of news content by AI becomes a fait accompli.