General Monitoring and Constructive Knowledge? Issues of Automated Content Moderation by Hosting Service Providers Under Japanese Law

Toru Maruhashi

General Monitoring and Constructive Knowledge? Issues of Automated Content Moderation by Hosting Service Providers under Japanese Law

Toru Maruhashi[1]

[1]Dept of Law, Meiji University
torumaruhashi@meiji.ac.jp

Abstract. This article reviews the current state of Japanese legal system regarding Content Moderation including automated or AI-based system operated by Hosting Service Providers (HSPs) and its relationship with knowledge-based limitation of liability of and injunctions against HSPs. We have been careful enough not to legislatively impose HSPs general monitoring obligations due to concerns on de facto censorship, significant chilling effect on freedom of expression and cost burden on HSPs. However, recent discussion was leaning towards proactive monitoring of illegal information with expectation that introduction of Content Moderation equipped with AI will become easier due to the progress of technology diffusion and cost reduction. This Japanese legal status and discussions are evaluated in comparison with global legal system especially recent EU legal system updates and judicial precedents. Finally, it concludes with extraction of issues such as prohibition of general monitoring and imposing excessive cost burdens on HSPs and need for confirming effect of proactive Content Moderation on knowledge-based limitation of liability.

Keywords: Automated Content Moderation, Intermediary Liability, Hosting Service.

1 Introduction[1]

An internet intermediary service classified as a hosting service provider, especially a platform service such as a large-scale Social Network (collectively, an "HSP") which moderates a large amount of content uploaded by users. This so-called “Content Moderation[2]” operation may involve monitoring all uploads of users, suppressing

[1] This article is a product of further elaboration of my previous articles, "Intermediaries' Liability - Transformation of Limitation of Liability Legislation", 1554 Jurist 19, and "Content Moderation and the Use of AI, and Injunction - Comparison with EU Trends", 132 Hanrei-Jiho no. 2508 (both in Japanese).

[2] This terminology has become used in recent years. *See* Grimmelmann J, "The Virtues of Moderation" (2015) 17 Yale Journal of Law and Technology 42, which analyzed generally online moderation practices, but does not use the two-word phrase "content moderation".

before their publication, deleting them after publication, and/or freezing and deleting the users' accounts.

It may be a global trend[3], as a policy agenda for coping with illegal and harmful content[4], to impose the de jure or de facto requirement for HSP to implement automated Content Moderation i. e. automatic screening and filtering or blocking content process using automated tools implementing advanced technology including artificial intelligence (such advanced technology is hereinafter called collectively "AI" and AI-based Content Moderation is called "ACM")[5].

In fact, it is certainly an urgent global common issue that as a measure against illegal or harmful information, whether to allow, promote, obligate, or conversely to limit use of ACM by HSP legislatively or administratively, taking into consideration that using ACM is becoming more efficient, but still leaves limits in the accuracy and trustworthiness. Furthermore, the issue extends to the way of enforcement by courts or administrative authorities, especially how far they can/shall award injunctive orders reflecting the HSPs' ability of suspension of illegal content using ACM.

There have been various discussions on the risk of excessive deletion of legal information by Content Moderation and mitigations of this risk[6] and how to avoid actions affecting fundamental rights such as private censorship[7] and the overall chilling effect on freedom of expression by Content Moderation.

This article does not go deep into the policy principles and legislation to protect against these over- or erroneous removals and mitigate their adverse systemic effects.

Instead, since HSPs' use of advanced technologies to avoid under-removal for the measures against illegal information is expected, we mainly focus on the issue of whether HSP can be required to generally (and comprehensively) monitor their users' content uploads and how HSP's limitation of liability is affected by the Content Moderation, especially ACM. When the illegal information is detected accurately by ACM, if such detection is considered as knowledge or awareness of an HSP, the HSP

---

3 G. Frosio and S. Mendis, "Monitoring and Filtering: European Reform or Global Trend?" in Oxford Handbook of Online Intermediary Liability, G. Frosio, Ed. May 2020. doi: 10.1093/oxfordhb/9780198837138.013.28.

4 Both terminology of illegal content and harmful content harms society at large or individuals. In this article, "harmful content" is used to mean "legal but harmful" or "harmful but not illegal or unlawful" in terms of criminal or civil proceedings. For example, not all discriminately content is unlawful or illegal, but unethical. *See* e. g. ACM Code of Ethics and Professional Conduct, where "harm" is broadly defined as negative consequences, especially when those consequences are significant and unjust.

5 *See* e. g. R. Gorwa, R. Binns, and C. Katzenbach, "Algorithmic content moderation: Technical and political challenges in the automation of platform governance," Big Data & Society, vol. 7, Art. no. 1, Jan. 2020, doi: 10.1177/2053951719897945.

6 Envisaged remedial actions include how to achieve transparency and accountability related to the Content Moderation, how to mitigate technical limitations of accuracy and reliability of the diagnostic results of AI used in ACM especially for context-dependent expressions, the availability of remedial scheme for users whose post and/or account is suspended or deleted due to erroneous diagnosis and determination in the Content Moderation process.

7 *See* e. g. K. Langvardt, "Regulating Online Content Moderation" Georgetown Law Journal, vol. 106, Art. no. 5, Jun. 2018.

is forced to delete the detected illegal information. If ACM is reasonably effective (with little systemic risk of over- and erroneous removals and other biproducts), naturally it becomes an option for the legislature to require or for the court to order the HSP to implement and utilize the ACM within the range of its effectiveness by a legislation or an injunction.

In this article we first define problem involving Content Moderation, then review the current state of Japanese legal system regarding Content Moderation including ACM by HSP and its relationship with limitation of liability of and injunctions against HSP. Next, the Japanese status will be evaluated in comparison with recent EU legal system updates and judicial precedents. Finally, it concludes with extraction of issues.

2 Problem

2.1 Definition of Content Moderation

The proposed EU Digital Services Act[8] ("DSA") Article 2(p) defines Content Moderation as follows:

> "… the activities undertaken by [HSPs] aimed at detecting, identifying and addressing illegal content …, provided by recipients of the service, including measures taken that affect the availability, visibility and accessibility of that illegal content …, such as demotion, disabling of access to, or removal thereof, or the recipients' ability to provide that information, such as the termination or suspension of a recipient's account[9]".

In the DSA, Content Moderation is not required to be automated, but as we will see later, it is assumed to be ACM that effectively uses AI.

2.2 General monitoring

In the process of ACM, an HSP systematically determines (flags) targets or candidates for filtering or blocking from all texts, images, and other information.

---

[8] Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM(2020) 825 final

[9] The definition of Content Moderation covers not only illegal content, which directly affects the civil and criminal liability of HSPs, but also information incompatible with their terms and conditions. Wilman F, The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US (Edward Elgar Publishing 2020) paras. 8.21-8.33 discusses intermediaries contractually prohibiting users from providing certain types of content and actively enforcing the prohibitions but distinguishes it from 'privatized' enforcement of illegal information by intermediaries for satisfying their legal requirement or for obtaining limitation of liability.

If we would allow voluntary ACM by HSPs, that means whether we accept HSPs' general, exhaustive and comprehensive monitoring all such information posted by us and following automated judgment.

On the other hand, even if we can rely on the ability of HSPs through ACM to make accurate judgments, when such a general monitoring is a legal requirement, constitutional or other higher norms become issues.

Although the Japanese Provider Liability Limitation Act (hereinafter "PLLA")[10] does not provide explicitly for whether there is a duty to generally monitor, it is interpreted that HSPs do not owe such a duty by confirming that there is "no awareness of information distribution" as an eligibility for limitation of liability under Article 3(1) PLLA[11].

The DSA maintains (Article 7) the prohibition of the general monitoring obligation in Article 15(1) of the Electronic Commerce Directive[12] ("ECD") which states: no general obligation to monitor the information which [HSPs] transmit or store, nor actively to seek facts or circumstances indicating illegal activity shall be imposed on those [HSPs] The first subparagraph of Article 17(8) of the Digital Single Market Copyright Directive[13] ("DSMCD") also states that: the application of [Article 17] shall not lead to any general monitoring obligation.

Under Section 512(m) of the U.S. Digital Millennium Copyright Act [14]("DMCA"), which is entitled "Protection of Privacy", nothing in that Section should not be interpreted as a condition for the application of the exemption for monitoring of services or active detection of infringement by HSPs.

In other words, neither Japan, the EU member states, nor the U.S. imposes a general monitoring obligation on HSPs.

Nevertheless, in fact, "voluntary" Content Moderation conducted by HSPs are prevalent, and specific laws e. g. regulation on terrorist content[15] require the proactive removal of certain specific information.

2.3 Effect on Recognition by Detection

According to the DSA definition, Content Moderation is a content monitoring activity that starts with the detection of illegal information. Since the monitoring is done by

---

[10] Act on the Limitation of Liability for Damages of Specified Telecommunications Service Providers and the Right to Demand Disclosure of Identification Information of the Senders (Act No. 137 of November 30, 2001)

[11] *See* Explanatory Note (in Japanese) by Ministry of Communication and General Affairs at https://www.soumu.go.jp/main_content/000671655.pdf (last accessed 2022/5/15)

[12] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 (OJ L 178/1). *See* generally Wilman (fn.9) Chapter 3.

[13] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (OJ L 130/92).

[14] 17 U. S. C. s512.

[15] *See* e.g. Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online (OJ L172/79).

human or automated systems, the question is whether it is possible or not, and the legal effect of the recognition of facts on the distribution of the information and detection and judgment of its illegality on the part of HSPs.

If HSPs are made aware of illegal information by the Content Moderation but leave it undeleted, they can be held liable for tort of omission against the victim or criminal liability for the omission. As a general consequence, HSPs are more likely to be reluctant to detect illegal information in order to avoid encountering illegal information. HSPs face less legal risk if they do not encounter illegal information in order to avoid obtaining knowledge or awareness of illegal information. Also, in the case of a legal system in which the provider's liability is limited if there is no knowledge or awareness of illegal information (knowledge-based-liability), the provider loses its immunity upon obtaining it. This is a situation described as the Good Samaritan paradox or the moderator's dilemma, but if we want to strengthen the countermeasures against illegal information via HSPs, we need to maintain the immunity and allow HSP's Content Moderation with certain monitoring activities.

Article 6 DSA prohibits HSPs from being deemed ineligible for immunity solely due to carrying out voluntary own-initiative investigation. Preamble 25 DSA states policy reason to clarify this effect "[i]n order to create legal certainty and not to discourage activities aimed at detecting, identifying and acting against illegal content that providers of intermediary services may undertake on a voluntary basis".

In addition, it would be useful to limit the liability of HSPs to the user such as removal or disabling access to content posted by the user. Article 3(2) PLLA, and Section 230(c)(2) of the U.S. Communications Decency Act are examples of legislation that limits the liability of HSPs against information content provider. However, if there is a contractual relationship with the user, the effect is limited to confirming the validity of the exercise of contractual rights.

3 Content Moderation and Interference Preventive Injunction Claims in Japan

3.1 The Prohibition of Censorship and the Inviolability of the Secrecy of Communications under the Japanese Constitution and the Telecommunications Business Act

The Constitution of Japan guarantees the freedom of expression in Article 21, paragraph 1, prohibits censorship in the first half of paragraph 2, and states in the second half that the secrecy of communications shall not be violated.

Our Supreme Court has narrowly construed that a censorship prohibited under the Constitution is limited to prohibition after the administrative authorities' conduct of the comprehensive and general examination of the specific matters of expression prior to its publication[16].

[16] Sup. Ct., December 12, 1984, 1982 (Gyo Tsu) 156, 38 Minshu 12, p.1308).

In the Hoppo Journal Case[17], the Court applied that precedent to determine whether a preliminary injunction on publication could be issued and held that a court injunction on publication does not constitute the censorship, but as a prior restraint, the strict scrutiny standard applies.

In response to Article 21 Constitution, the Telecommunications Business Act[18] (hereinafter "TBA") prohibits censorship of communications handled by telecommunications carriers in Article 3 and does not allow anyone to violate the secrecy of such communications in Article 4 (1).

3.2 Provider Liability Limitation Act

The PLLA was enacted referring the US CDA, DMCA and ECD as mother laws.

An HSP cannot be exempted for damages under article 3(1) PLLA[19] if an HSP has knowledge of the distribution of information by telecommunications intended to be received by unspecified persons and knows that the rights of others have been infringed (actual knowledge) or has reasonable grounds for believing that he or she could have known that the rights of others were being infringed (constructive knowledge). It is not an article that forces HSPs to engage in private censorship, violates Article 21(1) Constitution, nor conflict with Article 3 TBA.

Article 3(2) PLLA is a safe harbor clause that exempts an HSP from liability for damages if the HSP takes measures such as deleting the sender's postings but only to the extent necessary to prevent transmission (i) where the HSP has reasonable grounds to believe that the rights of others have been unjustly infringed by the distribution of the information via the specified telecommunications, and (ii) where the HSP inquires the sender of the infringing information as to whether the sender agrees to take down the information when the person who claims that his or her rights have been infringed requests the HSP to take down the information. The former is so-called Good Samaritan Takedown immunity, while the latter is Notice-Notice & Takedown (NN&T) immunity.

Although the PLLA does not exempt HSPs from criminal liability, it is understood that they will not be held criminally liable unless there is a special involvement with a crime such as opening a BBS specializing in child pornography.

There is no statute that explicitly allows a victim to request for an injunction against HSPs from publication of user postings that infringes the victim's rights, but as described below, future interference by such a posting may be prevented from recurring by an injunctive order of the courts.

---

[17] Sup. Ct., June 11, 1986, 1981 (O) 609 of 1982, 40 Minshu 4, p.872).

[18] Telecommunications Business Act (Act No. 86 of December 25, 1984)

[19] Even if an HSP is not exempted for damages under PLLA, it can still be held that the HSP is not liable under general tort law principle.

3.3 Guidelines for the PLLA and Illegal Information and Model Terms and Conditions

HSPs in Japan are not held liable for damages[20] for leaving or removing infringing information as long as the Content Moderation complies with the PLLA.

In Japan, there are soft-law-guidelines and standard terms and conditions that can be referred to in each field (Table 1).

Regarding information infringing rights covered by PLLA, in 2002 a self-regulatory body was established by rightholders' organizations and business organizations such as providers': guidelines on defamation and privacy, copyright infringement, trademark infringement[21].

In 2006, telecommunications-related organizations established a self-regulatory consortium[22] and formulated the "Guidelines for Dealing with Illegal Information on the Internet," and their member providers have implemented self-regulatory measures to takedown illegal information, including information that infringes on social interests. In addition, the "Model Terms and Conditions for Internet Services Concerning Response to Illegal and Harmful Information"[23] has been formulated, which lists information that is offensive to public order and morals in addition to illegal information and describes how to respond to such information.

Table 1. Soft-law-Guidelines relating to Content Moderation in Japan

Subject of Content Moderation by HSP	Source of rights/violations	Source of Hard Law	Style of softlaw/self-regulation
Infringing information covered by PLLA	defamation/ privacy violation	Civil Code	Codes of Conduct
	copyright infringement	Copyright Act	Codes of Conduct (Notice and Takedown)
	trademark infringement	Trademark Act	Codes of Conduct (Notice and Takedown)
Illegal Information other than infringing information	Obscene materials etc.	Penal Code	Codes of Conduct
harmful information	----------	Civil Code	Model Terms and Conditions

However, these softlaw-guidelines do not assume that HSPs will actively engage in voluntary Content Moderation or utilize AI in Content Moderation.

---

[20] In the EU, as defined by the DSA, Content Moderation is not only for information that infringes civil rights, but also for information that is criminally illegal or harmful (legal-but-harmful) that does not fall under these categories.

[21] *See* these Guidelines (in English) Provider Liability Limitation Act Guidelines Review Council at https://www.telesa.or.jp/consortium/provider/pconsortiumproviderindex_e-html

[22] *See* https://www.telesa.or.jp/consortium/illegal_info

[23] *Ibid.* English version of model terms and conditions and explanatory guide is available.

3.4 Recommendations for the Decade Review of PLLA

The Recommendation for the Decade Review of PLLA (hereinafter "Review Recommendation")[24] states that, since the ECD does not impose a general monitoring obligation on providers and such monitoring is not a condition for liability limitation under the US DMCA, it is necessary to consider the following points:

1. it is not legally appropriate to require [HSPs] to monitor information in circulation, including by technical means, as it may result in de facto censorship, and be liable to invite a significant chilling effect on freedom of expression,
2. may infringe the secrecy of communications, and
3. it is often impossible for providers to bear the burden of monitoring.

Review Recommendation continues: if imposing the monitoring obligation is denied, the introduction of technical monitoring means should not be a requirement for the limitation of liability under the PLLA.

Further, as for the monitoring of information for which there was a demand to prevent retransmission of the infringing information, it is sufficient to take measures to prevent transmission when the information becomes known to be circulated *ex post* because of chilling effect on freedom of expression and the reality of feasibility of implementation of such monitoring measure.

In addition, technical measures at [that] point can only confirm the identity of the copyright infringing material, and it is difficult to make the introduction of such measure mandatory from the viewpoint of cost.

Furthermore, it is not reasonable to easily expand the interpretation that [HSPs] who does not know the circulation of individual information is liable as a sender under PLLA as a person who placed the information in circulation[25].

The Review Recommendation also rejected the voluntary monitoring and deletion, introduction of technical measures such as filtering of illegal uploads, and implementation of reasonable measures such as suspension of service for users who repeatedly and continuously commit infringement as the requirements for limitation of liability of providers.

3.5 Cases requesting for preventive injunction against HSP

In Japan, there is no judicial precedent that an HSP was ordered to prevent its users from posting in the future while holding that it is a typical HSP whose liability for damages are limited under PLLA.

As to specificity of injunction, although it is technically closer to a search engine than HSP, we shall refer File Rogue (Napster-like Hybrid P2P system) case[26]. The file for which the P2P file exchange should be suspended is copied in the MP3 format with the file information in which both the "original title" column and the "artist"

---

[24] Published in July 2011

[25] Expressing concerns on TV Break case (fn.27) below.

[26] Tokyo Dist. Ct. December 17,2003, 2002 (Wa) 4237

column of the music list are described. It seems to be a simple specification to pay attention to, but it is necessary to use advanced text filtering technology because it includes all the combination of letters (Kanji, Hiragana, Katakana, and uppercase and lowercase letters of the alphabet are not specified in the judgment). On the other hand, the suspension of transmission and reception of all the managed works requested by the plaintiff the Japanese Society for the Rights of Authors, Composers, and Publishers (JASRAC) was not allowed as the request for abstract prohibition of copyright infringement itself.

In the TV Break (initial name was Pandora TV) video posting service case[27], the defendant was technically an HSP like YouTube, but the IP High Court held that copies were made by the HSP using the copying act by the user and deemed a subject who directly infringes the copyrights of plaintiff JASRAC for making the public transmission of the music listed and provided by JASRAC. The court also enjoined future reproduction and public transmission without narrowing down the attributes of the target files of the copyrighted work. Defendant=Appellant HSP argued that such an injunction was a comprehensive injunction effectively ordering deletion of legitimate files and has a chilling effect on the freedom of expression of video posting sites and the development of the culture that is the purpose of Copyright Act. However, the court dismissed it because it limited to video files that were copies of JASRAC managed works uploaded without a license. In addition, the fact that technical infringement avoidance measures and deletion measures such as hash matching of video files that infringe copyright and audio fingerprint matching are not adopted by HSP was taken into consideration as a negative factor in the judgment of infringement. In other words, it is a case where the injunction order is affected by not doing ACM using automated technology.

As to personality rights infringement, in the Animal Hospital case[28], it was confirmed that an infringement preventive injunction against a bulletin board operator was possible by citing the Hoppo Journal case[29], but under the fact of the case, the request for injunction of reappearance of the same wording as the defamatory statements was rejected.

3.6 Urgent Recommendations on How to Deal with Slandering on the Internet

"Urgent Recommendations on How to Deal with Slandering on the Internet" (August 2020)[30](hereinafter “Urgent Recommendation”) calls for proactive voluntary

---

[27] IP High Ct. September 8, 2010, 2009 (Ne)10078

[28] Tokyo Dist. Ct. June 26, 2002, 2001 (Wa) 15125. On appeal, Tokyo High Court also found that the deletion operation of the bulletin board, including the deletion guidelines set by the operator, was extremely inadequate as a remedy for victims and does not affect his liability. Tokyo High Ct. December 25, 2002, 2002 (Ne) 4083

[29] (n.17)

[30] The Study Group on Platform Services of the Ministry of Internal Affairs and Communications.

efforts in combating slander and defamation, considering the role that HSPs play in daily life and social and economic activities.

Regarding Content Moderation (deleting or hiding illegal[31] posts and suspending accounts by HSPs), the following responses and ideas are given (emphasis by the author):

- In addition to setting up an easy-to-understand system for reporting deletion requests, etc., [HSPs] will take prompt action such as deletion in response to reports from affected users.
- Since a large amount of information is circulated on [HSP's] service, it is assumed that [HSPs] will **at their own initiative generally and always monitor a large amount of information**, **find infringing information** (illegal information), and **take prompt action without waiting for reports from users**.
- Although it is not appropriate to uniformly require implementation of [AI-based ACM], **when** in the future technology that utilizes **AI** algorithms **becomes widespread and progresses**, and **costs are reduced**, making it easier to introduce it, it is also expected that **information will be deleted at its own initiative based on the rules and policies** stipulated in the freely designed service of the platform operator **without a report from the user or a third party** ...
- [HSPs] will **promptly determine whether to remove information** in response to reports from users as well as **reports from government agencies with legitimate authority and expertise**.

The government will work with HSPs and develop an environment to support the smooth implementation of various initiatives related to slander and libel in HSPs, including a certain legal framework.

In relation to Content Moderation, although making [content removal through ACM] mandatory will require extremely careful judgment, the application of the PLLA will be reconsidered in a timely manner, considering the spread and progress of [AI], the accompanying changes in the cost burden of HSPs, and the changes in users' expectations of the roles required of HSPs.

3.7 Summary：Content Moderation in Japan

Except for the File Logue and TV Break cases, there is no case law suggesting the relationship between Content Moderation and injunctive relief, but courts have no hesitation in issuing an injunction which would require general monitoring once the defendant is found to be out of the scope of HSP under PLLA, in other words, is

[31] "In the case of defamatory information that is short of constituting infringement of rights (legal-but-harmful information), while taking measures **to prevent the chilling effect on freedom of expression due to over-removal** etc., and to avoid **unreasonable private censorship**, it is expected that various countermeasures will be taken **autonomously** by HSPs according to the scale and specifications of the service." Surprisingly, it names HSPs as the entity who should consider and care the balance between merits of countermeasures and, "chilling effect on freedom of expression" or "unjustified private censorship.

deemed a publisher, even if the injunction in effect requires the HSP's implementation of AI.

As for the legislative mandate for general monitoring of HSPs, the reluctance of the Review Recommendation to impose such a mandate because it is "de facto censorship" due to the significant chilling effect on freedom of expression, the risk of infringement of the secrecy of communications, and the cost of monitoring HSPs, has not been completely reversed until now.

On the other hand, regarding the voluntary Content Moderation of illegal information by HSPs, it is expected that it will be accompanied by the constant monitoring of a large amount of information in circulation, provided that the introduction of AI, will become easier due to the progress of technology diffusion and cost reduction.

4 Recent EU Legislation and Case Law

For comparison, we shall examine below recent legislative and judicial trends in the EU that relate to Content Moderation or envisage the use of AI.

4.1 DSMCD Article 17

Article 17(3) DSMCD excludes the application of the limitation of liability under Article 14(1) of the ECD to certain HSPs.

Article 17(3) DSMCD excludes certain HSPs from the limitation of liability under Article 14(1) ECD. Article 17(4)(b) of DSMCD states that HSPs must use their best endeavours, with professional diligence and high industry standards, to implement upload blocking of works for which relevant and necessary information is provided by the right holder. In effect, the right holders are expected to provide data or fingerprints of their works for automatic matching by HSP. In addition, the Article (4)(c) imposes on HSPs a best-effort obligation for prompt removal of individual works from websites (takedown) and future upload blocking (stay-down) in accordance with (4)(b) upon receipt of a sufficiently specific notice from the right holder.

However, according to Article 17(7), upload blocking of non-infringing works due to exceptions or limitation must be avoided. For this reason, in its guidance on the application of Article 17 based on Article 17 (10)[32], the European Commission stated that automated blocking should in principle be limited to cases of manifest infringement, and that other cases should in principle be first published subject to ex-post human review based on notification by the right holder.

As mentioned above, since the first subparagraph of Article 17(8) states that it should not lead to the obligation of general monitoring, it clarifies the position that ACM is "specific" monitoring as opposed to "general" one.

---

[32] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL Guidance on Article 17 of Directive 2019/790 on Copyright in the Digital Single Market COM/2021/288 final

4.2 *YouTube and Cyando* joined Cases[33]

In the *YouTube and Cyando* joined cases, CJEU held that Article 8(3) Information Society Copyright Directive[34] (hereinafter "ISCD"), which obliges Member States shall ensure that rightsholders are in a position to apply for an injunction against [HSPs] whose services are used by a third party to infringe a copyright or related right does not preclude the fulfilment of a prerequisite for a claim for an injunction under the national law of a Member State, such as German interferer liability (Störerhaftung), which before court proceedings are commenced, that infringement has first been notified to [an HSP] and [the HSP] has failed to intervene expeditiously in order to remove the content in question or to block access to it and to ensure that such infringements do not recur.

The rightsholder will be given the opportunity to terminate the infringement out of court and prevent its recurrence and will not lose the right to seek an injunction. In addition, HSPs do not need to proactively monitor uploaded user content to avoid injunction claims, thus complying with Article 15(1) ECD.

Article 17 DSMCD seems to be formulated by incorporating German interferer liability (Störerhaftung) system which requires the private enforcement of out-of-court removal and staydown request against HSPs before court proceedings. Article 17 DSMCD will facilitate such an out-of-court enforcement against HSPs, when it is expected that they have efficient ACM systems.

4.3 *Glawischnig-Piesczek* Case[35]

Eva Glawischnig-Piesczek (EGP), a senior official of the Green Party in Austria, requested Facebook Ireland (FB) to remove the defamatory remarks commented with the portrait photo of EGP. The Court of First Instance issued an interim order to remove the words that had the same or equal meaning as the comments, and after an appeal, the Austrian Supreme Court referred the case to the CJEU for a preliminary ruling.

CJEU found that the ECD, in particular, Article 15(1), did not preclude national courts from (emphasis added):

–ordering [an HSP] to **remove** information which it stores, the content of which is identical to the content of information, which was previously declared to be unlawful, [(**stored identical information [Ii]**)] or to block access to **[Ii]**, **irrespective of who requested the storage** of that information,

–ordering [an HSP] to remove which it stores, the content of which is equivalent to the content of information which was previously declared to be unlawful [(**stored equivalent information [Ie]**)], or to block access to **[Ie],**

---

[33] *YouTube and Cyando*, Joined Cases C-682/18 and C-683/18, ECLI:EU:C:2021:503, 22 June 2021.

[34] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (OJ L 167/10)

[35] CJEU, C-18/18, *Glawischnig-Piesczeck*, ECLI:EU:C:2019:821, 3 October 2019

- provided that the monitoring of and search for the information concerned by such an injunction are **limited to information conveying a message the content of which remains essentially unchanged** compared with the content which gave rise to the finding of illegality and **containing the elements specified in the injunction**, and
- provided that the **differences in the wording** of that equivalent content, compared with the wording characterizing the information, which was previously declared to be illegal, **are not such as to require [HSP] to carry out an independent assessment** of that content.

As there is a genuine risk that information which was held to be illegal is subsequently reproduced and shared by another user of that network, an injunction ordering the deletion or blocking of **[Ii]** is justified, irrespective of who requested the storage of that information[36].

For granting an effective remedy, the Court held that the injunction must be extended to **[Ie]** but avoid imposing excessive burden on HSP in so far as the monitoring of and search for information which it requires are limited to information containing the elements specified in the injunction, and its defamatory content of an equivalent nature does not require the host provider to carry out an independent assessment, since the latter **has recourse to automated search tools and technologies**[37].

The injunction of both **[Ii]** and **[Ie]** cannot be regarded as general monitoring in violation of ECD Article 15(1)[38].

The significance of CJEU findings regarding ACM is that the court seems to believe that AI is enough advancing, and it is easy for a national court to understand AI algorithm and can easily depend on HSP's recourse to AI specifying injunction order.

4.4 DSA

As mentioned above, the DSA defines Content Moderation and maintains the Article 15(1) ECD which prohibits the Member States from requiring HSPs to conduct general monitoring.

4.5 Summary: Content Moderation in the EU

We believe both YouTube and Cyando and Glawischnig-Piesczek[39] judgements will facilitate out-of-court request for removal and staydown content identical to and equivalent to illegal information using AI.

The main legal interest protected by Article 15(1) ECD is the freedom to conduct of business by avoiding excessive burden on HSPs, but it is also considered to func-

---

[36] Id. para. 36-37

[37] Id. para. 46

[38] Id. paras 37, 47

[39] The Glawischnig-Piesczek case is also cited in the DSA and remains valid even after the law is enacted as an interpretation of the legality of the injunction in Article 7 DSA.

tion to protect, at least indirectly, the fundamental rights of users' personal data, privacy and freedom of expression[40].

Article 6 and preamble 25 DSA will be expected to create legal certainty and not to discourage good faith voluntary own-initiative Content Moderation by HSPs for illegal information or for compliance. We need to see if this level of assurance for immunity works well.

5 Comparison: ACM environment between the EU and Japan

This section compares the legal systems from Content Moderation to injunction and their discussions in the EU and Japan, as, again, discussion in Japan has referred to EU status.

What has controlled legitimacy of monitoring and filtering in a mandatory fashion in Content Moderation process is, in the case of the EU, Article 15(1) ECD, which protects the legal interests of freedom to conduct business of HSPs and (indirectly) fundamental rights of users, while in the case of Japan, "de facto censorship," "significant chilling effect on freedom of expression," "secrecy of communication," and "burden on [HSPs]" have been also lined up.

The EU has so far decided not to make Content Moderation general obligation except for individual legislations, but Japan will consider it according to the spread and progress of AI, the reduction of cost burden of HSPs, and changes in users' expectations, although it requires "extremely careful scrutiny".

In Japan, if we take the direction to make upload blocking best effort obligation and to incorporate private out-of-court injunction into Content Moderation practice as private law enforcement measures like Article 17 DSMCD, an extremely careful scrutiny is necessary. Even if it is not mandatory, a government mandated system that promotes "voluntary" deletion by Content Moderations but follows administrative authorities' *ex post* control to avoid under-removal would have similar effects. The current debates can be regarded as a legitimate direction to the extent that it aims to increase the transparency and accountability of HSPs regarding the risk of over- and erroneous removals by ACM and to ensure relief for users, at the same time it aims to promote ACM in order to make measures against a large volume of illegal information more effective.

While Article 18(1) ECD has limited the scope of injunction requests for infringing information, there is no case in Japan where such a restriction has been applied to typical HSPs. We may expect few cases will continuously be reported, if ACM advances. As seen from YouTube and Cyando judgment, the spread and advancement of ability of AI-based Content Moderation will increase the ratio of out-of-court dispute resolution. That means courts will handle only difficult cases, which requires issuing injunction with detailed (algorithmic) instruction to prevent infringement. In order to avoid the risk of over- and erroneous removals by AI on the part of HSPs, as

[40] *See* Wilman, fn.9, paras.3.25-3.31

a natural result from the Glawischnig-Piesczek precedent, judges and courts will be required to have literacy on and technology related to AI preferably inhouse.

6 Conclusion: Issues for Japanese Law

6.1 General Monitoring

In Japanese law, general monitoring by HSPs is not mandated by the Constitution or TBA but rather prohibited in principle. However, we have not discarded the option of ordering HSPs to be proactive in Content Moderations, as discussed in our Urgent Recommendations. It would be necessary to reconfirm the necessity of proactive Content Moderation. Since the Japanese Constitution prohibits censorship and judicial precedents require that prior restraint meet the strict scrutiny standard, "extremely careful scrutiny" is required, but referring to the global trend especially in the EU, the following issues need to be cleared for such consideration.

6.2 Confidentiality of communications

There is no issue of secrecy of communication in the first place if the Content Moderation takes place only *ex post* (after posting). If the proactive *ex ante* Content Moderation can be legalized after strict scrutiny, the violation of the secrecy of communication under TBA should be exempted.

6.3 Excessive burden on HSPs

HSPs shall not be required to implement specific type of AI for Content Moderation, which requires excessive burden on and significant investment from HSPs, whether based on an individual legislation or injunctions ordered by a court.

6.4 Development of tools to limit liability for proactive measures

It is desirable to clarify legislatively the impact of the execution of ACM on knowledge-based civil and criminal liability.

More specifically, Article 3(1) PLLA clarifies that without awareness or constructive knowledge of infringing information, there is no liability for damages. In this regard, if it is confirmed that awareness under Article 3(1) PLLA do not arise solely because of the execution of voluntary Content Moderation, the range of options for HSPs will be further expanded, because they are free to design up to the stage of mechanical detection of illegal information using AI. For example, based on the "manifestly" illegal criterion, i.e., the certainty that the information is illegal, as stated in the Article 17 DSMCD Guidance[41], HSPs will be able to distinguish between cases where

[41] fn. 32.

the information should be removed immediately, cases where it should be flagged and manually checked again, and cases where it should not[42].

The above knowledge and design standard should be supplemented by existing softlaw-guidelines in detail preferably with the participation of very large global HSPs.

[42] The limitation of liability to the sender for erroneous removals may be considered already covered in Article 3(2) PLLA. It would be desirable to clarify the relationship between "reasonable grounds" in Article 3(2) and erroneous-removals in AI-based Content Moderation in the soft-law guidelines.