Risks of Filtering Requirements for Online Expression and Privacy

By: Corynne McSherry and Christoph Schmon, Electronic Frontier Foundation

Case:	2021Heonma290 Constitutional complaint regarding Article 22-5(2) of the Telecommunications Business Act, etc.
Claimants:	Mr. Kim and 3 others
Interested party:	Korea Internet Corporations Association

The Electronic Frontier Foundation ("EFF") is a member-supported, non-profit civil liberties organization that works to protect digital rights. For over 30 years, EFF has represented the public interest in ensuring that law and technology support human rights. In the US and abroad, we have worked to ensure that internet policy, legislation, and technological measures appropriately balance the rights of all internet users. As a legal services organization, we also counsel individuals and companies whose legitimate activities may be undermined by filtering proposals and requirements.

It is our understanding that the challenged provisions of the Telecommunications Business Act and the Enforcement Decree requires business to take technical measures to prevent circulation of "illegally filmed contents". In our experience, as a practical matter service providers will attempt to meet their legal obligations by reviewing all user content, often called "general monitoring", and using technical measures such as content filters. Below, we outline several principal concerns regarding the potential impact of such requirements on human rights, particularly online expression and privacy, as well as competition. To be clear, we are not experts in South Korean law nor the legislation at issue. We submit this declaration in the hopes that our experience with comparable regulations may be useful to the Court's determination in this case.

A. General Monitoring, Filtering and Online Expression

1. Risk of censorship due to expansive content moderation obligations

The increasingly powerful role of service providers in modern society has prompted a host of policy concerns. One key policy challenge is defining online intermediaries' legal liability for harms caused by content generated or shared by—or activities carried out by—their users or other third parties.

Unfortunately, laws that impose such liability inevitably result in the censorship of lawful and valuable expression. Stringent liability laws for online intermediaries encourage service providers to affirmatively monitor how users behave; filter and check users' content; and remove or locally filter anything that is controversial, objectionable, or potentially illegal to avoid legal responsibility. The effects are especially present where service providers obligations are unclear or broadly defined.

Faced with expansive and vague moderation obligations and major legal consequences if they guess wrong, companies inevitably overcensor. Stricter regulation of and moderation by platforms also results in self-censorship, as users try to avoid negative repercussions for their artistic and political expression. Numerous studies document that when people believe their communication is being monitored, they self-censor both their expression and the content they seek out and read.[1]

2. International experiences – the example of European internet legislation

EU governments confronted this problem in the negotiations over the EU's new internet bill— the Digital Services Act or DSA. The DSA seeks to articulate clear responsibilities for online platforms, with strong enforcement mechanisms, while also protecting users' fundamental rights. The DSA encourages "good Samaritan" content moderation and sets out type- and size-based due diligence obligations, which include obligations relating to transparency in content moderation practices, algorithmic curation, and notice and action procedures.

For instance, the DSA's transparency requirements mandate that users be informed about a platform's content moderation practices. Terms of service of online platforms must contain details about the utilization of automated decision-making processes and the extent of human oversight. In order to address illegal content online, all providers of hosting services, regardless of their size, must put in place notice and action mechanisms that facilitate the notification of potentially illegal content, after which platform providers can decide whether to take action with regard to the notified content. In all cases where content is removed, whether or not the removal decision was based on a notification or on platform's own investigations, users are entitled to know the rationale behind the decision and must be given options to appeal the decision. The DSA sets out several safeguards to ensure that content removal decisions are targeted and consider users' rights to freedom of expression and of information as well as their right to privacy and non-discrimination. Users also enjoy a right to reinstatement if platforms wrongly remove their content.

Several national bills in Europe were presented ahead of the DSA negotiations such as the controversial "Avia Bill" in France. The new law required social media intermediaries to remove obviously illegal content within short time frames and was met with criticism from experts and civil society. An intervention before the French Supreme Court, co-organized by the Electronic Frontier Foundation, proved ultimately successful, as the Court struck down the law's requirements to remove infringing content within 24 hours[2], recognizing that platforms would be encouraged to remove perfectly legal speech. Other national bills, such as the Austrian hate speech law also prompted question about 'overblocking'[3] of legitimate expression and non-compliance with overriding EU principles.

The draft versions of the Digital Services Act showed sympathy for a variety of legal solutions to address problems of online safety and the sharing of illegal content. The initial proposal

[1] *See generally* N. Richards, *Why Privacy Matters*, Oxford Press (2022)
[2] https://www.eff.org/press/releases/victory-french-high-court-rules-most-hate-speech-bill-would-undermine-free-expression.
[3] https://www.euractiv.com/section/data-protection/news/austrias-law-against-online-hate-speech-question-marks-in-the-home-stretch/.

suggested to let substantiated user notices suffice to trigger removal obligations for online platforms thus resulting in potential secondary liability for user content. Other versions suggested short deadlines for the removal of content. Ultimately, considering the negative experience with national initiatives and striving to achieve a fair balance between the various interests at stake, including freedom of expression rights by users, the final version abstained from any filter mandates and short inflexible deadlines for the removal of content.

The final legislative deal is widely considered being in line with the principles that underpin the e-Commerce Directive, the EU's previous backbone legislation for internet regulation. The DSA upheld the important principle that States should never mandate platforms to monitor user communication as this would inevitably lead to over-removal of content, undermine free speech rights and ignore users' right to privacy. Addressing the spread of illegal content online and the societal risks posed by the dissemination of disinformation or other harmful content, EU lawmakers opted for a harmonized notice and action system. Online platform providers must establish mechanisms that allow for the submission of detailed and substantiated notices. Upon receiving a notice from users or entities alerting them to the presence of illegal content, they are obligated to act. Platform providers must make a timely decision regarding the restriction or removal of such content and inform users if removal decisions are made. If notices contain detailed and precise information for diligent hosting service providers to identify the content as illegal without conducting a thorough legal examination, these providers will lose the DSA's liability exemption if they fail to act.

From this follows that under EU online platform rules, liability for speech continues to rest with the speaker and not with platforms that host what users post or share online. However, online platform providers are required to put in place processes that help tackle the dissemination of illegal content online.

More recent national bills that don't live up to this standard, such as the draft UK online safety bill, are widely attacked[4] as violating the fundamental rights of user.

3. Human Rights Principles: the Need for Check and Balances

Freedom of expression and online privacy are fundamental rights under the EU Fundamental Rights Charter, the European Human Rights Charter, and they are also protected under other instruments of international human rights law. Any State measure that aims to interfere with these rights for the sake of protecting another public value, such as the avoidance of harmful content online, must seek to strive a fair balance between these objectives.[5] As a general principle, States measures should have a clear legal basis and be necessary in a democratic society and proportionate, meaning that they cannot go beyond of what is necessary to achieve the objectives. They should also be sufficiently safeguarded.

As far as the core of the relevant contested provisions under the Korean Telecommunication Business Act is concerned, it is our understanding that there is a legitimate concern that not enough attention has been paid to fundamental rights aspects, including the right to freedom of

[4] https://www.eff.org/deeplinks/2022/08/uks-online-safety-bill-attacks-free-speech-and-encryption.
[5] ECHR, Perinçek v. Switzerland [GC], § 274; Case C-275/06 Promusicae [2008] ECR I-271.

expression. The bill sets out a positive obligation to monitor and restrict content if they match content reviewed and decided to be illegal by a specific body.

Without sufficient check and balances, such an obligation will likely lead to the indiscriminate general monitoring of all user content and the deployment of error-prone matching technology, such as automated filter systems that automatically prevent relevant content from being uploaded. Under human rights doctrines, any legal framework that provides for or demands blocking measures should ensure that measures strictly target the illegal content and has no arbitrary or excessive effects.[6]

4. Likelihood of Error

Technical measures to screen content are also prone to error, because no algorithm can replace or do the kind of contextual and legal analysis required to distinguish lawful from unlawful uses. This is one of the lessons of the EU Copyright Directive. EU lawmakers tried to mandate use of filters that can block infringement while also permitting lawful use. Everyone who testified in the implementation process, including filter vendors, said they could not do this. This means platforms will have to choose between over- or under-blocking.

Filters can also make purely technical errors, falsely identifying material as a duplicate of a protected work. Even the most expensively developed filters, like YouTube's ContentID, have problems like classical music recordings being falsely matched. Technical analysis of other filters has also found matching problems.

Even when filters perform as intended, they very frequently remove legal content because of human error, usually in the form of rightsholders claiming the wrong content. (Like claiming an entire nightly news broadcast or the film clip used in a movie review.) Many errors with YouTube's ContentID fall in this category.

B. General Monitoring, Filtering, and Privacy

General monitoring also undermines users privacy rights by requiring companies to collect abundant data about users, often without users' knowledge. International experiences show that any bill that requires online platforms to systematically filter user content create serious privacy and security risks. For example, in the realm of copyright enforcement mandated technical measures have shown to be inadequate to deal with risks of data breaches and in the worst case even compel platform operators to break encryption in order to scan the content of messages.[7]

These actions violate human rights standards. The UN High Commissioner for Human Rights has emphasized that any interference should be carried out only when authorized by an independent judicial body, on a case-by-case basis.[8]

---

[6] Cf. ECtHR, Application no. 10795/14 – *Kharitonov v Russia* (website blocking), at 45.
[7] https://www.eff.org/deeplinks/2022/11/filter-mandate-bill-privacy-and-security-mess.
[8] https://www.ohchr.org/en/press-releases/2022/09/spyware-and-surveillance-threats-privacy-and-human-rights-growing-un-report.

Finally, the EU's General Data Protection Regulation (GDPR) recognizes that filter system can have adverse effects on users, whose content is automatically removed or whose data are collected for the purpose of profiling.[9] Data protection principles thus require state measures to ensure safeguards for users' privacy, freedom of speech and other fundamental rights before any uploads are judged, blocked or removed.,

C. General Monitoring, Filtering and Competition

Monitoring requires intensive resource investments, which in turn discourages new companies from entering the field. Aside from the human resources require, technical measures such as upload filters are expensive to build and expensive to license. YouTube's ContentID had already cost that company over $100 million dollars as of five years ago. Audible Magic, which was promoted to EU lawmakers as an affordable solution, is widely reported to cost more in practice than that company represented.

Filters also reduce competition by creating technical lock-in. As Engstrom and Feamster report, a major consequence of adopting particular filtering technology is that companies design their other systems around that technology. Design choices, investments, or new lines of business can be precluded or rendered prohibitively costly as a result. It is one thing for companies to voluntarily put themselves in this position -- but it is quite another for it to be created by government mandate.

D. A Framework for Transparency and Redress

Adverse effects on human rights may be alleviated by a voluntary human rights framework for content moderation. Any decision about which content should or should not be shared online has serious human rights implications and online platforms can benefit from instructions to help operators to act more responsibly. EFF has long worked to provide guidance: In 2015, EFF, as part of an international coalition, helped launch the "Manila Principles on Internet Liability"[10], a framework of baseline safeguards and best practices based on international human rights instruments and other international legal frameworks. In 2018, EFF and partners then launched the "Santa Clara principles[11] on Transparency and Accountability in Content Moderation", which call on intermediaries to voluntarily adopt better practices. In 2021, a new version of the principles was developed, with a focus on adequately addressing fundamental inequities in platforms' due process and transparency practices for different communities and in different markets.

However, EFF recognizes that there's a need to strike a balance between addressing the very real issue of platforms hosting and amplifying illegal content while simultaneously providing enough protection to those platforms so that they are not incentivized to remove protected user speech, thus promoting freedom of expression. Our recommendations on how to best achieve this are guided by the rationale that it is in the best interest of all parties to focus on the regulation of processes on platforms rather than on speech and to make sure that mandatory content

[9] Article 22 GDPR.
[10] https://manilaprinciples.org/index.html.
[11] https://santaclaraprinciples.org/.

restrictions are always ordered by a judicial authority and are applied without resorting to intrusive filter systems.[12]

E. Conclusions

Intermediaries are vital pillars of internet architecture, and fundamental drivers of free speech, as they enable people to share content with audiences at an unprecedented scale. International experiences with internet legislation show the challenges for regulators and legislators to choose the right toolbox when addressing illegal content online. The EFF believes that the adoption of moderation frameworks that are consistent with human rights can best help to meet that challenge.

[12] https://www.eff.org/deeplinks/2022/05/platform-liability-trends-around-globe-conclusions-and-recommendations-moving.

필터링 의무의 온라인 표현과 개인정보 보호와 관련 위험성

작성자: 전자 프런티어 재단(Electronic Frontier Foundation)의 Corynne McSherry 및 Christoph Schmon

사건: 2021헌마290 전기통신사업법 제22조의5 제2항 등에 관한 헌법소원

청구인: 김○○ 외 3인

이해관계인: 한국인터넷기업협회

참고자료 2 번역본

전자 프런티어 재단("EFF")은 구성원이 지원하는 비영리 인권단체로, 디지털 권리 보호를 위해 노력합니다. 지난 30 년이 넘는 기간 동안, EFF 는 법과 기술이 인권을 옹호하도록 함에 있어 공공의 이익을 대변하였습니다. 미국 및 미국 외에서 본 재단은 인터넷 정책, 입법 및 기술적 조치가 모든 인터넷 이용자의 권리와 균형을 이루도록 노력해왔습니다. 법률서비스단체로서, 본 재단은 또한 필터링 제안 및 의무에 의하여 정당한 활동이 저해될 수 있는 개인 및 기업에게 자문을 제공합니다.

본 재단은 전기통신사업법 및 동법 시행령의 청구대상 조항이 기업에게 "불법촬영물등(illegally filmed contents)"의 유통을 방지하기 위한 기술적 조치를 취할 것을 의무화하는 것으로 알고 있습니다. 본 재단의 경험에 비추어, 현실적으로 서비스 제공자는 모든 이용자 콘텐츠를 검토하는 방식인 소위 "일반 모니터링(general monitoring)"을 실시하고 콘텐츠 필터 등 기술적 조치를 이용하여 법적 의무를 충족하고자 할 것입니다. 아래에서 본 재단은 그러한 의무를 부과하는 것이 인권, 특히 온라인 표현과 개인정보 보호를 비롯하여 경쟁 등에 미치는 잠재적 영향에 관한 몇 가지 주요한 우려를 기재하였습니다. 분명히 하자면, 본 재단은 한국의 법이나 쟁점 법안에 대한 전문가가 아닙니다. 본 재단은 유사한 규정과 관련한 본 재단의 경험이 본 사건에서 귀 법원의 결정에 유용할 수 있기를 바라며 본 진술서를 제출합니다.

A. 일반 모니터링, 필터링, 그리고 온라인 표현

1. 콘텐츠 조정의무 확대로 인한 검열위험

현대사회에서 서비스 제공자의 역할 강화는 일련의 정책적 우려사항을 유발하고 있습니다. 정책상 주요 난관 중 하나는 이용자 또는 기타 제 3 자가 생성 또는 공유하는 콘텐츠 또는 그들이 수행하는 활동에 의하여 발생하는 피해에 대한 정보매개자(intermediary)의 법적 책임을 정의하는 것입니다.

안타깝게도 그러한 책임을 부과하는 법률은 합법적이고 가치 있는 표현을 검열하는 불가피한 결과를 낳습니다. 정보매개자에 대한 엄중한 책임을 지우는 법은 서비스 제공자가 이용자 행동방식을 적극적으로 모니터링하고, 이용자의 콘텐츠를 필터링 및

검사하고, 논쟁적이거나, 불쾌하거나, 불법적인 잠재성을 가진 모든 사항을 삭제하거나 지역별로 필터링하여 법적 책임을 피하도록 유도합니다. 그러한 효과는 특히 서비스 제공자의 의무가 불명확하거나 광범위하게 정의된 경우 발생합니다.

확장적이고 모호한 조정의무와 오판 시 엄중한 법적 결과에 직면한 기업들은 불가피하게 과도한 검열을 실행합니다. 또한 플랫폼 규제 및 조정이 엄격해지면 이용자가 예술적 표현과 정치적 표현에 대한 부정적 결과를 피하고자 함에 따라 자기검열이 발생합니다. 자신이 사용하는 통신이 모니터링된다고 생각하는 사람은 자신의 표현 및 자신이 찾아보고 열람하는 콘텐츠를 모두 자기검열한다는 사실이 수많은 연구를 통해 밝혀졌습니다.[1]

2. 국제 사례 – 유럽 인터넷 법령 사례

EU 정부들은 EU 의 신규 인터넷 법안인 디지털 서비스법(DSA)에 대한 협의과정에서 동 문제에 직면하였습니다. DSA 는 강한 집행 체계를 통해 온라인 플랫폼의 명확한 책임을 규정하는 한편, 이용자의 기본권을 보호하고자 합니다. DSA 는 "착한 사마리아인(good Samaritan)" 콘텐츠를 장려하고 유형 및 규모에 근거한 실사 의무를 규정하며, 여기에는 콘텐츠 조정 실무, 알고리즘 큐레이션, 신고 및 조치 절차의 투명성과 관련한 의무가 포함됩니다.

예컨대, DSA 의 투명성 요건에 따르면 이용자는 플랫폼의 콘텐츠 조정 실무에 대한 사항을 통지받아야 합니다. 온라인 플랫폼의 서비스 약관에는 자동화된 의사결정 절차 활용에 대한 세부사항 및 인간에 의한 감독 범위가 포함되어야 합니다. 온라인상의 불법 콘텐츠에 대응하기 위하여 규모를 불문한 모든 호스팅 서비스 제공자는 잠재적인 불법성을 가진 콘텐츠를 신고할 수 있도록 하는 신고 및 조치 체계를 마련하여야 하며, 이후 플랫폼 사업자는 신고된 콘텐츠와 관련한 조치를 취할지 여부를 결정할 수 있습니다. 콘텐츠가 삭제되는 모든 경우에, 삭제 결정이 신고에 의한 것인지 또는 플랫폼의 자체 조사에 따른 것인지를 불문하고, 이용자는 해당 결정의 근거를 알 권리를 가지며 해당 결정에 이의를 제기할 수 있어야 합니다. DSA 는 콘텐츠 삭제 결정이 특정 대상을 목표로 하도록 하고, 콘텐츠 삭제 결정에 이용자의 표현의 자유 및 정보의 자유를 비롯하여 개인정보 보호 및 차별금지에 대한 권리가 고려되도록 몇 가지 보호조치를 마련하고 있습니다. 또한 이용자는 자신의 콘텐츠를 플랫폼이 착오로 삭제한 경우 해당 콘텐츠의 복구를 청구할 권리를 보유합니다.

DSA 협의에 앞서 몇 개의 나라에서 법이 발의되었고, 그 중에는 논란이 되었던 프랑스의 "아비아 법(Avia Bill)"이 있습니다. 아비아 법은 소셜미디어 정보매개자가 명백히 불법인 콘텐츠를 짧은 시간 내에 삭제하도록 의무를 부과하였으며, 이에 대하여는 전문가 및 시민단체의 비판이 있었습니다. EFF 는 공동으로 기획한 소송으로 프랑스 대법원이 24 시간 내에 권리침해 콘텐츠를 삭제하도록 한[2] 해당 법 요건이 플랫폼으로 하여금

[1] N. Richards, *Why Privacy Matters*, Oxford Press (2022)
[2] https://www.eff.org/press/releases/victory-french-high-court-rules-most-hate-speech-bill-would-undermine-free-표현.

완전히 합법적인 표현을 삭제하도록 유도할 수 있음을 인정하고 이를 폐기하도록 함으로써 법원에 의한 성공적인 개입을 이끌었습니다. 호주의 혐오표현법 등 다른 나라의 법에 관하여도 정당한 표현의 '과도한 차단(overblocking)'[3] 및 선행하는 EU 원칙과의 상충에 관한 문제의식이 형성되었습니다.

DSA 초안은 온라인 안전 문제와 불법 콘텐츠의 공유에 대응하기 위한 다양한 법적 해결책에 대한 지지를 표시하였습니다. 최초 발의안은 이용자 신고에 근거가 있는 경우 온라인 플랫폼으로 하여금 해당 이용자 콘텐츠를 삭제할 의무를 부과하여 그에 따라 이용자 콘텐츠에 대한 2 차 책임이 발생하도록 하였습니다. 다른 안은 콘텐츠 삭제를 위한 기한을 짧게 부여하기도 하였습니다. 결과적으로, 각 국가별 법안에서의 문제점들을 고려하고, 이용자의 표현의 자유에 대한 권리를 포함하여, 다양한 이해관계들 간의 공정한 균형을 달성하기 위하여, 최종 법안은 필터링 의무와 콘텐츠 삭제 기한을 짧게 하는 내용을 배제하였습니다.

법안의 최종안은 인터넷 규제에 대한 EU 의 과거 기본 법령인 전자상거래 지침(e-Commerce Directive)의 근거가 되는 원칙과 일치하는 것으로 널리 알려져 있습니다. DSA 는 국가가 플랫폼으로 하여금 이용자 통신을 모니터링하도록 강제하는 경우 이는 필연적으로 콘텐츠의 과다삭제를 초래하고, 언론의 자유를 저해하며, 이용자의 개인정보 보호권을 침해하게 될 것이므로 절대로 허용되어서는 된다는 중요 원칙을 확인하였습니다. 온라인 불법 콘텐츠의 확산과 허위정보 또는 기타 유해 콘텐츠의 전파로 발생하는 사회적 위험에 대응하기 위하여 EU 입법가들은 조화로운 신고 및 조치 체계를 마련하는 방법을 선택했습니다. 온라인 플랫폼 사업자는 이용자가 세부내용과 근거를 함께 첨부하여 콘텐츠를 신고할 수 있는 방법을 마련하여야 합니다. 플랫폼은 이와 같이 이용자 또는 기타 주체로부터 불법 콘텐츠의 존재를 신고 받는 경우 적절히 조치를 취할 의무를 집니다. 플랫폼 사업자는 그러한 콘텐츠의 제한 또는 삭제와 관련하여 적시에 결정을 내리고 삭제를 결정할 경우 이를 이용자에게 알려야 합니다. 성실한 호스팅 서비스 제공자라면 철저한 법적 조사를 수행하지 않고도 신고된 콘텐츠를 불법으로 식별할 수 있을 정도로 상세하고 정확한 정보가 신고 내용에 포함되어 있는 경우, 해당 사업자에 대하여는 DSA 책임 면제 조항이 적용되지 않아 해당 사업자가 아무런 조치를 취하지 않는 경우 책임이 면제되지 않습니다.

이는 표현에 대한 책임은 이용자가 온라인으로 게시 또는 공유하는 콘텐츠를 호스팅하는 플랫폼이 아니라 표현자가 계속해서 진다는 EU 온라인 플랫폼 규칙을 따른 것입니다. 그러나 온라인 플랫폼 사업자는 불법 콘텐츠의 온라인 전파를 저지하기 위한 절차를 마련하여야 합니다.

이러한 기준에 부합하지 않는 영국 온라인안전법 초안 등 최근의 여러 나라의 법은 이용자의 기본권을 침해하는 것으로 널리 비판받고[4] 있습니다.

---

[3] https://www.euractiv.com/section/data-protection/news/austrias-law-against-online-hate-speech-question-marks-in-the-home-stretch/.

[4] https://www.eff.org/deeplinks/2022/08/uks-online-safety-bill-attacks-free-speech-and-encryption.

3. 인권 원칙: 견제와 균형의 필요성

표현의 자유와 온라인 개인정보 보호에 관한 권리는 EU 기본권 헌장(EU Fundamental Rights Charter), 유럽 인권 헌장(European Human Rights Charter)에 따른 기본권이며, 또한 기타 국제 인권법 제도에 따라 보호됩니다. 그러한 권리를 온라인 상의 유해 콘텐츠 회피 등 다른 공적 가치의 보호를 위하여 제한하고자 하는 국가적 조치는 해당 목표들 간의 공정한 균형을 추구하여야 합니다.[5] 일반 원칙으로, 국가적 조치는 명확한 법적 근거가 수반되어야 하며 민주사회에서 필요한 것이어야 하며 목표 달성을 위한 필요한 범위를 초과하지 않는 균형성을 담보하여야 합니다. 또한 그러한 조치는 충분히 보호되어야 합니다.

문제가 되고 있는 한국 전기통신사업법 관련 조항과 관련하여, 표현의 자유에 대한 권리를 포함하는 기본권 측면이 충분히 고려되지 않았다는 정당한 우려가 존재한다고 본 재단은 이해하고 있습니다. 해당 법안은 특정 콘텐츠가 특정 기관에 의하여 불법인 것으로 검토 및 결정되는 콘텐츠와 일치하는지 모니터링하고 제한할 적극적 의무를 규정합니다.

충분한 견제와 균형이 없다면 그러한 의무는 모든 이용자 콘텐츠에 대한 무분별한 일반 모니터링 및 관련 콘텐츠의 업로드를 자동으로 방지하는 자동 필터링 시스템 등 오류에 취약한 매칭(matching) 기술의 이용으로 이어질 가능성이 높습니다. 인권주의에 따라, 차단조치를 규정 또는 요구하는 법적 체계는 해당 조치가 엄격히 불법 콘텐츠를 목표로 하며 어떠한 자의성이나 과도한 효과를 가지지 않도록 하여야 합니다.[6]

4. 오류 가능성

콘텐츠를 차단하기 위한 기술적 조치 또한 오류에 취약한데, 그 이유는 어떠한 알고리즘도 합법적 이용을 불법적 이용과 구별하기 위하여 필요한 이 같은 맥락적 분석과 법적 분석을 대체할 수 없기 때문입니다. 이는 EU 저작권 지침(EU Copyright Directive)의 교훈 중 하나입니다. EU 입법가들은 권리 침해를 차단하는 한편 합법적 이용을 허용할 수 있는 필터 사용을 의무화하려고 하였습니다. 필터 제공업자를 포함하여, 필터 조치 이행과정에 관여한 모든 이들은 이는 불가능하다고 증언하였습니다. 이는 플랫폼이 과도차단이나 과소차단을 해야만 함을 의미합니다.

또한 필터는 순수하게 기술적 오류를 발생시켜, 어떠한 자료를 보호대상 작업의 복제물로 오인할 수 있습니다. YouTube 의 ContentID 와 같이 가장 고비용으로 개발된 필터마저도 클래식 음악 기록물을 잘못 매칭시키는 등의 문제가 있습니다. 또한 기술적 분석 결과 다른 필터에서도 매칭 문제가 발견됩니다.

필터가 의도된 대로 작동한다고 하여도, 통상적으로 권리자가 적법한 콘텐츠를 권리 침해 콘텐츠로 잘못 주장하는 등의 인적 오류로 인하여 그러한 필터가 합법적 콘텐츠를

[5] ECHR, Perinçek v. Switzerland [GC], § 274; Case C-275/06 Promusicae [2008] ECR I-271.
[6] Cf. ECtHR, Application no. 10795/14 – *Kharitonov v Russia* (website blocking), at 45.

삭제하는 경우도 매우 잦습니다(저녁 뉴스 방송의 영화 소개에 사용된 영화 클립만이 아니라 뉴스 방송 전체를 신고하는 등) YouTube 의 ContentID 와 관련한 많은 오류가 이 범주에 해당합니다.

B. 일반 모니터링, 필터링, 그리고 개인정보 보호

또한 일반 모니터링은 기업들로 하여금, 많은 경우 이용자에게 알리지 않은 채, 이용자에 대한 대량의 데이터를 수집하도록 강제함으로써 이용자의 개인정보 보호권을 제한합니다. 여러 국가에서의 경험에 따르면 온라인 플랫폼이 이용자 콘텐츠를 체계적으로 필터링하도록 강제한 법률은 심각한 개인정보 보호 및 보안 위험을 발생시킵니다. 예컨대, 저작권 분야에서 시행이 강제된 기술적 조치는 데이터 유출 위험 대응에 부적절한 것으로 나타났으며, 최악의 경우 플랫폼 사업자가 메시지 내용을 파악하기 위하여 암호화를 풀도록 하기도 하였습니다.[7]

그러한 행위는 인권기준을 위반합니다. UN 인권고등판무관은 어떠한 간섭행위도 사건별로 독립적인 사법기관에 의하여 승인된 경우에 한하여 수행되어야 한다고 강조하였습니다.[8]

끝으로, EU 의 일반데이터보호규칙(GDPR)은 필터 시스템은 이용자의 콘텐츠를 자동으로 삭제하거나 이용자 데이터를 프로파일링 목적으로 수집할 수 있어 이용자에게 악영향을 줄 수 있음을 인정하였습니다.[9] 따라서 데이터 보호원칙은 어떠한 콘텐츠가 검토, 차단 또는 삭제되기 전에 이용자의 개인정보 보호, 언론의 자유 및 기타 기본권에 대한 보호조치가 확보되도록 국가적으로 조치를 취할 것을 요구하고 있습니다.

C. 일반 모니터링, 필터링 그리고 경쟁

모니터링은 대규모 자본 투자를 요하고, 이는 결국 신생 기업의 해당 분야 진입을 저해합니다. 요구되는 인적자원 외에도, 업로드 필터 등 기술적 조치의 구축 및 인허가에는 높은 비용이 소요됩니다. YouTube 의 ContentID 로 인하여 Google 은 이미 5 년 전 1 억 달러 이상의 비용을 지출하였습니다. EU 입법가들에게 저렴한 해결책으로 홍보된 Audible Magic 의 이용에는 해당 기업이 진술하는 것보다 실제로 더 많은 비용이 소요된다는 것이 널리 보도되었습니다.

또한 필터는 기술적 락인효과(technical lock-in)를 발생시킴으로써 경쟁을 감소시킵니다. Engstrom 및 Feamster 보고서에 따르면 특정 필터링 기술 채택의 주요한 결과로 기업들이 해당 기술을 중심으로 다른 시스템을 설계하게 된다는 점이 있습니다. 그 결과로 설계

[7] https://www.eff.org/deeplinks/2022/11/filter-mandate-bill-privacy-and-security-mess.
[8] https://www.ohchr.org/en/press-releases/2022/09/spyware-and-surveillance-threats-privacy-and-human-rights-growing-un-report.
[9] Article 22 GDPR.

선택, 투자, 또는 신규 사업군이 배제되거나 불가능할 정도로 많은 비용이 필요해질 수 있습니다. 이 경우 기업들이 자발적으로 그러한 선택을 하는 것이 아니라 정부 규제에 의하여 강제로 그러한 선택을 하여야 하는 상황으로 내몰린 것이라는 점을 주목하여야 합니다.

D. 투명성 및 시정을 위한 체계

인권에 대한 악영향은 콘텐츠 조정을 위한 자발적 인권 체계를 통하여 경감될 수 있습니다. 어떤 콘텐츠가 온라인으로 공유되어야 하는지 여부에 대한 결정은 중대한 인권적 함의를 가지며 온라인 플랫폼은 그러한 결정을 함에 있어 사업자가 더욱 책임감 있게 행위하도록 돕기 위한 가이드라인을 활용할 수 있습니다. EFF 는 이와 같은 가이드라인을 제공하기 위해 오랫동안 노력하였습니다. 2015 년, EFF 는 국제연합의 일원으로서 국제인권문서와 기타 국제 법률체계에 근거한 기본 보호조치 및 모범관행 체계인 "인터넷 책임에 관한 마닐라원칙(Manila Principles on Internet Liability[10])"을 공포하는 데에 일조하였습니다. 2018 년, EFF 와 파트너들은 "콘텐츠 조정의 투명성 및 책임에 관한 산타클라라 원칙(Santa Clara principles[11] on Transparency and Accountability in Content Moderation)"을 발표해 정보매개자가 자발적으로 더 나은 실무관행을 채택하도록 촉구하였습니다. 2021 년에는 다양한 공동체와 다양한 시장에 대한 플랫폼의 정당한 절차 및 투명성 관행의 기본적 불평등에 적절히 대응하는 것에 초점을 맞춘 해당 원칙의 신규 버전이 개발되었습니다.

그러나 EFF 는 플랫폼이 불법 콘텐츠를 호스팅하고 증폭시키는 매우 현실적 문제에 대응하는 동시에 해당 플랫폼이 보호대상 이용자 표현을 제거할 유인을 얻지 못하도록 해당 플랫폼에 충분한 보호를 제공하여 표현의 자유를 진흥하도록 하는 것 사이에서 절충점을 찾을 필요가 있음을 인식합니다. 이를 달성할 최선의 방법에 관한 본 재단의 권고사항은 표현의 자유보다 플랫폼 상의 처리 규제에 중점을 두는 것과 의무적 콘텐츠 제한은 어떠한 경우에도 사법당국이 명령에 의하여야 하고, 침입적인 필터 시스템에 의존함이 없이 적용되도록 하는 것이 당사자 모두의 최선의 이익에 부합한다는 점에 근거를 두고 있습니다.[12]

E. 결론

정보매개자는 대중이 전례없는 규모의 청중과 콘텐츠를 공유할 수 있도록 하는, 인터넷 아키텍쳐의 필수적 요소이자 표현의 자유의 근본적 동력입니다. 인터넷 법령과 관련한 여러 나라에서의 경험에 비추어 보면 규제당국과 입법당국은 온라인 불법 콘텐츠 대응을

---

[10] https://manilaprinciples.org/index.html.
[11] https://santaclaraprinciples.org/.
[12] https://www.eff.org/deeplinks/2022/05/platform-liability-trends-around-globe-conclusions-and-recommendations-moving.

위한 적절한 도구 선택에 어려움을 겪고 있습니다. EFF 는 인권에 부합하는 조정체계 채택이 그러한 난관을 극복하는 데에 큰 도움이 될 수 있을 것이라고 생각합니다.