platform accountability in digital age for what and how?

Kyung Sin ("KS") Park
Open Net/ Korea University
Director/Professor
kyungsinpark@korea.ac.kr

Liability exemptions in general

- Samsung Heavy Industries – Hebei Spirit Oil Spill (2008) → Shipowners' liability limitation
- Warsaw Convention on airline luggage liability limitation
- Exempts liability for negligence but only for reckless disregard of danger - → Why? Not for ordinary drivers, landowners, doctors → knowledge v privacy? The number of ppl using it? Or common carrier doctrine (right to refuse services?)
- How about digital intermediaries - e.g., using mobile communication to conspire on a bank robbery → probably no telecom liability?
- How about internet companies? How different are they from telecoms?

Civilizational significance of the Internet

- Giving powerless people the tools of mass communication
- What is mass communication? TV and newspaper (reaching multitude)
- TV weak to government, newspapers weak to corporations, also LIMITED SPACE → features usually elite (vs people)
- Formal democracy vs. **substantive democracy** (requires equality in communication
- Does internet give us that?

2012 Korean Constitutional Court on Internet real name law : "overcome hierarchy offline in age, gender, social status"

2011 Korean Constitutional Court on election restriction: "online communication requires AFFIRMATIVE conduct of receiver, so not easily affected by financial power of candidates (publishers)"

Crux of internet's civilizational significance:

Inclusivity to mass communication – everyone can produce massively available contents without approval from middlemen (legacy media outlets) who must triage for moderation and limited space → **Reflects true popular opinion**

Massive contents -→ search engines – everyone can be on massive information search → **egalitarian access to knowledge**

Server-client model (90% of web traffic) → why pageviews, likes, followers are important → **fully consensual communication**

Intermediary Liability Safe Harbor?

- Unapproved power of uploading -→ ability for bad users to abuse the power → **Internet is bound to transmit bad contents** : How to keep up the civilizational value of internet while reducing bad contents
- Intermediaries: ISPs/ Social media platforms/ Web hosts/ Search engines → **When should intermediaries be held liable** for "aiding and abetting" online illegal content?
- SAFE HARBOR: **No liability as long as not aware of illegal contents**, why? → if not, GENERAL MONITORING or Prior Censorship → Internet becomes like TV and newspaper subject to gate keeping → People lose the power of speaking to one another without approval (Crux of Internet's success)

World's response: best practices

- EU e-Commerce Directive Article 13-15 – **"information society services"**
- Japan Provider Liability Act, Article 3 (1)-(2)
- Indian IT Act and its rules
- US Digital Millennium Copyright Act section 512 (difference with EU & Japan: You MUST DO SOMETHING! – NOTICE AND TAKEDOWN BUT ONLY IF YOU WANT TO
  - Cf. US CDA 230

→Liability-Exempting, Not liability-imposing: "shall NOT be liable if not aware of illegal contents. . ."

→Expectation: notice and takedown

→not liable if you do X (DMCA) or if not aware (the rest). Not doing X or awareness does not mean liability but just falls back on ordinary torts → Incentive: Companies' bright line rule of exemption for unknown contents

Not the best practice

- Broad immunity: Communication Decency Act Section 230

- Exemption even for liability for contents that platforms are given full awareness of.

- **Intermediary Liability Safe Harbor** - EU, Japan
- Liability-imposing regime – Many Asian countries
  - Thailand's Computer Crimes Act 2007 (CCA 2007) Article 14-15 criminal sanctions imposed, inter alia, for allowing publication of information on public computers in circumstances where the disseminated information is false and is likely to cause damage to a third party or the country's national security
  - China – liability for failing to monitor, remove, sometimes even without notice
  - Korea – liability for failure to remove upon notice **(mandatory notice and takedown! Why is it bad?)**

Intermediary liability safe harbor as international standard

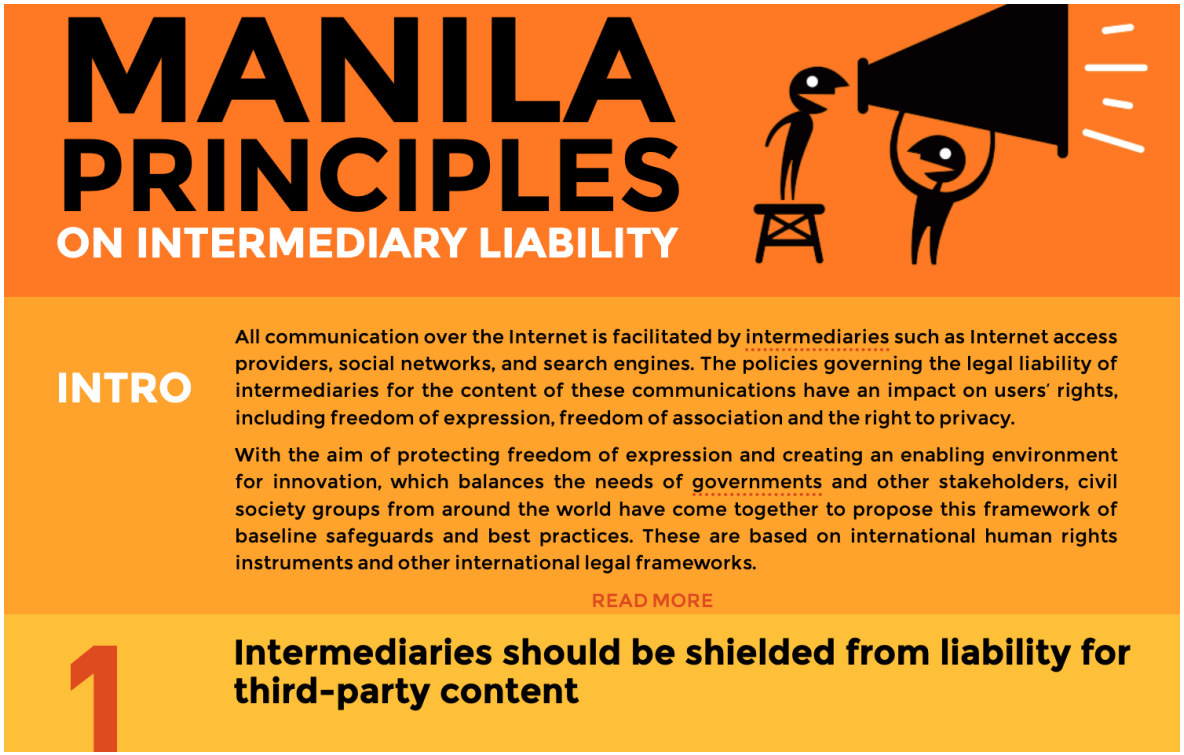
Consideration should be given **insulating intermediaries from liability for content** produced by others where liability should only be incurred if the intermediary has specifically intervened in the content, which is published online or fails to take down content following a court order (contrary to the practice of notice and takedown).

2011 Joint Declaration of UN, OAS, OSCE, and ACHPR on Freedom of Expression and the Internet, June 2011

[N]o one should be held liable for content on the internet of which they are not the author. Indeed, no State should use or force intermediaries to undertake censorship on its behalf.

Report of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 16 May 2011, A/HRC/17/27, para. 43.

International Soft Law



Ethical questions and safe harbor's canonical answers

- "Platforms should be held liable for illegal contents on their services. They make money off of it. Why not?"

→ Telecoms, airlines?

- "Platforms should monitor their services for illegal contents. They are doing it anyway. Why not make it mandatory?"

→ Incentives?

- "Platforms should be held liable for illegal contents someone asked to take down. No incentive to shut down space. Why not?"

→ Who is that someone? Or platforms as adjudicators?

Not All Quiet on Western Front:
CJEU case law on the scope of Article 14 ECD

Cases relate to the possibility of different kinds of online platforms to invoke the hosting safe harbor;

- Google France (C-236/08) – selling of search keywords – no liability
- L'Oréal/eBay (C-324/09)- liability if generally aware
- Uber (C-434/15): no information society service (qualifies as transport service);
- Airbnb (C-390/18): information society service;

Recital 42 of ECD – Does this really apply to hosting services;

> ECD recital (42): . . .***this activity is of a mere technical, automatic and passive nature, which implies that the information society service provider has neither knowledge of nor control over the information which is transmitted or stored.***

Q: Arent' platforms actively involved in curating and monetizing content and user activity?

Moving beyond reactive responsibilities? CJEU case law on injunctions and general monitoring ban

Injunctions: ECD allowed for possible injunctions, including prohibitory, and national level duties of care (recital 48 ECD);

CJEU in Scarlet/Sabam and Netlog case: no general obligation can be imposed to monitor all data (with heavy reliance on fundamental rights of users);

CJEU in Eva Glawischnig-Piesczek: possibility to impose (**stay down**) injunction to remove "equivalent content" after a judicial finding of illegality (NB. No explicit filtering obligation);

CJEU in Article 17 CDSM case (C-401/19): filters need to be able to distinghuish between lawful and unlawful content due to their interference with freedom of expression; Injunctions need to respect **fair balance** between fundamental rights;

General monitoring ban remains and stands in the way of (general) proactive duties of care;

Sabam/Scarlet

Preventive monitoring of this kind would thus require active observation of all electronic communications conducted on the network of the ISP concerned and, consequently, **would encompass all information to be transmitted and all customers using that network.**

Eva Glawischnig-Piesczek

[..] **an obligation** such as the one described […] above, on the one hand — **in so far as it also extends to information with equivalent content** — appears to be sufficiently effective for ensuring that the person targeted by the defamatory statements is protected. On the other hand, that protection is not provided by means of an excessive obligation being imposed on the host provider, 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**.

Something happens in 2016

- Buzzfeed 1: : fake news gone viral (e.g., Pope Endorses Trump) past real news.
- Buzzfeed 2: 40% of Trump voters believing in Democrats' child sex slave ring." 36% believing in Kenyan birth of Obama

➔ **"Fake news is affecting world history!!!"**

- "Fake News = NOT just false news but false news from FAKE SITES. Fake news is the offsprings of digitalization.
- "'METANESS' about fake news: once believed to have been picked up by reputable media, goes viral again not because people believe the story but people believe the fact of coverage by reputable media. → that alone does the magic e.g., casting a cloud of doubt on Hilary's candidacy

2016 Trump election → "fake news" debate

- German Network Enforcement Act (NetzDG) 2017
- Australia's 2019 "abhorrent violent content" law
- France 2020 Avia law

→**"Mandatory" notice and takedown law**

- Liability-imposing: ". . shall be liable if fails to take down within 7 days/24 hours/1 hr"
- Technically ok under SAFE HARBOR because applies only to

NOTICED and KNOWN ILLEGALITY but. . .

→Platform's tendencies to err on the side of **taking down** vs. **keeping it up**

→MANY lawful postings taken down

My answer: Are fake news really a problem?

1. Do we know whether the stories were believed by people who shared them on Facebook? Maybe, **fake news were shared just for fun not because the substance were believed**. Look at Fake sites like WorldPoliticus.com, ABCNews.com.co. NOT distinguishable from supermarket tabloids (*The National Enquirer*, *Star*, *The Globe*, *National Examiner*) *or "red tops" – (e.g. Alien Endorses Trump).* Will we regulate tabloids as well?

2. Harmful controversies believed by people (i.e. Obama's Kenyan birth) ARE NOT FAKE NEWS shared through social media but **false information shared by POLITICIANS.**

Deep dive on Germany's NetzDG (2017)

- Over 2 million registered users
- Take down illegal contents defined by Germany's Criminal Code.
- flagged by individuals.
- "manifestly unlawful" : within 24 hours, all other "unlawful content", within 7 days.
- Or face fines of up to 50 million euros

→**On face, no violation of international standard on safe harbor but exploits the grey area by requiring "noticed" contents to be taken down --**"over-implementation" by providing an incentive to err on the side of caution rather than free expression

- "privatizing" online censorship because of the scalability issue. No public control but reliance on platforms' decisions

Mandatory notice and takedown, we had it all along in Korea and Asia!

- Korean Copyright Act – attempt to copy DMCA 512 but break into 2 sentences – “not liable if take down” + “must take down if noticed”
- Network Act - ”must take down if noticed” (2007)
  - -Missing ”liability-exempting” language and only “liability imposing”
- Problems of over-blocking – many lawful contents taken down.
- Other Asian adaptations of mandatory notice and takedown:
  - Malaysian Copyright Act (2012)
  - Indonesian commerce platform circular (2016)

But remember the best practices(liability-exempting)!

  - Japanese Provider Law (2001)
  - Indian IT Act (2011)

Spread of NetzDG : Mal-adapted into Administrative Censorship

- 2019 Philippines Anti False Content Act – mentions NetzDG
- 2018 Malaysia Anti Fake News Act – mentions NetzDG
- 2018 Vietnam Cybersecurity law – “propaganda against Vietman”, etc., - similar to NetzDG, e.g., 24 hours
- 2019 Singapore Protection from Online Falsehoods and Manipulation Act – false statement of facts - mentions NetzDG
- 2021 Indonesia MR5 – “prohibited content”, e.g., 4 hours, 24 hours of **flagging by Ministry** - similar to NetzDG
- 2022 Myanmar Cybersecurity Bill – must remove ”timely” all prohibited content **after flagging by the department** including “complained of stmts damaging another’s social standing and livelihood”

Push-back: France Avia Law struck down– Constitutional Conseil (June 2020)

- Part I: Terrorist content, child pornography AS **notified by administrative authority** – within 1 hr
- Part II: Other "manifestly illegal" content – within 24 hours of notice
- Unconstitutional b/c time too short, extrajudicial → criminal penalty not proportionate
- Probably considered impact on free speech – "false positives"
- Conseil national du numerique (French Digital Commission), la Commission nationale consultative des droits de l'homme (French Human Rights Commission) opposed.
- Notice that Part I is not even a liability law but **direct administrative censorship.**
- 2022 October **Spanish Supreme Court** on blocking order on womenonweb.org – administrative censorship unconstitutional!
- 2009 June **French Supreme Court** on HADOPI law – **administrative cut-off** of internet access unconstitutional
- 2014 **Philippines Supreme Court** on Cybercrime Law– administrative censorship is like "search and seizure". Unconstitutional without warrant!
- Of course, then you already had **Bantam Books v Sullivan (US Sup Ct** 1963)

DSA: Overview – Europeanizing DMCA

- Intermediary services (incl. conduit/cashing) → hosting services → online platforms → VLOP (45M users)/VLOSE
- Liability exemption: same as e-Commerce directive → No basis for liability but only exemption from it → Change: **MUST DO SOMETHING TO GET EXEMPTION like DMCA. No more voluntarism under ECD**
- Good Samaritan rule more clearly established (no more passivity requirement)
- Platform accountability for bad content
  - Judicial/administrative order on specific content takedown allowed? Not prohibited
  - Hosting services – notice and action – **illegality** must be clear -- Non-arbitrary and objective action – Cf. DMCA 512 – "trusted flaggers" prioritized
  - Duty to report terrorism/trafficking
  - Deplatform bad users
- Due diligence to "users": terms of restriction clear – point of contact - "dark patterns" - advertising transparency – appeal mechanism

A tiered approach in the design of **due diligence** obligations

Most basic obligations: all intermediary services; **single point of contact, or EU legal representative**

Additional due diligence requirements: hosting services; **notice and takedown mechanism; inform authors**

Additional due diligence requirements: consumer-facing hosting services that make things public (‘online platform’ definition); **internal complaint-handling mechanism; non-binding out-of-court dispute settlement; “trusted flaggers”; habitual publishers; dark patterns; ads labeling; no targeted advertising on minors; info on recommender systems; Online Market places(Know Your Business Customer)**

Additional due diligence requirements (asymmetric): largest online platforms and search engines, including risk-management, auditing and crisis protocol.

DSA evaluated: Good for Asia?

- Safe harbor preserved
- Non-mandatory Notice and Takedowns - notice-and-takedown mandated as a process, not as an action - no content-by-content liability – avoid false positives
- Greater transparency on notice and takedown process
- Does not establish administrative censorship but explicitly condones such power by MS → risk of pro-incumbent bias and its harm on democracy

Full Disclosure: First World Moving backward?

- In the meantime. . .
  - 2021 Australia Online Safety Act - eSafety Commissioner's removal notice to be acted upon
  - 2021 Canada Online Harms bill – upload filtering, 24 hours mandatory (private) notice and takedown.
  - 2021 UK Online Safety bill – **duty to detect and remove** and Ofcom able to penalize if not fulfilled → general monitoring obligation

So how to make internet safer?

- Theory: Social media companies need Information Integrity because they need more eyeballs and more peer created contents
- Problem: Some bad contents generate more eyeballs and peer contents than good contents → Conflict of Interest
- Solution: More mandatory regulation? Empowering or Democratizing? Platform Diversity?
- Metaphor: Is internet a stadium or a rhizome (à la **Lacan**) or an aggregate of rooms? People make 2 choices for communication to be consummated. Choice of Platform and choice to press the link. Defamation happens in disrespect of the person targeted between the speaker and the listener but isn't the listener also already pre-dispositioned to listen?

FOSTA (Fight Online Sex-Trafficking Act)

- CDA 230: no liability even for known contents.
- Exception from CDA 230: “knowingly assist[], support[], or facilitate[]” activity violating federal sex trafficking law
- Difference b/w trafficking and voluntary prostitution→ sex workers at increased risk b/c no longer distance advertising → balloon effect
- Good Samaritan aspect of CDA 230 excepted also: Moderation does not impose knowledge on you
- **“knowing” assistance** = **knowing illegality or knowing existence of contents?**

Korea: Anti-”Nth Room” Law

- Nth Room: Forced sexual activities, not rape, using threats of disclosing nude photos – broadcast over various chatrooms on Telegram for fee varying depending on which “N”th room
- Law requires platforms to take “filtering” to discover and prevent circulation of illegally filmed material
- Illegally filmed material – nonconsensual sexual video, sexual “deep fakes”, ”child pornography(CSAM)”,
- All videos uploaded in major apps/webs in Korea subject to filtering to compare against administratively pre-curated database of ‘illegally filmed material’
- General monitoring obligation? ”Function creep”?
- Copyright Act Article 104: ‘technical measure’ to filter out copyright infringing material
- Telecommunication Business Act Article 22-3 (1) ‘technical measure to filter out obscene material