

An aerial photograph of Zurich, Switzerland, showing the city's layout with the Limmat river, various buildings, and green spaces. The ETH Zurich campus is prominently featured in the center, with its large domed building and surrounding structures. The text is overlaid on a blue rectangular area on the left side of the image.

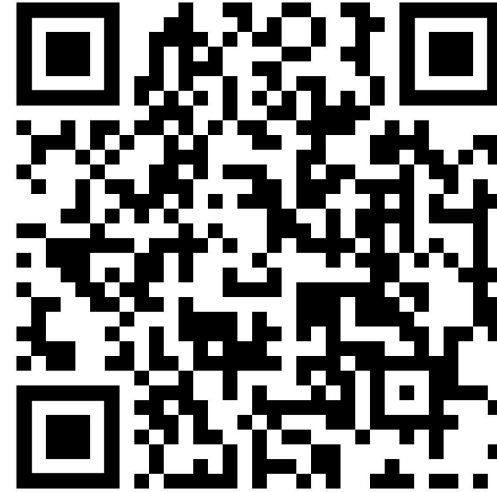
Moderating Digital Platforms *Week 2: Definitions, History, and Case Study*

Raquel De Haro & Luka Nenadic

23/09/2025, Zurich

Agenda

1. Admin
2. Mentimeter Quiz
3. Defining Content Moderation
4. History of Content Moderation
5. Interactive Case Study



GitHub Repository

Admin

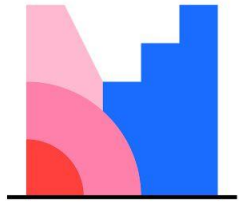
We have decided to make ***all readings voluntary***.

Assignments:

- **By 10.10.25:** Create your groups & fill your names [here](#).
- **By 25.10.25:** Upload a brief outline (max. 1 page) explaining your chosen platform
- **By 19.12.25:** Upload the full content moderation policy & the group video presentation
- **By 16.12.25 (Optional, max +0.25 of the final grade):** Upload 2 news relating to the content of the course and comment briefly on the news uploaded by, at least, 2 other colleagues. Do so by opening a new discussion on Moodle.

Submissions must be received before midnight (23:59) on the stated date. Assignments submitted after 00:00 will not be reviewed.

Now It Is Your Turn!



Mentimeter

Defining Content Moderation

Moderation as any form of *curating, filtering, or deleting specific content*

Gillespie, 2018¹: **platforms** are online sites and services that

- 1) host, organize, and circulate users' shared content or social interactions
- 2) without having produced or commissioned (the bulk of) that content,
- 3) built on an infrastructure, beneath that circulation of information, for processing data for customer service, advertising, and profit.
- 4) *Platforms do, and must, moderate the content and activity of users, using some logistics of detection, review, and enforcement.*

1. Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*, New Haven: Yale University Press, 2018. <https://doi.org/10.12987/9780300235029>

History of Content Moderation: Before the the Internet

- Content moderation predates the internet:
 - Roman forum
 - State censorship
 - Editors and publishers
 - Academic journals
- Dualist model of speech/content moderation: governments and speakers (Balkin, 2018)²
- Most importantly: regulators and legacy media (newspaper, radio, television)

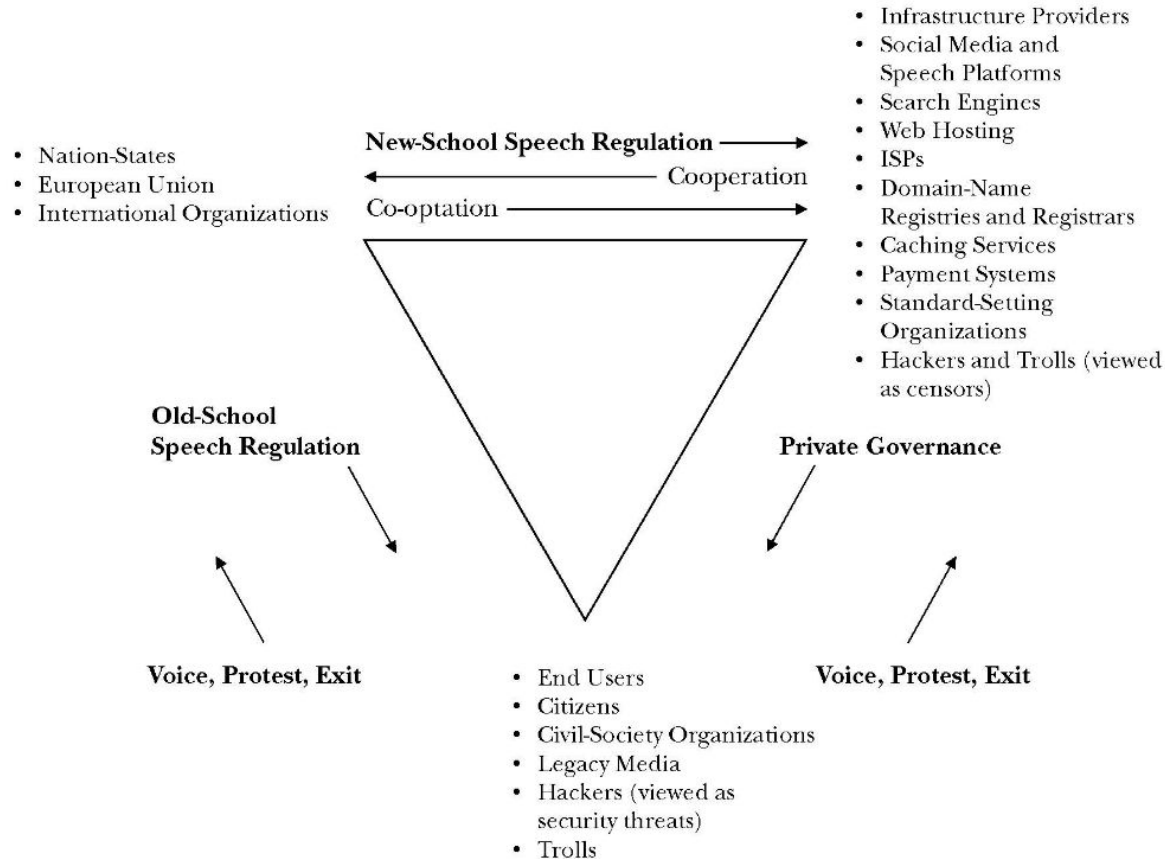
2. Balkin, Jack M. "Free Speech is a Triangle." *Columbia Law Review* 118, no. 7 (2018): 2011-2056.
<https://teaching.globalfreedomofexpression.columbia.edu/resources/free-speech-triangle>

History of Content Moderation: Early Internet Era

- “Utopist” early online communities with relatively few users self-regulating
- Shift to individual platform/forum moderators (akin to Reddit moderators today)
- Need for clearer rules and automated enforcement as platforms grow larger
- This leads to the “Triangle of Speech:”³ states, users, *and third-parties*

3. Balkin, Jack M. “Free Speech is a Triangle.” *Columbia Law Review* 118, no. 7 (2018): 2011-2056.
<https://teaching.globalfreedomofexpression.columbia.edu/resources/free-speech-triangle>

History of Content Moderation: Pluralist Model (Balkin, 2018)



From Legacy Media to Social Media: New Model, Same Old Problems?

- 1972: Vietnam war photo (The Terror of War) by Nick Ut wins Pulitzer Prize
- Historical Impact: Image helped turn American opinion against Vietnam War
- 2016: Facebook censored photo on Norwegian newspaper's page
- Protests: Norwegian PM and ministers reposted image in defiance
- Facebook Reversal: Platform reinstated the photo after the backlash

Source: Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*, New Haven: Yale University Press, 2018. <https://doi.org/10.12987/9780300235029> (pp. 1-23)

History of Content Moderation: Early Platform Regulation in the US

- Background: Legal uncertainty of platforms regarding content moderation
- Core provision: Section 230 of the Communications Decency Act (1996)
- Broad legal immunity to platforms for third-party user content (exceptions exist)
- Simultaneous protection of content moderation decisions conducted in good faith
- Rationale: Promote free speech, while shielding good faith moderation
- Impact: Foundation of the internet as we know it today

When content moderation discussions went global...

- CompuServe Case: German executive Felix Somm faced charges for CSAM (1998).
- Bonn Ministerial Declaration on Global Information Networks (1997): EU-15 MS and EFTA countries + guest countries (USA, Canada, Japan, Russia).

21. Ministers agree that any regulatory framework for electronic commerce should be clear and predictable, pro-competitive, strike the right balance between the freedom of expression and the protection of private and public interests, in particular the protection of minors, and ensure consumer protection.

- This is how liability exemptions for providers were born both in the US and in the EU.
 - Another exemption was introduced in the US Digital Millennium Copyright Act of 1998 for copyright infringements.

Content Moderation History in the EU:

- The US Digital Millennium Copyright Act (DMCA) became the model for the EU's e-Commerce Directive (2000).



- The liability exemptions of the e-Commerce Directive are now part of the Digital Services Act (DSA) (chapter 2).
- The DSA builds on that incorporating due diligence obligations (chapter 3).

The DSA redistribution of power (Husovec, 2024):

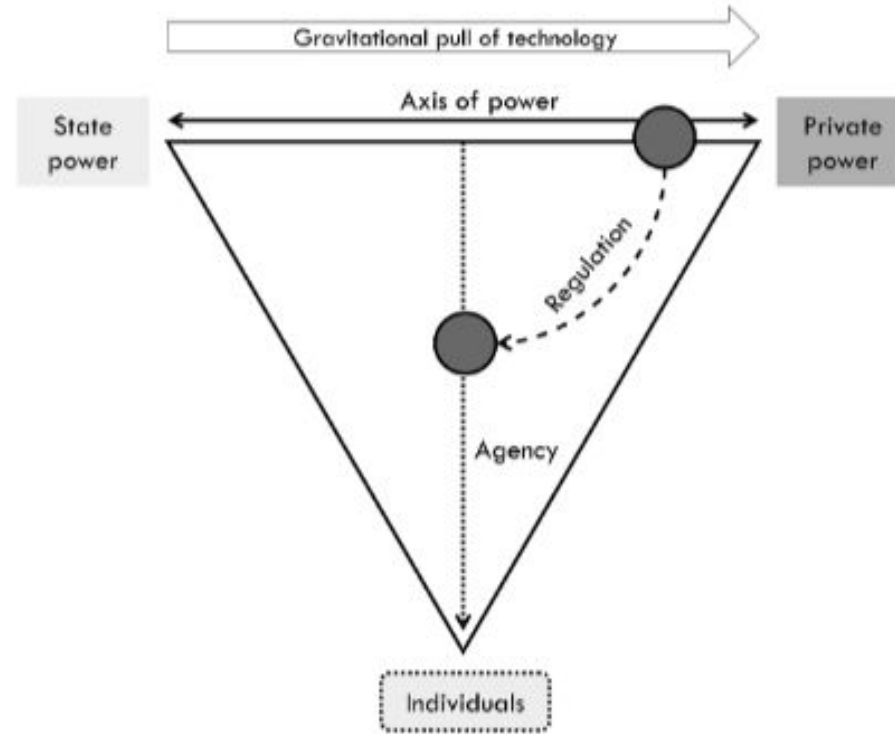


Figure 1.1 Redistribution of power

Why do definitions matter?

- Liability involves civil, criminal and administrative liability.
- Key to remember → **liability exemptions protect providers from liability for User-Generated Content (UGC)**.
 - Hence, editorial content is NOT protected from liability.
- Liability only emerges under certain conditions:
 - Knowledge of the illegality (exceptions for good faith of providers). Knowledge needs to be specific and NOT require a detailed legal examination (i.e. CSAM).
 - Neutrality (neutrality requirements can also be different for each technical service). In summary, neutrality of a provider can be lost if:
 1. Providers co-determine user's content.
 2. Providers present user's content as the provider's own.
 3. Providers deliberately instruct the uploader's wrongs.
 4. Providers control or jointly collaborate with the users.

Source: Husovec, M. (2024). *Principles of the Digital Services Act*. Oxford University Press.



Skip Ad



Viewership

Subscribe

- Key: Do NOT change the meaning of the content created by users.



Why do definitions matter?

- Interplay of liability vs. accountability in the DSA:
 - Chapter 2: Who can be liable, and when is the provider acting unlawfully?
 - Chapter 3: When can the provider be held responsible for not designing their services and procedures according to a certain diligent standard? → does NOT require neutrality.

Technical service (Chapter 1)	Liability exemption (Chapter 2)	Due diligence (Chapter 3)
Mere conduit	Regardless of knowledge	Light touch
Caching	Knowledge + removal at source	Light touch
Hosting	Dependent on knowledge	Submission & Explanation NTD
Online platform (hosting + public)	Dependent on knowledge	Core focus
Search engines	[Open]	<i>VLOSEs only</i>

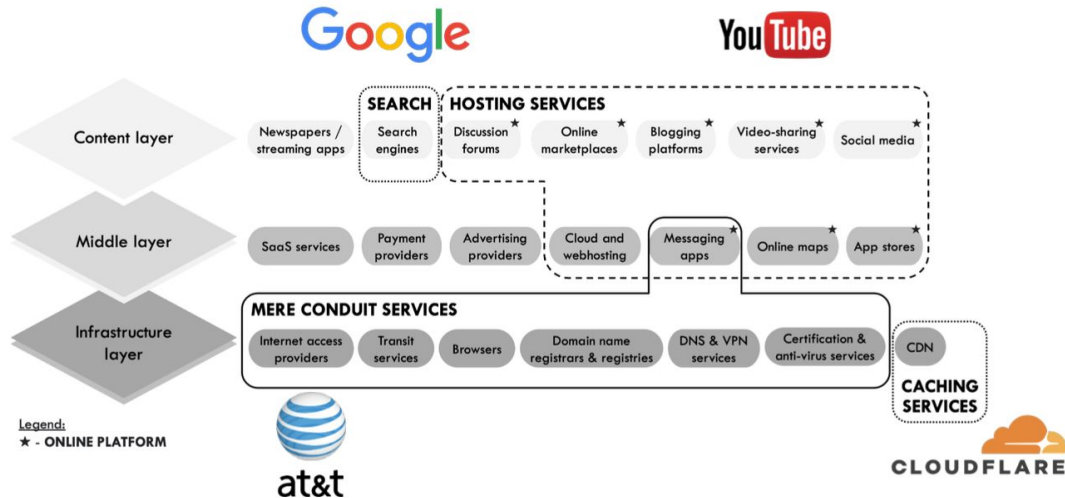
Source: Husovec, M. (2024). *Principles of the Digital Services Act*. Oxford University Press.

Chart: Husovec, M. (2025). Digital Services Act Specialist Masterclass (Module 2).

Note: the chart is ONLY for illustrative purposes. You are not expected to learn it for the exam.

Why do definitions matter?

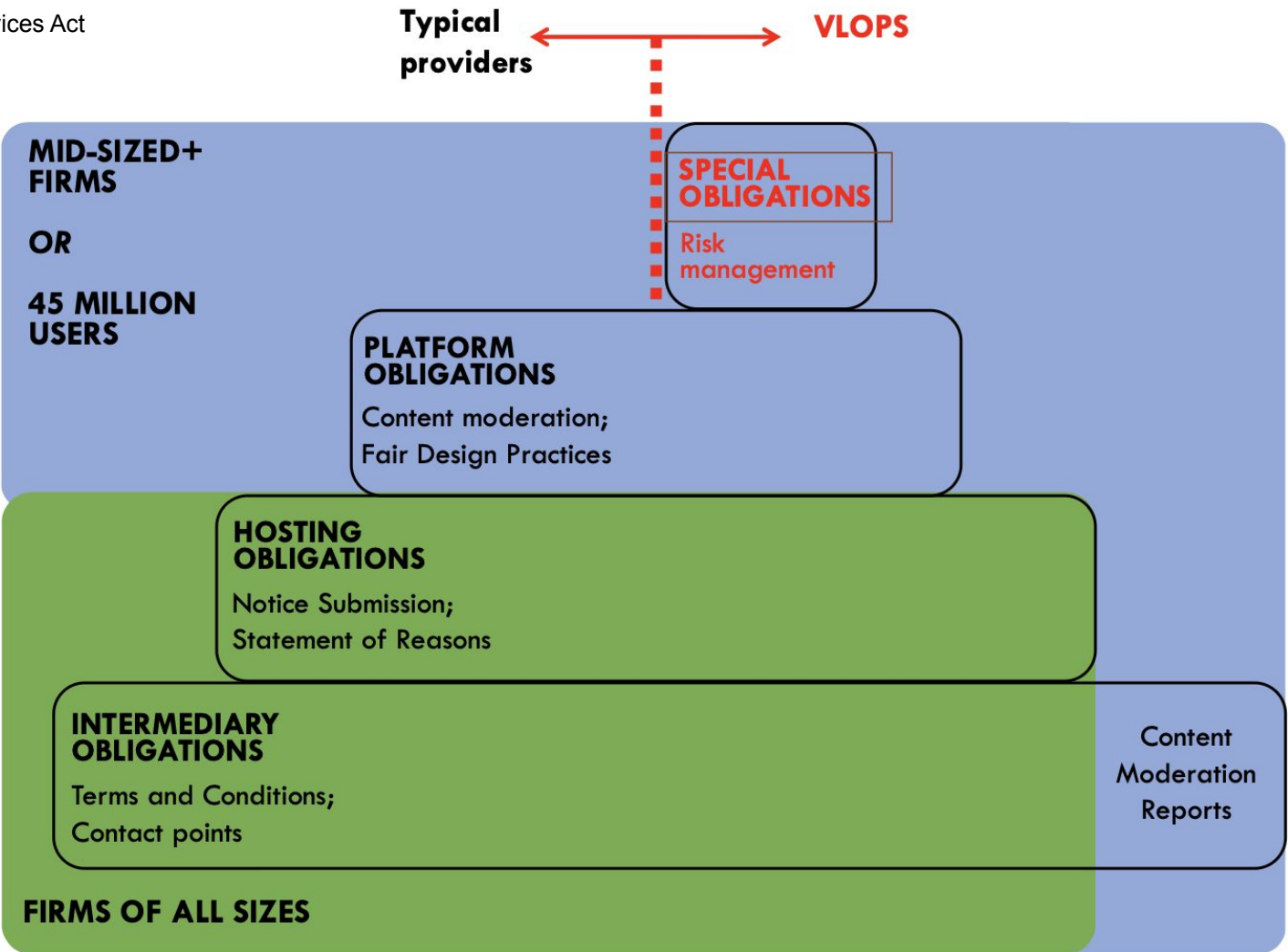
- The DSA sets out different due diligence obligations depending on the technical service offered by the provider.
- The main regulatory areas are content moderation, risk management and transparency.



Source: Husovec, M. (2024). *Principles of the Digital Services Act*. Oxford University Press.

Chart: Husovec, M. (2025). Digital Services Act Specialist Masterclass (Module 1).

Note: the chart is ONLY for illustrative purposes. You are not expected to learn it for the exam.



So, what is an online platform?

- Online Platforms are hosting services.
- Hosting Services are defined as follows: storage of information provided by, and at the request of, a recipient of the service.

Art. 3(i):

‘online platform’ means a hosting service that, at the request of a recipient of the service, **stores and disseminates information to the public**, unless that activity is a minor and purely ancillary feature of another service or a minor functionality of the principal service and, for objective and technical reasons, cannot be used without that other service, and the integration of the feature or functionality into the other service is not a means to circumvent the applicability of this Regulation.

Online Platforms = storage + dissemination of information [& be a mid-sized company]

Complex cases:



- To define online platforms, integration is KEY.

Simulation Case: *ClipTok* and the Urban Movement Lessons

- In-class activity.

Raquel De Haro i Pérez

PhD Student

raquel.deharo@gess.ethz.ch

ETH Zürich

Center for Law & Economics

Building IFW

Haldeneggsteig 4

8092 Zurich

<https://lawecon.ethz.ch/>

Luka Nenadic

PhD Student

luka.nenadic@gess.ethz.ch

ETH Zürich

Center for Law & Economics

Building IFW

Haldeneggsteig 4

8092 Zurich

<https://lawecon.ethz.ch/>