

# Synthetic content moderation

From algorithmic content moderation to algorithmic  
moderation of synthetic content...?

29.10.2025

WIPO Conversation on IP and Frontier Technologies

sebastian.f.schwemer@bi.no

@schwemer.bsky.social

BI Norwegian Business School



For the last decade or so (algorithmic) content moderation has been a central issue on online platforms, now being addressed in **several EU laws**...

From algorithmic content moderation to  
algorithmic moderation of synthetic content...?



How adaptable are those EU frameworks to challenges posed by synthetic media?

EU IP/copy-right angle

Synthetic media (incl. AI-generated or manipulated images, videos, audio, and text) is rapidly evolving; these technologies can realistically but falsely depict individuals, replicating voices, likenesses, and personal styles.

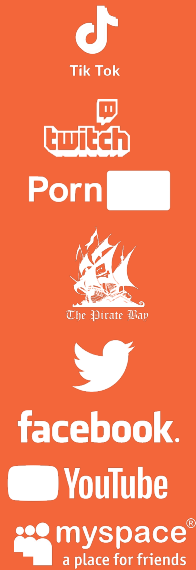
**The current EU  
framework of  
content  
regulation\***



2000



**E-Commerce Directive**



**EU Commission:  
Communication  
Towards a modern,  
more European  
copyright framework**

May 2015

December 2015

**EU Commission:  
Strategy for the digital  
single market**

May 2016

September 2016

**EU Commission:  
Proposal for a Directive  
on copyright in the  
Digital Single Market  
COM(2016)0593**

**EU Commission:  
Proposal for AVMSD refit  
COM(2016)**

**Code of Conduct  
Countering illegal hate  
speech**





“provide clarifications to platforms on their liability when they take **proactive** steps to detect, remove or disable access to illegal content”

## Communication: Tackling Illegal Content Online, Towards an enhanced responsibility of online platforms

- ◆ Gemini
- ✦ Claude
- ⌘ DALL·E
- ⌘ ChatGPT
- ud.io
- stability.ai
- Midjourney

## AI Act



2017

2018

2018

2019

2022

2024



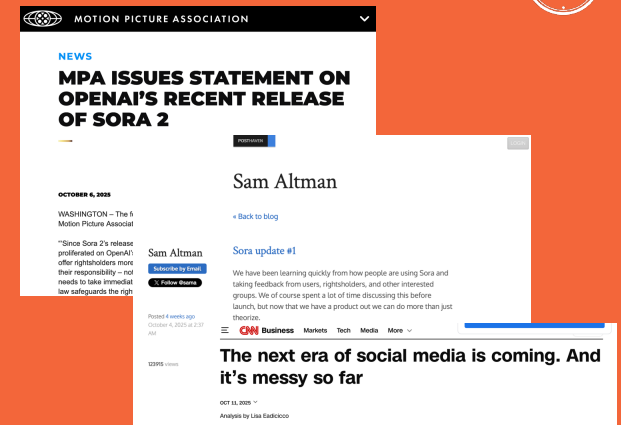
Recommendation on  
measures to effectively  
tackle illegal content  
online

EU Commission:  
Proposal for regulation  
on terrorist content

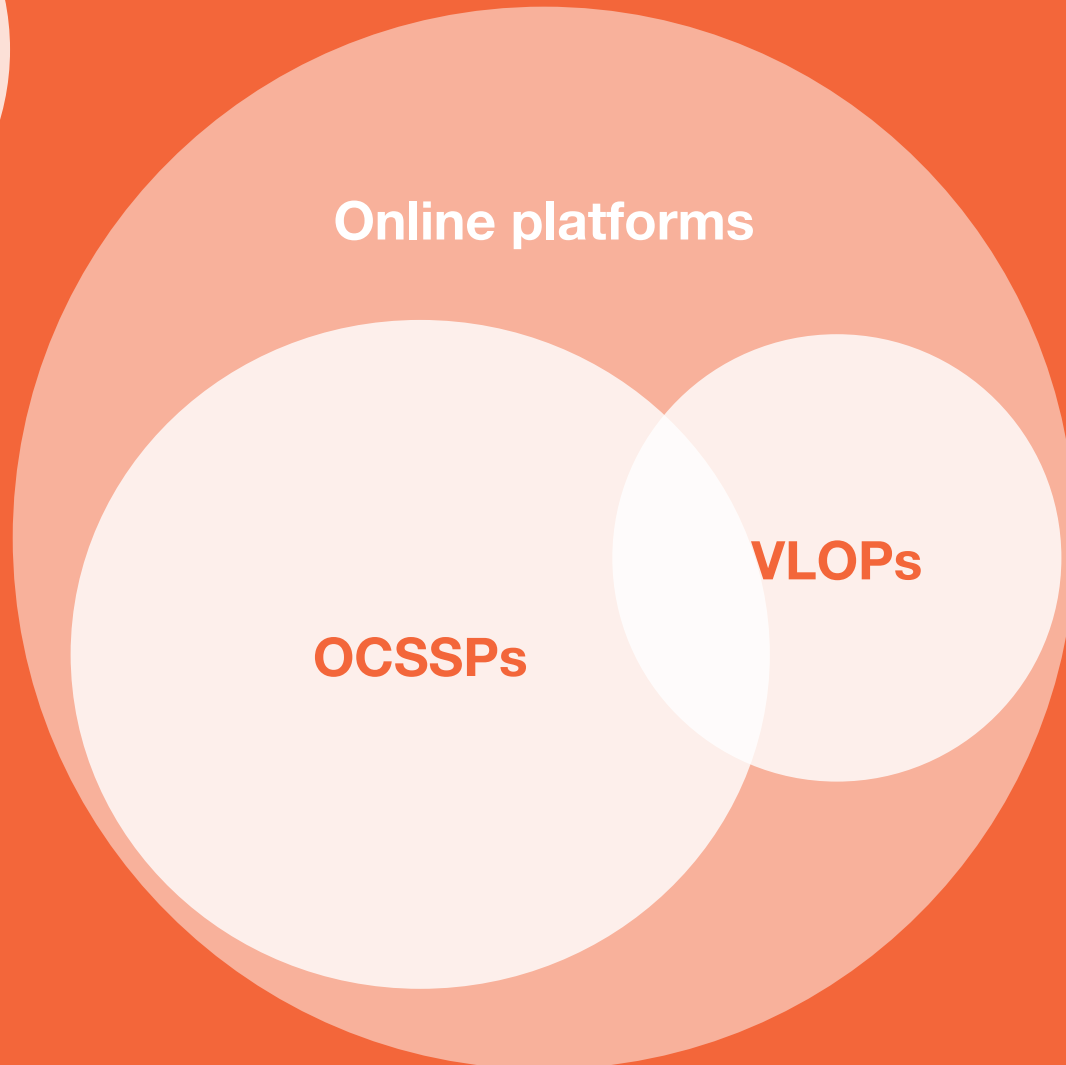


Art. 17 CDSM  
Directive

Digital Services Act  
Regulation (EU)  
2022/2065



**genAI  
platforms?**



**Digital Services Act**



**Art. 17 DSM Directive**

\*Quintais & Schwemer (2022). The Interplay between the Digital Services Act and Sector Regulation: How Special Is Copyright?, EJRR, doi:10.1017/err.2022.1

## Content moderation



## Recommendation

means the activities, whether automated or not, undertaken by providers of intermediary services, that are aimed, in particular, at **detecting, identifying and addressing illegal content or information incompatible with their terms and conditions**, provided by recipients of the service, including measures taken that affect the availability, visibility, and accessibility of that *illegal* content or that information, such as demotion, demonetisation, disabling of access to, or removal thereof, or that affect the ability of the recipients of the service to provide that information, such as the termination (...) of a recipient's account;

## Illegal content: Member State or EU law



UDKAST

Forslag  
til  
Lov om ændring af lov om oplysnings-  
(hæftelse af en præstationsbeskyttelse og beskyttelse mod digital gene-  
renselse af digitale tjenester)

§ 1

1. Lov om oplysning, if. lovhæftingsparagraf nr. 105 af 20. marts  
2018 og lov nr. 476 af 11. juni 2024, fastsættes følgende lov:

1. Oversat i lovens indledning eller vedvarende kommentare: «  
eller « advarende kommentar: » eller artikel, efter som indvænt  
eller en artikel og efter vedvarende kommentar eller artikel

2. I § 47, stk. 2, indføres efter § 53 og § 87 A, stk. 3, a:

3. I § 47 indføres efter stk. 3 som nyt stykke:  
« 308. A. Regeringen kan fastsætte regler om digital be-  
medt forvalt, herunder om anvendelse af bestemte IT-systemer  
glade tjenester og tjenester og den til enhver tid gældende re-  
levning af lign.»

**BRED AFTALE OM DEEPPAKES GIVER  
ALLE RET TIL EGEN KROP OG EGEN  
STEMME**

Regeringen og et bredt udsnit af Folketingets partier er enige om, at der skal være en  
stærkere beskyttelse mod digitale efterligninger af personlige kendetegn.

Regeringen (Socialdemokratiet, Venstre og Moderaterne).

## Digital Services Act (“general framework”)

Hosting (≈ platform) liability exemption  
(Article 5 DSA): **notice and takedown**,  
starting point illegal (or contrary to T&C’s)

Good Samaritan clause (Article 7 Voluntary  
own-initiative investigations)

Assymmetric due diligence **obligations**

## Article 17 CDSM Directive (“copyright”)

Online content sharing service providers

4. OCSSPs **liable** (...), **unless**:
- (a) made **best efforts** to obtain authorisation,
  - (b) made, in accordance with high industry standards of professional diligence, **best efforts to ensure the unavailability** of specific works for which the rightholders have provided (...) necessary information, and in any event
  - (c) notice and staydown.

(C-18/18 – Facebook Austria?)



Prohibition of general monitoring obligation, Art. 8 DSA

synthetic vs. authentic content  
≠  
illegal vs. legal content

# Basic tension remain: Over-removal vs. under-removal, but what is the benchmark?

		Infringing	
		yes	no
Takedown	yes	True positive (TP)	False positive (type-I error)
	no	False negative (type-II error)	True negative (TN)

Quality  $\approx$  How much “error” is acceptable?

**Art. 17(7) CDSM:** Cooperation between OCSSPs and rightholders **shall not result** in the prevention of the availability of works, which do not infringe copyright

MS shall ensure that users are able to rely on exceptions or limitations when uploading and making available content *generated by users* on OCSS: **quotation, criticism, review; caricature, parody or pastiche.**

DSA recital 25: Automation technology must be “sufficiently reliable to **limit** to the maximum extent possible the **rate of errors**” [also: reporting on accuracy and error rate]

C-401/19

Good luck, Article 17 – it won't be getting easier just now



C. Michael Gibson MD  
@CMichaelGibson

Folgen

Determining if an image is a Chihuahua or muffin is a tough problem in artificial intelligence



16:53 - 14. Mai 2017

7.665 Retweets 9.993 „Gefällt mir“-Angaben



210 7,7 Tsd. 10 Tsd.

# Procedural safeguards and (much) more...

## Complaint handling

**Art. 17 (9) CDSMD: Complaint-handling + human review.**

**Art. 20 DSA: Internal complaint-handling system + human review.**

## Transparency + fundamental rights

**Art. 14 DSA – Terms and conditions:** shall include information on any policies, procedures, measures and tools used for the purpose of content moderation, including algorithmic decision-making and human review (...). + providers shall act in diligent, objective and proportionate manner in applying and enforcing the restrictions (...), with due regard to the rights and legitimate interests of **all** parties involved (...)

**Art. 15 DSA: Transparency reporting**

**Art. 17(3) DSA: Statement of reasons (≈ “explanation”?)**

## Systemic risks (VLOPs/VLOSE)

Dissemination of illegal content, negative effects for exercise of fundamental rights (assessment and mitigation, **Arts. 35/36 DSA**)

# The EU content moderation rules - fit for synthetic media? Well...

“What” is being moderated?

Synthetic content shared by *humans* vs. shared by *machines*.... “traditional” copyright questions? Synthetic ≠ illegal per se (deceptive?)



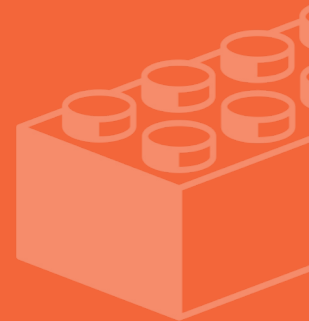
“Where” is moderation happening?

Moderating “output” (i.e. genAI under DSA/CDSM)? vs. Moderating “use” of synthetic content on platforms?

Synthetic content challenges traditional assumptions BUT frameworks (transparency, T&C, etc.) remain relevant (with adjustments?)

NB: Copyright and IP issues but there’s other angles too: personality rights, privacy...

CJEU: “the need for safeguards ‘is all the greater where the interference [with fundamental rights] stems from an automated process.’”\*



\*Case C-401/19, *Republic of Poland v. European Parliament and Council*, ECLI:EU:C:2022:297, para. 67; C-311/18, *Facebook Ireland and Schrems*, EU:C:2020:559 para. 176, Opinion 1/15 of the Court, ECLI:EU:C:2017:592, para. 141, Joined cases C-293/12 and C-594/12, *Digital Ireland and Others*, ECLI:EU:C:2014:238, para. 55, case-law of the ECtHR...

# Discussion and Comments

## Further reading

- Schwemer, S.F. (2024) Decision Quality and Errors in Content Moderation. *IIC* 55, 139–156. <https://doi.org/10.1007/s40319-023-01418-4>
- Quintais, J.P., Katzenbach, C., Schwemer, S.F. et al. (2024) Copyright Content Moderation in the European Union: State of the Art, Ways Forward and Policy Recommendations. *IIC* 55, 157–177. <https://doi.org/10.1007/s40319-023-01409-5>
- Quintais, JP. & Schwemer, S.F. (2022). ‘The Interplay between the Digital Services Act and Sector Regulation: How Special Is Copyright?’ *European Journal of Risk Regulation* 1-31 <https://doi.org/10.1017/err.2022.1>
- Margoni, T, Quintais, JP and SF Schwemer (2022) . ‘Algorithmic propagation: do property rights in data increase bias in content moderation? Part I + 2, <http://copyrightblog.kluweriplaw.com/2022/06/08/algorithmic-propagation-do-property-rights-in-data-increase-bias-in-content-moderation-part-i/>

## Creative commons attribution (all from the Noun Project)

Scale by Amelia, cassette tape by Alexander Skowalsky, AI by ic2icon, Stopwatch by Veronika Krpciarova, filter By Eucalyp, Denmark by Chameleon Design, yoga by Mariia Nisiforova, european union by Federico Panzano, IT, Copyright by Arthur Shlain, screws by Oleksandr Panasovskyi Algorithm by Eucalyp

29.10.2025

WIPO Conversation on IP and Frontier Technologies

sebastian.f.schwemer@bi.no

@schwemer

@schwemer.bsky.social