

# DSA and Content Moderation

AI, Free Speech, and Privacy



Group 14

Lavinia Benetollo

Alessandro Ferraiolo

Rachel Gabel

Lorenzo Giambra

Mark Maci

# TOPICS

- DSA: BALANCING FREEDOM AND SAFETY ONLINE
- DSA & FREE SPEECH
- RISK MITIGATION, PRIVACY DSA VS GDPR VS AI ACT
- AI AND CONTENT MODERATION
- DSA AND CONTENT MODERATION

# DSA: Balancing Freedom and Safety Online

The Digital Services Act (DSA) is an EU Regulation aiming to assure fairness, trust, and safety in the digital environment, while, at the same time, ensuring Freedom of expression and the development of ISPs (internet service providers) by exempting them from liabilities for third-party content.



2000  
Electronic  
Commerce  
Directive



16-12-2020  
Digital service  
Package  
(DSA & DMA) Proposal



27-10-2022  
Publication of  
DSA on  
Official Journal



25-08-2023  
Effective for  
VLOPs and VLOSEs



17-02-2024  
DSA effective  
in its entirety

ISPs: Companies providing services for accessing, using, or participating in the Internet

Application → Markets & Users in EU



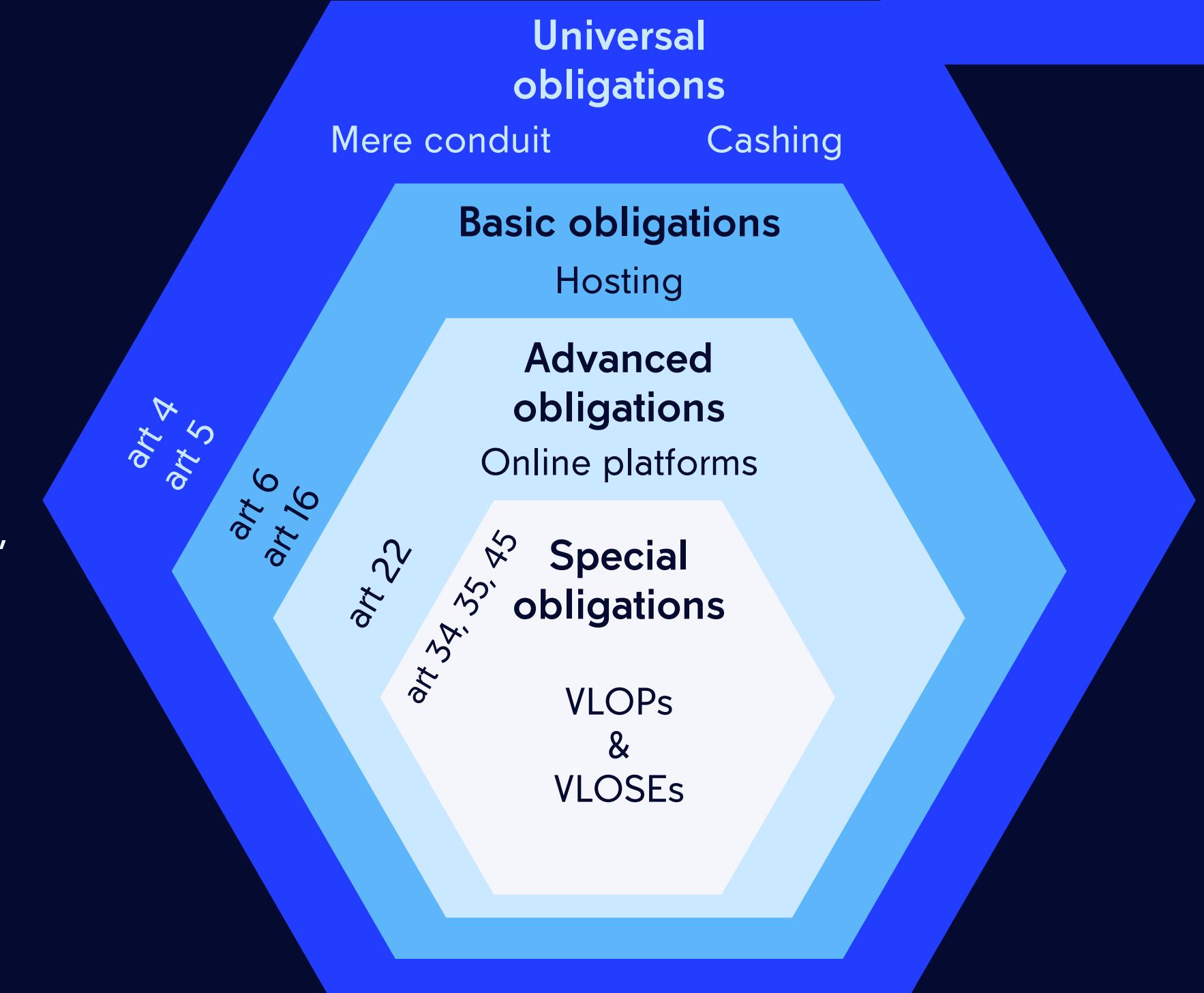
# DSA: Risk-based approach

The framework establishes tiered obligations for ISPs:

- The strictness of obligation increases with the ISP's complexity and size.
- Within a group of ISPs, each follows its specific group obligation and those of its parent group.

Relevant articles:

- art 4 : "Mere conduit"
- art 5 : "Cashing"
- art 6 : "Hosting"
- art 7 : "Voluntary own-initiative investigations "
- art 8 : "No general monitoring or active fact-finding obligations"
- art 16 : "Notice & Action"
- art 22 : "Trusted Flaggers"
- art 34 : "Risk Assessment"
- art 35 : "Risk Mitigation"
- art 45 : "Codes of conduct"



# DSA articles and free speech



---

Art. 11 DSA vs Art 10 of ECHR

---

Art. 12 (Exemption of liability for mere conduit)  
& Art 13 (Exemption of liability for catching)

---

Art 14 (Exemption of liability for hosting  
services)

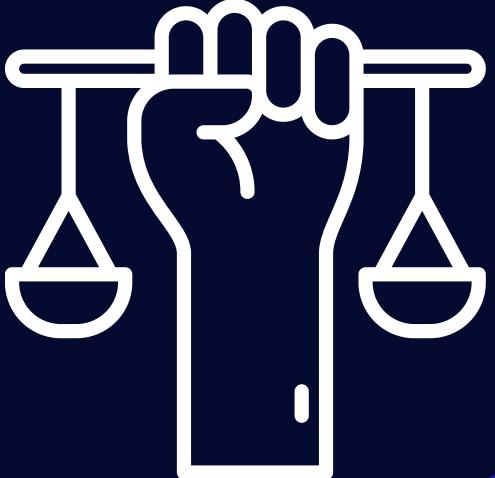
---

Art 15 prohibits general monitoring obligations

---

Art 34 Risk Assessment

# DSA - Focus on its relation with free speech



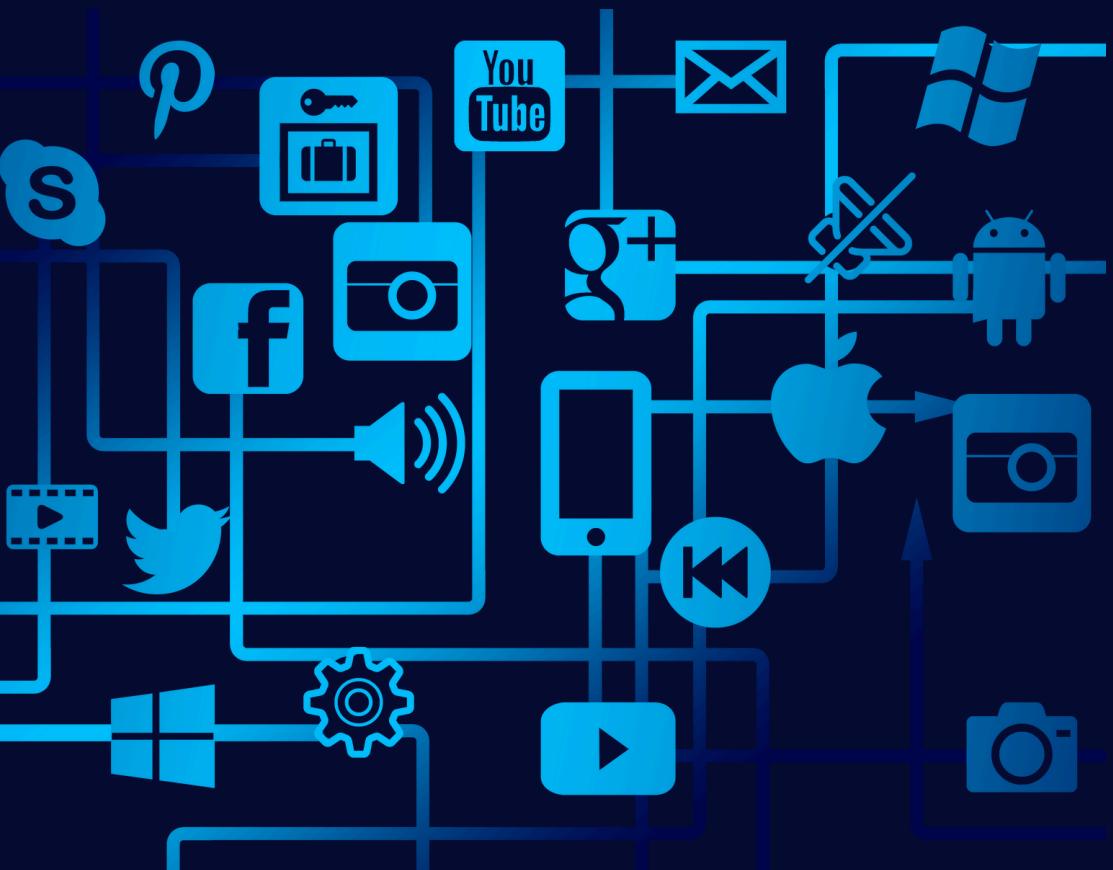
Transparency in  
content  
moderation

Protection  
against arbitrary  
censorship

Right to appeal:  
power to users

Prohibition of general  
monitoring obligations:  
Art 15 & general obligations  
on platforms to monitor the  
information they transmit or  
store

# DSA - Focus on its relation with free speech



# Legal content protection: illegal online ≠ illegal online

# Risk assessment and mitigation: Art 25

# Empowering users and researchers: Art 35

# Risk assessment mitigation

The DSA mandates VLOPs and VLOSEs, as defined in Article 25, to assess and mitigate systemic risks that could affect societal values, including FoE (Risk Management).

Articles 26 and 27 discuss the assessment and mitigation measures

- VLOPs and VLOSEs must undergo Independent Audits
- Crisis mitigation
- Systemic Risk Assessments

# DSA, Privacy and Risk Mitigation

- **Transparency Obligations:**

- Article 24 online advertising transparency.
- Article 30 online advertising transparency, publicly available repository for VLOPs and VLOSEs.
- Article 34 transparency, content moderation standards.

- **Data Management and Accountability:**

- Article 28 conducts risk assessments
- Article 33 mandates external audits and transparency reports.

- **User Complaints:**

- Article 17 sets internal complaint-handling systems

- **User Empowerment:**

- Article 18 outlines procedures for out-of-court dispute settlement



# GDPR-DSA-AI Act



# Online platforms and media



- inform the public
- participate in public life
- are a channel for external voices



They play a crucial role in the passive dimension of the FoE

“Right to be informed”

they guarantee the correct functioning of **democracy**

This migration of the information market to the Internet has had to face the typical goal of algorithmic systems for content curation which is that of maximizing users' engagement.

# Online platforms and media

Because of the enormous amount of content posted online everyday, internet intermediaries use AI systems to moderate content



## Hard moderation

filters out illegal content  
through hybrid moderation  
systems

combination of humans and AI tools

## Soft moderation

organization and suggestion of  
content through recommender  
systems      algorithms that analyze user  
preferences to generate personalized  
recommendations



# Impact of content curation

Regarding **content curation** (soft moderation), the use of algorithmic filters could lead to

the limited exposure of the user to diverse content and viewpoints

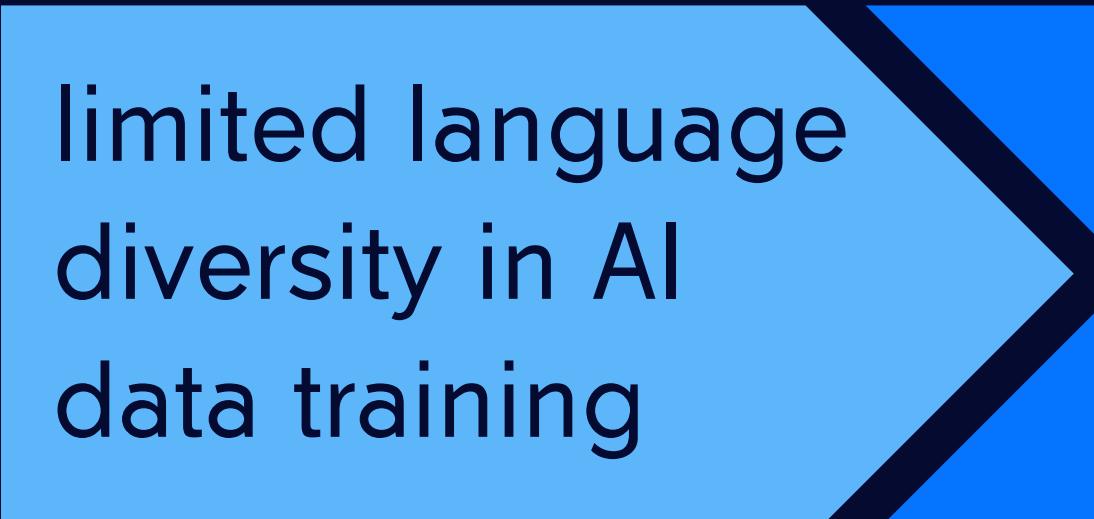
Example: On February 9 2024 Meta has introduced a new feature on Instagram to reduce the recommendation of political content. This setting, which can be changed from the menu of the app, is implemented without obtaining any consent from the user and without providing any explicit notification in-app.



# Impact of content moderation

For **hard moderation** the main problem is the errors AI commits:

- removing legitimate content
- missing illegal material



limited language diversity in AI data training

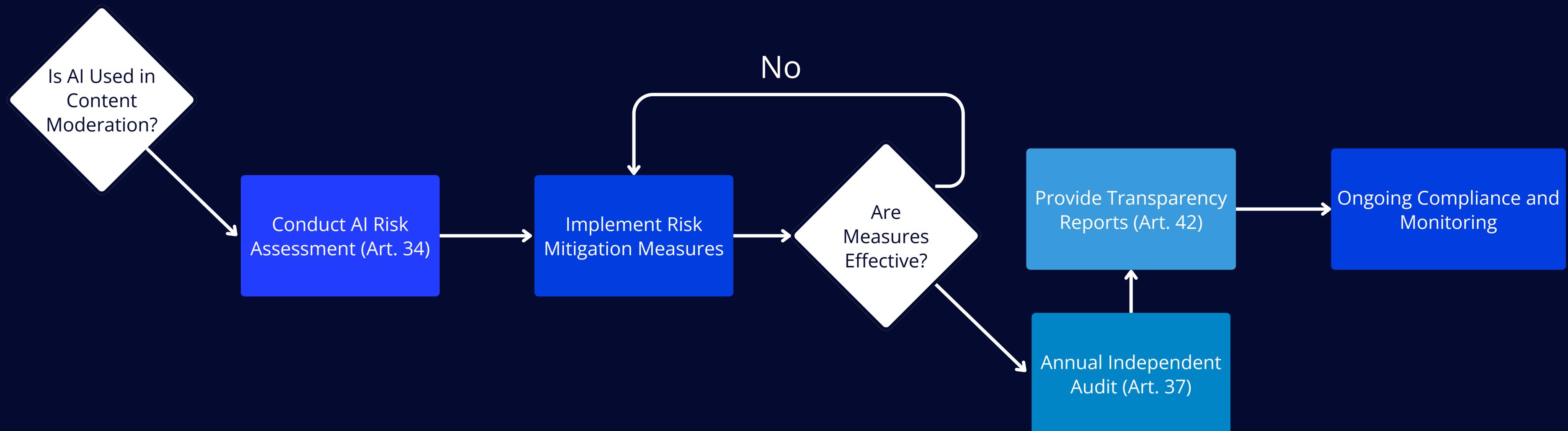


discriminatory outputs and censorship of marginalized groups



some countries have reacted to the lack of quality of content moderation with internet shutdowns and restrictions

# DSA's Regulatory Framework for AI in Content Moderation



# Commission Requests for AI Content Under DSA

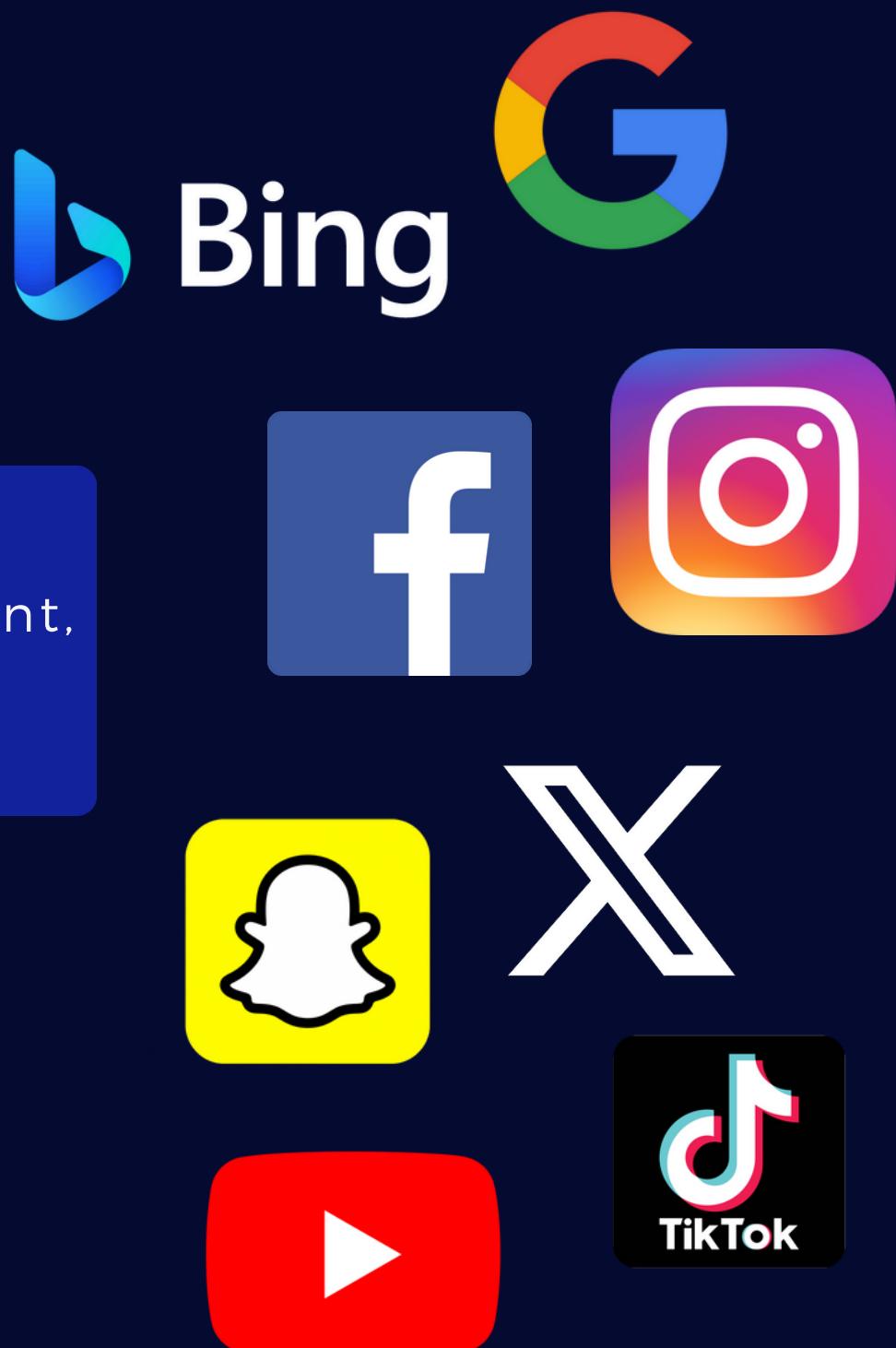
On March 24 2024, the EU Commission requested two VLOSEs and 4 VLOPs regarding generative AI risks



The requests stress the risks of AI hallucination, viral deepfake content, election interference among other concerns

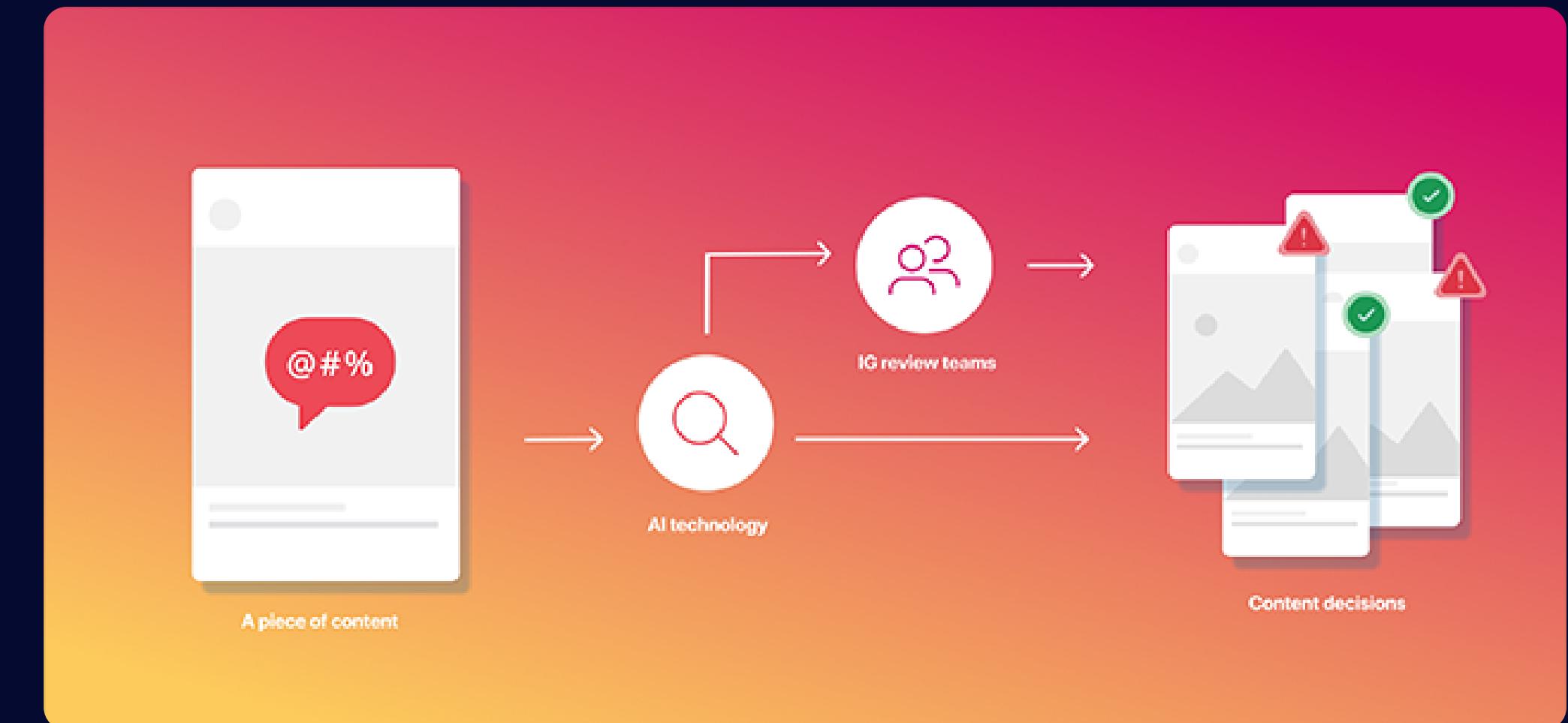


Pursuant to Article 74 (2) of the DSA, the EU can impose fines and periodic payments for failure to comply or provide information



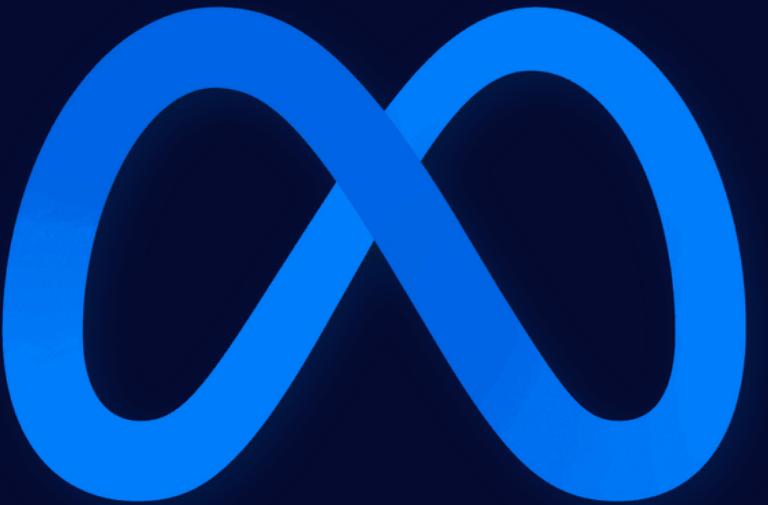
# Non-Obligatory Content Moderation: Article 8

- Under DSA art. 8, there is no general obligation for platforms to monitor for illegal content or practices
- Providers may monitor their platforms, but must disclose how AI models are used in content moderation (algorithmic accountability)



# Meta and TikTok vs. European Commission

- Under the DSA, VLOPs are to pay 0.05% of yearly profit to the Commission
- Meta and TikTok have brought legal action against the EU in protest of there required payments
- The EU has stated it needs €45.2 million to supervise over two dozen VLOPs
- Other platforms such as X, Snapchat, Wikipedia, etc. record small or no net income, leaving Meta and TikTok with bills of up to €11 million



# DSA - Final Message

- **DSA's pivotal role:** Ensures transparency, fairness, and fundamental rights protection in Europe's digital landscape.
- **Daily impact:** Crucial for safe and equitable online interactions among European citizens.
- **Challenges with AI:** Pre-AI era framework struggles to address complexities introduced by generative AI technologies.
- **Need for updates:** to maintain relevance and effectiveness amid digital advancements the EU Parliament enacted the AI Act.