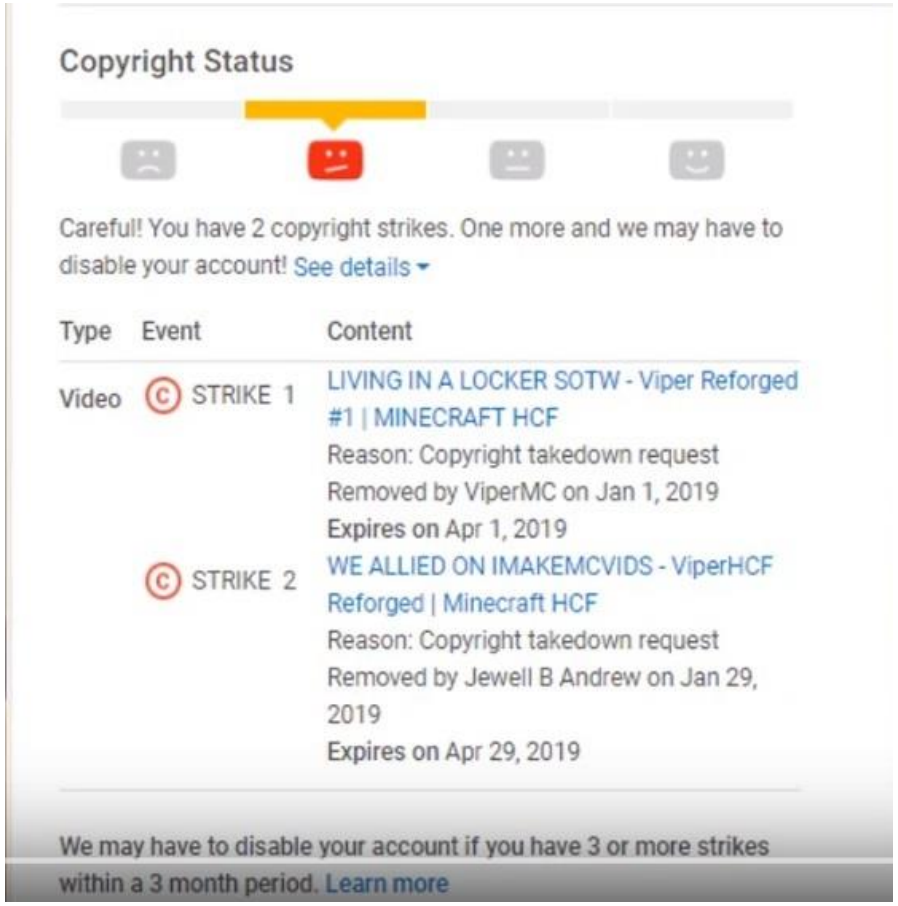
LSE Department of Law

How Does the EU's Digital Services Act Regulate Content Moderation? And Will it Work?

- Dr. Martin Husovec

Department of Law

YouTube & copyright infringers



LSE Department of Law

Facebook & breastfeeding

Department of Law

Twitter & Trump

Donald J. Trump
@realDonaldTrump

45th President of the United States of
Traducir la biografía

Washington, DC Vote.DonaldJ

49 Siguiendo **25,2 M** Seguidores

"The 75,000,000 great American Patriots who voted for me, AMERICA FIRST, and MAKE AMERICA GREAT AGAIN, will have a GIANT VOICE long into the future. They will not be disrespected or treated unfairly in any way, shape or form!!!"

Shortly thereafter, the President Tweeted:

"To all of those who have asked, I will not be going to the Inauguration on January 20th."

Department of Law

AWS & hate speech

TECH • AMAZON

Amazon Will Suspend Hosting For Pro-Trump Social Network Parler

Amazon's suspension of Parler's account means that unless it can find another host, once the ban takes effect on Sunday Parler will go offline.

John Paczkowski
Technology and Business Editor

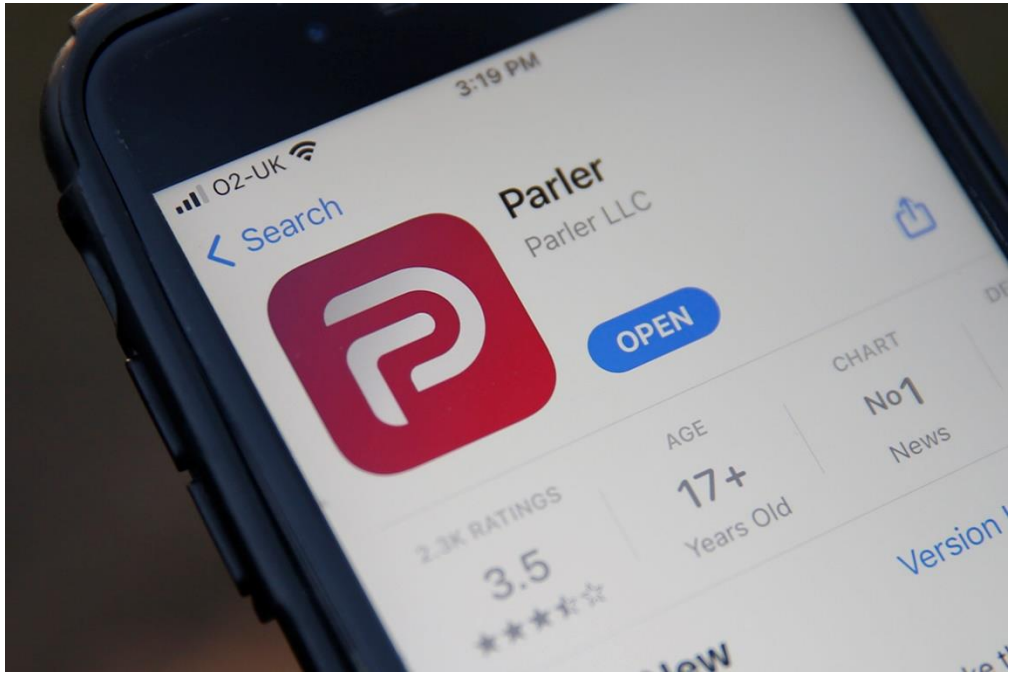
Ryan Mac
BuzzFeed News Reporter

Updated on January 10, 2021 at 3:08 am
Posted on January 10, 2021 at 2:07 am

View 126 comments

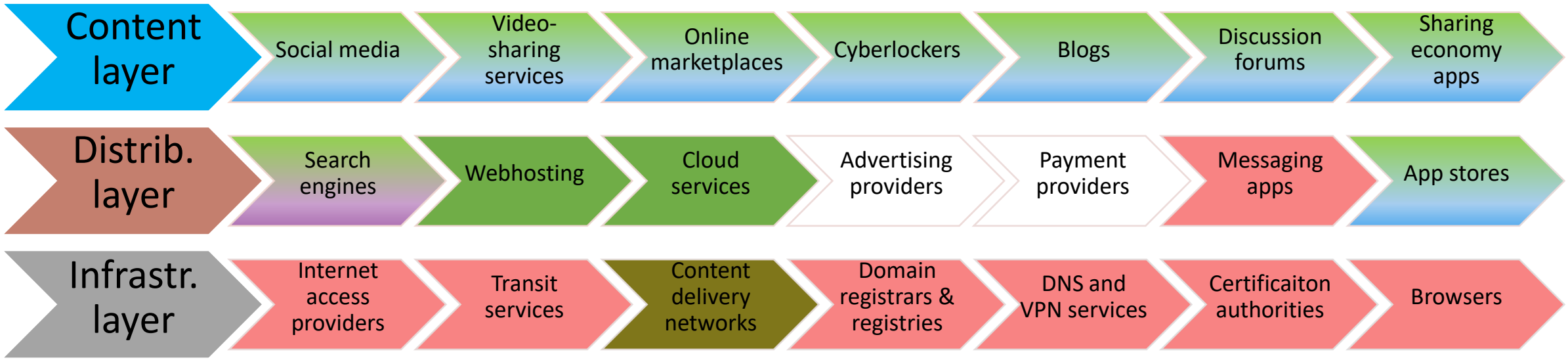
LSE Department of Law

Apple App Store



Department of Law

The main focus of regulation



Hosting, Online Platform, Mere Conduit, Caching, VLOSE

LSE Department of Law

Digital Services Act – from Feb 2024

Document 32022R2065

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance)

PE/30/2022/REV/1

OJ L 277, 27.10.2022, p. 1–102 (BG, ES, CS, DA, DE, ET, EL, EN, FR, GA, HR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV)

In force

ELI: http://data.europa.eu/eli/reg/2022/2065/oj

LSE Department of Law

A new generation of rules

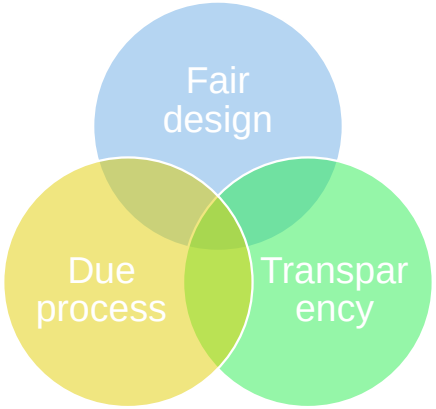
	United States	European Union
I. generation: 1996-2000	Sec 230 CDA; Sec 512 DMCA	Articles 12-15 ECD
II. generation: 2020-?	? [PACT]	**Digital Services Act**

- **I. generation:** breathing space for speech & industries
  - Liability exemptions to avoid strict (or any) liability
- **II. generation:** regulation of risks posed by services
  - Regulatory expectations that overlay the liability social contract

LSE Department of Law

Due diligence obligations vs liability exemptions

- The shift from liability for content to (legal) accountability for the design of services
  - Failing to comply does not lead to the illegality of service or liability for users' actions but targeted non-compliance with the stand-alone obligations
  - Even services which are *not* protected anymore by exemptions remain subject to due diligence obligations



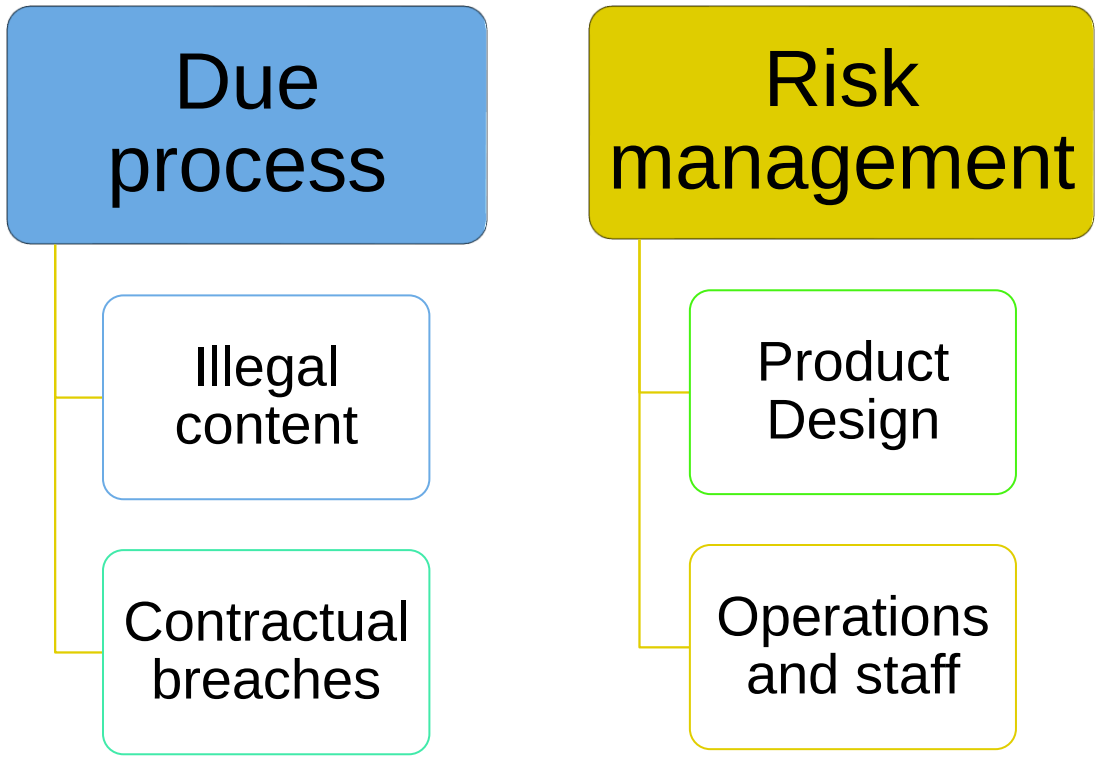
LSE Department of Law

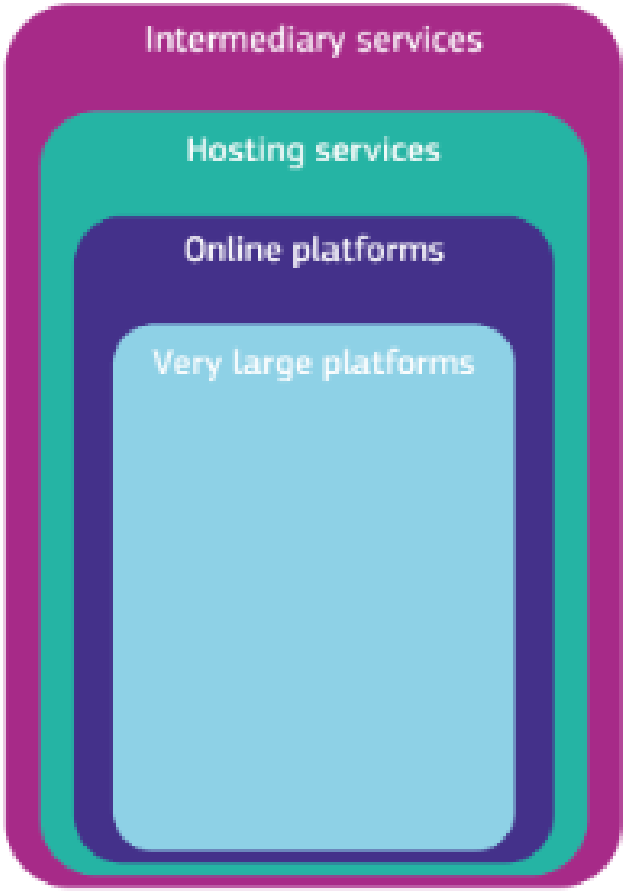
The guts of the DSA

Obligations	Universal *All providers of conduit, caching, hosting services*	Basic *all hosting services*	Advanced *medium-to-large[1] online platforms*	Special *VLOPs & VLOSEs*
Content Moderation	**Art 14** (fair content moderation)	**Art 16** (notice) **Art 17** (statement of reasons)	**Art 20** (internal redress); **Art 21** (out-of-court mechanism); **Art 22** (trusted flaggers); **Art 23** (anti-abuse provisions); **Art 30-32** (specific rules on B2C marketplaces)	**Art 34-35** (risk mitigation assessment) **Art 36** (crisis response mechanism)
Fair Design (user interfaces, recommender systems, advertising and other parts)	**Art 14** (fair content moderation)	**Art 16** (user-friendly notice and action)	**Art 25** (fair design of user-experience); **Art 26(3)** (advertising); **Art 27** (recommender systems); **Art 28** (protection of minors); **Art 30** (traceability of traders); **Art 31** (facilitating design for traders)	**Art 38** (recommender systems) **Art 39** (risk mitigation assessment)
Transparency	**Art 15** (annual reporting)	**Art 17(5)** (database of all the statements of reasons)	**Art 22** (reports by trusted flaggers); **Art 24** (content moderation reports); **Art 26** (advertising disclosure)	**Art 39** (advertising archives); **Art 42** (content moderation transparency)
Oversight	**Art 11** (regulator's contact point); **Art 12** (recipient's contact point); **Art 13** (legal representative)	**Art 18** (notification of suspected relevant crimes)	(-)	**Art 37** (auditing); **Art 40** (data access/scrutiny); **Art 41** (compliance function)

LSE Department of Law

DSA's two main tools:





Overview of the DSA's due diligence obligations

Obligations	Universal	Basic	Advanced	Special
	All providers of conduit, caching, hosting services	*all hosting services*	*medium-to-large[1] online platforms*	*VLOPs & VLOSEs*

- Online platforms = medium-to-large firms (50+ employees or turnover 10+ million EUR)
- VLOPs/VLOSEs = 45+ mil average active monthly EU users

Department of Law

VLOPs / VLOSEs

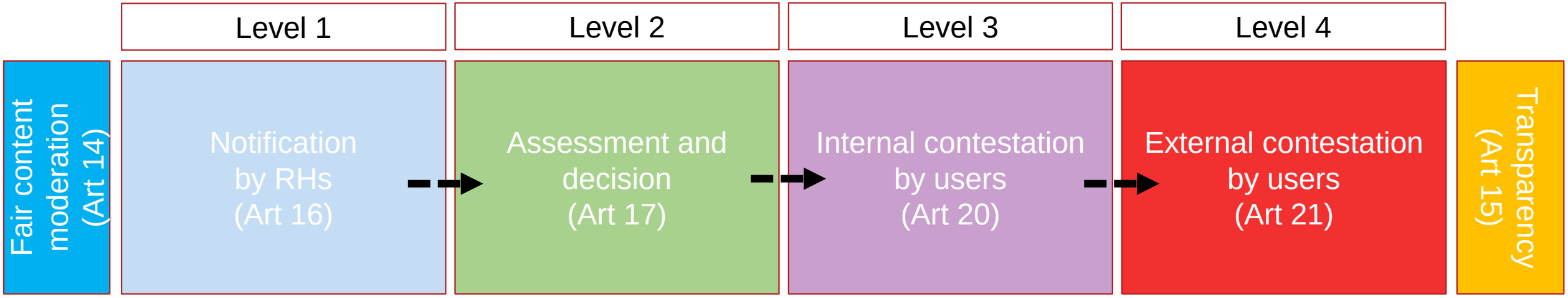
	Company	Digital Service	Type	Est. (cc)	Users (mil)	User-generated-content components
Search	Alphabet[11]	Google Search	VLOSE	IE	332+	Paid and unpaid search results
	Microsoft[12]	Bing	VLOSE	IE	107	Paid and unpaid search results
Social media	Alphabet	YouTube	VLOP	IE	401+	Videos, sound, photos & text
	Meta[13]	Facebook	VLOP	IE	255	Videos, sound, photos & text
	Meta	Instagram	VLOP	IE	250	Videos, sound, photos & text
	Bytedance[14]	TikTok	VLOP	IE	125	Videos, sound, photos & text
	Microsoft	LinkedIn	VLOP	IE	122	Videos, sound, photos & text
	Snap[15]	Snapchat	VLOP	?	96+	Videos, sound, photos & text
	Pinterest[16]	Pinterest	VLOP	?	n/a	Videos, sound, photos & text
	Twitter[17]	Twitter	VLOP	?	100+	Videos, sound, photos & text
App stores	Alphabet	Google App Store	VLOP	IE	274+	Mobile apps
	Apple[18]	Apple App Store	VLOP	IE	n/a	Mobile apps
Wiki	Wikimedia[19]	Wikipedia	VLOP	?	151+	Mostly text and photos
Markets	Amazon[20]	Amazon Marketplace	VLOP	LX	n/a	Sellers' offerings & users' reviews
	Alphabet	Google Shopping	VLOP	IE	74+	Sellers' offerings & users' reviews
	Alibaba[21]	AliExpress	VLOP	?	n/a	Sellers' offerings & users' reviews
	Booking.com[22]	Booking.com	VLOP	NL	n/a	Sellers' offerings & users' reviews
Maps	Alphabet	Google Maps	VLOP	IE	278+	Shop profiles, reviews, etc.

Department of Law

Content Moderation

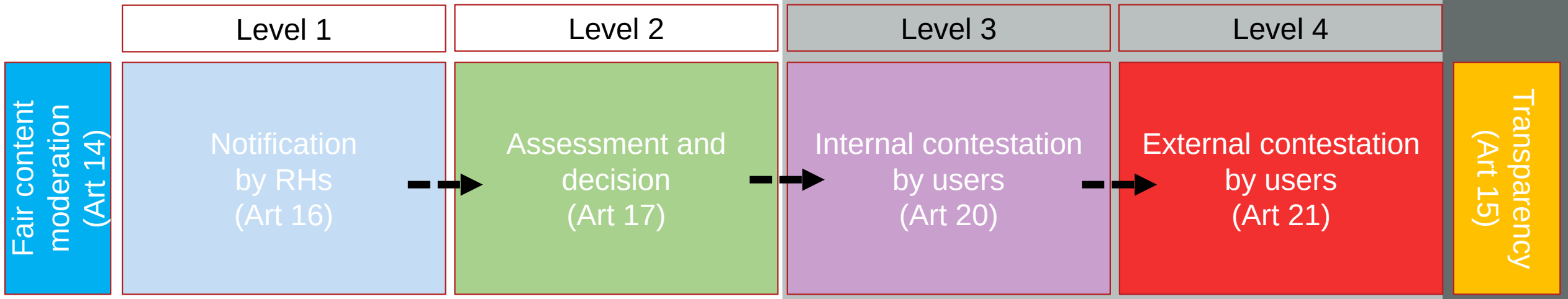
Department of Law

Content Moderation



LSE Department of Law

Content Moderation



Hosting services & any-size

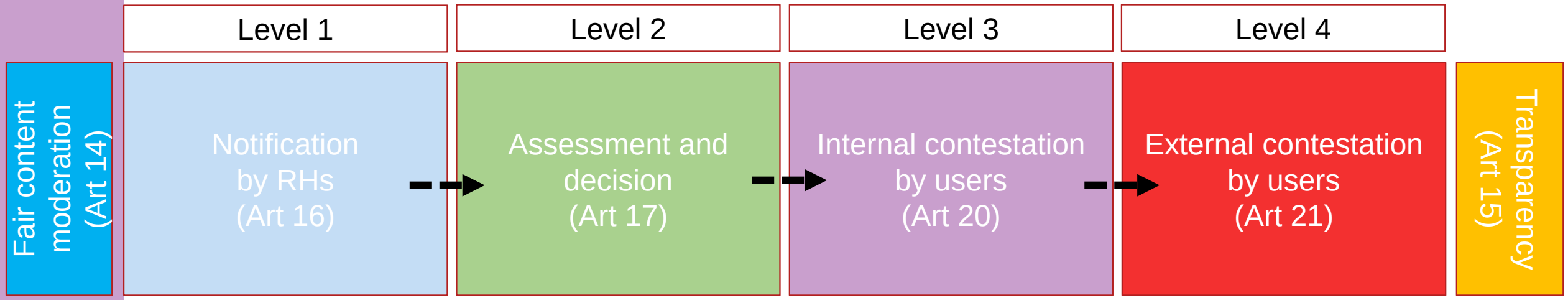
Online platforms & mid-size +

mid-size +

Department of Law

Rule formation

Content Moderation



LSE Department of Law

Two main obligations

- Codification & explanation of all restrictions
  - “any restrictions that they impose in relation to the use of their service in respect of” UGC content
- Conduct content moderation fairly
  - act “diligently, objectively and proportionately” with due regard to the fundamental rights of others

LSE Department of Law

Article 14(1)

"Providers of intermediary services shall include information on any restrictions that they impose in relation to the use of their service in respect of information provided by the recipients of the service, **in their terms and conditions**. That information shall **include** information on any policies, procedures, measures and tools used **for the purpose of content moderation**, including algorithmic decision-making and human review, as well as the rules of procedure of their internal complaint handling system. It shall be set out in clear, plain, intelligible, user-friendly and unambiguous language, and shall be publicly available in an easily accessible and machine-readable format."

LSE Department of Law

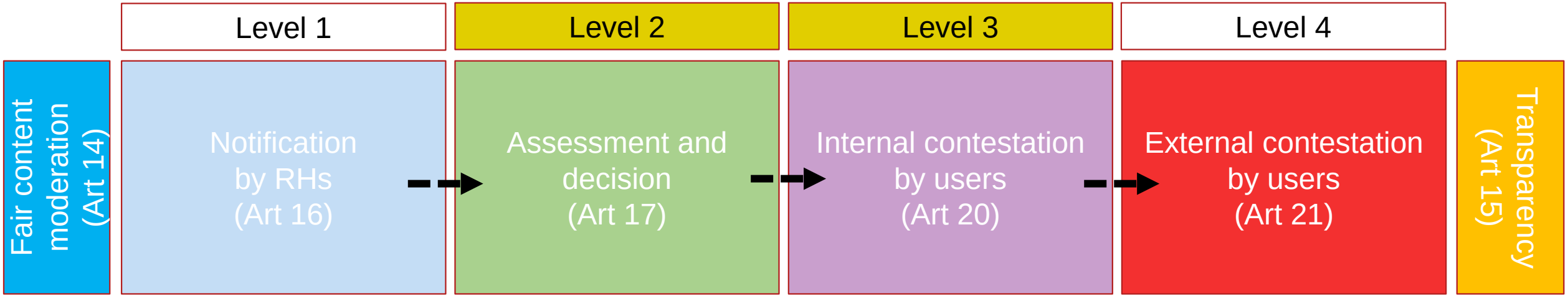
Article 14(4) & Recital 47

"Providers of intermediary services shall act in a diligent, objective and proportionate manner in applying and enforcing the restrictions referred to in paragraph 1, with due regard to the rights and legitimate interests of all parties involved, including the fundamental rights of the recipients of the service"

(47) When designing, applying and enforcing those restrictions (..).

LSE Department of Law

Content Moderation



LSE Department of Law

Level 2 & 3 (internal operations)

- Initial decision (Art 17)
  - providers must issue a statement of reasons
  - broad notion of COMO: visibility, monetization, etc.
  - specific explanation of reasons
  - can be automated (but can small companies do this without vendors?)
- Internal appeal mechanism (Art 20)
  - “easy to access and user-friendly”
  - “not solely automated” + timely, diligent and objective manner
  - inform affected parties

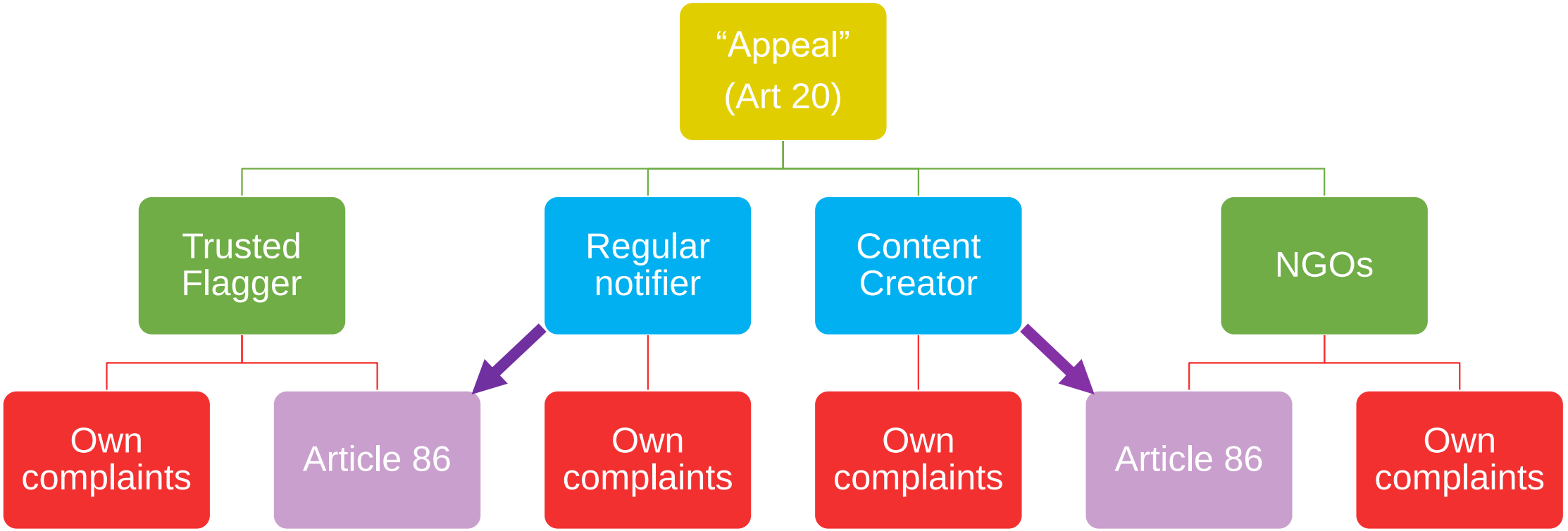
LSE Department of Law

A broad scope of relevant restrictions

Content visibility restrictions (17(1)(a)	Monetisation restrictions (17(1)(b)	Service restriction (17(1)(c)	Account restriction (17(1)(d)	Content moderation outside Article 17
Removing content	Forfeiting earnings	Blocking access to a service	Blocking accounts	Shaming
Suspending content	Suspending access to earnings	Suspending rights on a service	Suspending accounts	Community service
Relocating content	Disabling income on some content	Reducing speed on a service	Erasing accounts	Attach context
Redacting content	Reducing earnings	Limits puts on notifications	Blacklisting registration	Attach warning
Shadow banning or Blacklisting	Fining	Limits on internal complaints		Apology, labelling

Department of Law

Who can complain?



LSE Department of Law

Transparency for mid-sized companies

- Annual reports (or bi-annual for VLOPs) about the content moderation practices (Article 15)
- Aggregate Lumen-like database for all statements of reasons (Article 24(5)) for platforms
  - Will this work? How to anonymise this? But the structuring needed to comply with this can be helpful for standardisation, including annual reports, data access, and risk management

LSE Department of Law

External review: ADR bodies

- Follows broadly Fiala & Husovec 72 (2022) *International Review of Law and Economics* (re funding)
- The national regulators certify ADR providers
  - Up to market or states to create them
- Content creators & notifiers can complain
  - IF they win, provider reimburses the fee + costs
  - IF they lose, they pay the fee

Level 4

External contestation by users (Art 21)

LSE Department of Law

Article 21

- ADR (Art 21): out-of-court settlement bodies
  - Regulators certify entities (must be independent of Ps & users)
    - FB's Oversight Board is clearly not independent in this sense
  - Content creators & notifiers (and their reps) can use the option
  - ADR provider is complainants' choice; no need to exhaust appeals
  - ADR issue decisions: non-binding, Ps must engage in good faith
  - P compensates complainants who win (pays fees & possibly costs)
  - Complainants that lose pays their own fees & costs

LSE Department of Law

The effect of ADR

- External interpreter of platform's rules (= loss of power)
- Incentive for platforms to be clearer (= push to codify)
- Incentive for platforms to resolve internally (= costs)

BUT:

- Does not take away the power to make rules!

LSE Department of Law

Rule-making vs Interpretation

- The basic rule: the **proceduralist approach** constrains interpretation but not rule-making of platforms
- DSA takes mostly proceduralist approach (and a systems-design approach for risk management), with the exception of Art 14(4)
- But my view: Article 14(4): constrains only arbitrary & grossly disproportionate rule-making; all other legal rule-making is fine
- However, once a content rule is expressed by Ps, they don't decide its meaning unless they change it again (as we do with regular contracts)

LSE Department of Law

Case 1: VIP No-Moderation List

- A micro-blogging site decides to create a list of VIPs whose content is not moderated at all (regardless of illegality or contractual nature)
- VIPs are all top elected officials in all countries of the UN
- Is the policy in violation of Article 14(1)?
  - *IMO: not, if properly described.*
- Is the policy in violation of Article 14(4)?
  - *IMO: yes, due to the impact of illegal content (separate from Art 6!)*

LSE Department of Law

Case 2: Pay2Say

- A micro-blogging site has a new Pay2Say product
- For 5 EUR a month, you can say whatever you want on the service, as long it is legal in your country (= no contractual restriction on speech [e.g., disinformation, nudity, vulgar content], only illegality).
- Everyone else is moderated on ToS violations & illegality.
- Is the policy in violation of Article 14(1)? *No if disclosed.*
- Is the policy in violation of Article 14(4)? *Probably no.*

LSE Department of Law

Open issues 1

- Statement of reasons
  - Automation by small providers (part of licensed COMO solutions?)
  - Recommended system changes vs statement of reasons (individualised)
  - FB page owners (eg news orgs) with their own COMO as hosting services
  - Transparency reporting and its standardisation & real-time transmission
- ADR
  - Certification of fee structures, calculation of reasonable costs
  - Transparency obligations, oversight & abuse
  - Scope of specialisation by ADRs

LSE Department of Law

Open issues 2

- Rule-making:
  - Article 14(4) – what is disproportionate?
  - Article 14(1) – technical constrains?
  - Article 14(1) – what is “clear, plain, intelligible, user-friendly and unambiguous language” vs doable
    - Trade-off between: administrability (scalability) & explainability
- Private Enforcement of COMO due diligence obligations
  - Impact on contract law
  - Impact on private claims, e.g., copyright holders

LSE Department of Law

Risk management and its impact on COMO

- VLOPs / VLOSEs are subject to additional requirements
  - mostly extended or intensified reporting obligations
  - researchers' data access
  - unique obligations: profiling-free choice on recommender systems, or advertising archives
- MAIN: a regulatory dialogue about risk management

Department of Law

VLOP

	Company	Digital Service	Type	Est. (cc)	Users (mil)	User-generated-content components
Search	Alphabet[11]	Google Search	VLOSE	IE	332+	Paid and unpaid search results
	Microsoft[12]	Bing	VLOSE	IE	107	Paid and unpaid search results
Social media	Alphabet	YouTube	VLOP	IE	401+	Videos, sound, photos & text
	Meta[13]	Facebook	VLOP	IE	255	Videos, sound, photos & text
	Meta	Instagram	VLOP	IE	250	Videos, sound, photos & text
	Bytedance[14]	TikTok	VLOP	IE	125	Videos, sound, photos & text
	Microsoft	LinkedIn	VLOP	IE	122	Videos, sound, photos & text
	Snap[15]	Snapchat	VLOP	?	96+	Videos, sound, photos & text
	Pinterest[16]	Pinterest	VLOP	?	n/a	Videos, sound, photos & text
	Twitter[17]	Twitter	VLOP	?	100+	Videos, sound, photos & text
App stores	Alphabet	Google App Store	VLOP	IE	274+	Mobile apps
	Apple[18]	Apple App Store	VLOP	IE	n/a	Mobile apps
Wiki	Wikimedia[19]	Wikipedia	VLOP	?	151+	Mostly text and photos
Markets	Amazon[20]	Amazon Marketplace	VLOP	LX	n/a	Sellers' offerings & users' reviews
	Alphabet	Google Shopping	VLOP	IE	74+	Sellers' offerings & users' reviews
	Alibaba[21]	AliExpress	VLOP	?	n/a	Sellers' offerings & users' reviews
	Booking.com[22]	Booking.com	VLOP	NL	n/a	Sellers' offerings & users' reviews
Maps	Alphabet	Google Maps	VLOP	IE	278+	Shop profiles, reviews, etc.

Husovec, Martin, The DSA's Scope Briefly Explained (February 20, 2023). Available at SSRN: https://ssrn.com/abstract=4365029

LSE Department of Law

Risk Management Dialogue

- Regulatory dialogue put in place due to the opacity of the ecosystem & information asymmetry
- The regulator has no clear idea of risks, or contributing factors, and is in dark about solutions
- Forces providers to think about this, let themselves be reviewed by others (auditors, researchers, field NGOs), and then the regulator forms an opinion

LSE Department of Law

VLOP's risk management: Article 34(1)

Providers of very large online platforms and of very large online search engines shall diligently identify, analyse and assess any systemic risks in the Union stemming from the design or functioning of their service and its related systems, including algorithmic systems, or from the use made of their services. This risk assessment shall be specific to their services and proportionate to the systemic risks, taking into consideration their severity and probability, and **shall include** the following systemic risks: (..)

LSE Department of Law

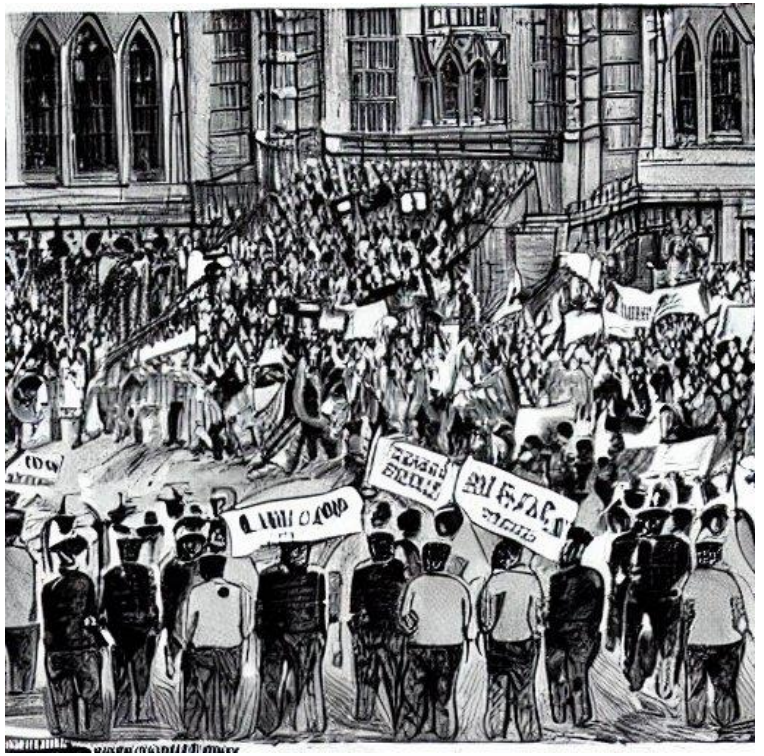
VLOP's risk mitigation

Risk Areas & Categories	Recommender systems	Content moderation	Terms and conditions	Advertising	Data practices	Other areas
Illegal content (Art 34(1)(a))	*Examples:* (a) terrorist content; (b) child sexual abuse; (c) illegal hate speech; (d) intellectual property infringements; (e) defamation; (f) sale of unsafe products; (g) cyberstalking or grooming; or (h) *any other* areas of illegal content or behaviour.					
Fundamental rights (Art 34(1)(b))	*Examples:*[97] (a) human dignity; (b) freedom of expression and information, including media freedom and pluralism; (c) right to private life; (d) data protection; (e) right to non-discrimination; (f) rights of the child; (g) consumer protection; or (h) *any other* fundamental rights.					
Public security and elections (Art 34(1)(c))	*Exhaustive subcategories:*[98] (a) civic discourse; (b) electoral process; and (c) public security.					
Health and well-being (Art 34(1)(d))	*Exhaustive subcategories:*[99] (a) gender-based violence; (b) public health; (c) rights of Minors; (d) physical well-being; and (e) mental well-being.					

LSE Department of Law

Metaphor: safety regulation of public protests

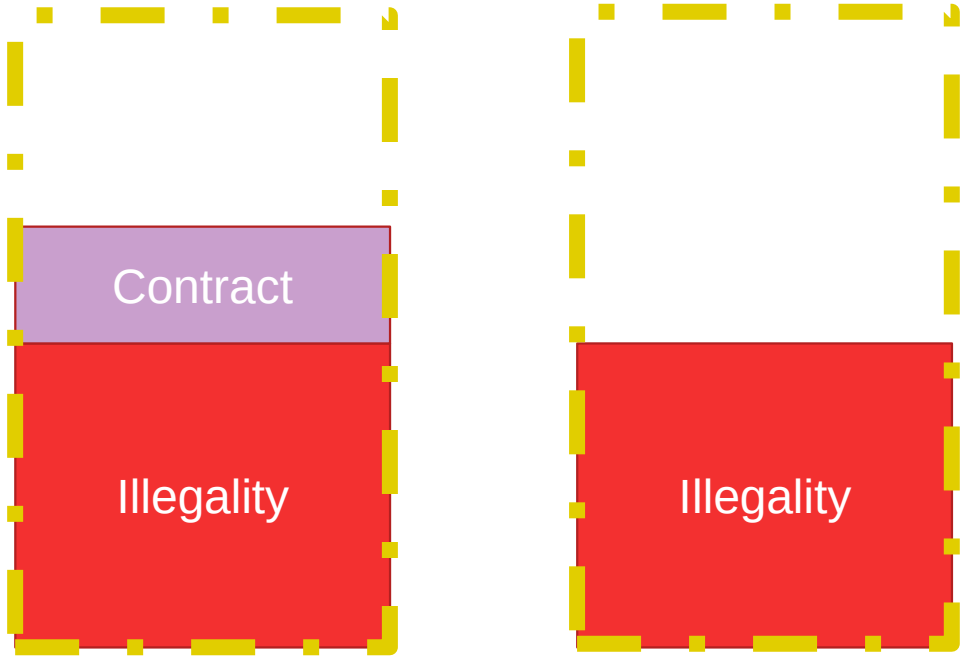
The authorities can (and do) partly restrict how and when protest activities take place (eg streets, hours, or use of amplification tools), and take measures to prevent harm to protesters or others (eg boost police presence), but cannot select speakers, or dictate the content



LSE Department of Law

Thus

- Two social media companies A & B set their baseline of rules [eg disinformation] and mitigate the risks on that basis
- Since parliaments permit both models of rules, they also permit two mixtures; risk mitigation cannot negate the existence of such choice



LSE Department of Law

Risk Mitigation Measures

Type of content	Priority by type of intervention	Examples
1. Illegal content (e.g., hate speech, terrorist content, copyright infringement)	1. Content removal	removal of content
	2. Visibility restrictions	age-gating of content or recommendations
	3. Nudges and incentives	demonetization of content or de-ranking borderline content
	4. Empowerment	flagging systems; rating systems; help lines; information
2. Legal content (e.g., disinformation, sensitive content, nude content, gambling content)	1. Empowerment	choice on recommendations; parental consent; rating systems; hotlines; information; suggestion tools;
	2. Nudges and incentives	default on recommendations; costly super-sharing; verification; pre-publication notices; parental consent; demonetization of content
	3. Visibility restrictions	age-gating of content or recommendations; de-ranking
	4. Content removal	removal of content

LSE Department of Law

Want to know more?

husovec.eu/DSA

LSE Short Course on the EU Digital Services Act