LETS LAB

LAW, EMERGING TECH & SCIENCE

Content Regulation & the Digital Platform Economy:

Combating the "Infodemic"

Dr Argyro Karanasiou
Dr Aysem Diker Vanberg
Mr Charalampos (Harry) Kliaris

30/09/2020

—

School of Law & Criminology

—

University of Greenwich, London
United Kingdom

LETS Lab (Law, Emerging Tech & Science)

The Law, Emerging Tech & Science (LETS) Lab, is an interdisciplinary research cluster at the University of Greenwich that aims at providing a collaborative hub for scholars sharing expertise or research interests in digital aspects of a wider socio-legal substratum.

The LETS Lab brings together academics and students from various disciplines and backgrounds working on tech related aspects in their fields and provides a strong network that facilitates and supports synergies in future research ventures.

Table of contents

1. The platform economy and content regulation

The regulation of online content and the potential liability of internet intermediaries has been one of the earliest legal conundrums for information technology law. The initial analogies drawn with the roles and duties of newspaper editors or those of common carriers providing services akin to post/telephony, have been unable to provide convincing answers to the puzzling matter of online content regulation. The past two decades the ecosystem of online intermediaries has changed significantly: new types of intermediaries (e.g. search engines, news/price aggregators, cloud service providers etc.) with a broad scope of activities and moderation techniques (predictive moderation, ex post moderation, reactive moderation, user-only moderation, community driven moderation) have rendered the current legislative framework on intermediary liability outdated (namely, the EU E-Commerce Directive - adopted in 2000, and the US CDA/DMCA adopted in 1996 and 1998 respectively).

There is no doubt that online platforms are key for the digital economy and have always played an important part in boosting e-commerce since the mid-90s. Most importantly though, they act as gatekeepers determining data flows and access to information. This explains well the concept of limited intermediary liability (e.g. s230 CDA and art 14 E-Commerce Directive), which has been credited by many as the driving force behind the growth in digital economy. Twenty years later, we are now witnessing a change of attitude towards intermediaries: a shift from liability to a duty of care (Mac Síthigh, 2020). At the same time, digital platforms have gained significant powers over digital communications and have adopted data driven business models that monetise user generated content and personal data. This in turn translates to 1.7 MB of data created every second for every person on earth (DOMO report, 6th edition); data that can often contain racist/xenophobic/misogynistic/extreme hate speech, fake news, disinformation and other types of computational propaganda.

The phenomenon of such problematic instances of speech has been particularly prevalent after the new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was declared a pandemic. The WHO Director-General expressed his concern that ‘We’re not just fighting an epidemic; we’re fighting an infodemic. Fake news spreads faster and more easily than this virus and is just as dangerous’ (World Health Organization [WHO], 2020). Just as a pandemic, infodemic is a global phenomenon requiring global consensus on addressing the threats posed for free speech and democratic discourse (Radu 2020). However, the fragmented legal landscape of national laws and dated EU/US provisions paired with a patchwork of self-regulatory attempts from social media, are far from a global response to combat this infodemic. This report discusses the regulatory reforms in intermediary liability (EU/UK), explores self-regulation as a viable alternative and suggests recommendations towards a co-regulatory principle-based approach.

2. The current regulatory landscape: Towards stricter liability/duty of care

As digital communications are constantly evolving, public expectations for a safer and accountable technology have intensified. The current trend in the EU (and subsequently in the UK) is that platforms must be subject to tighter regulations and thereby assume greater responsibility. As a result, the UK has issued the Online Harms White Paper and the EU has put forward the Digital Services Act. These regulatory efforts follow a different approach but have a common denominator: they aim to move the focus from conditional intermediary liability to direct intermediary liability for activities involving illegal and harmful content.

2.1. The Digital Services Act (DSA)

The e-Commerce Directive has offered exemptions to liability for three types of online services – those that share (Art 12), store (Art 13) or host (Art 14) information under specific circumstances. Under Art 15, Member States must not impose a general obligation on Internet intermediaries to actively monitor information indicating illegal activity. The rationale of the Directive is that liability should not be assumed if the service provider has no knowledge of illegal content. The approach taken in the e-Commerce Directive protects freedom of expression as it enables online service providers to set their own rules for content moderation without fear of government intervention. Be that as it may, whether these rules are consistent with human right standards (such as the freedom of expression) is a different question.

In June 2020, the European Commission published its consultation paper on the proposed Digital Services Act (DSA). The DSA will provide the regulatory framework for digital services and online platforms in the EU and will replace the e-Commerce Directive. The Act aims to propose clear rules providing the responsibilities of digital services to address the risks faced by their users and to protect their rights (EC, 2020 Consultation). It also aims to propose ex ante rules to regulate online platforms that act as gatekeepers (EC, Shaping Europe's Digital Future, 2020).

2.1.1. Scope of compliance

The consultation paper clarifies that the DSA aims to go beyond regulating unlawful content as it will involve activities that are not necessarily illegal but may cause harm to users. That said, the action that online platforms are required to take against such content still remains unclear (EP, Collection of Studies for the IMCO Committee, 2020). As it stands, it is difficult to determine what risks may arise to one's freedoms and rights: If the DSA only requires platforms to be transparent with their terms of use and moderation tactics regarding what content is and is not allowed, then the regulation should be welcomed. If, however, the DSA requires platforms to use ex ante mechanisms of upload filtering and take downs, this could present a conflict with numerous, existing fundamental protections such as Art 10 ECHR (freedom of expression), and Art 8 ECHR (privacy). The current draft though lacks clarity on such matters.

2.1.2. Limited regard over emerging tech

It is apparent that the DSA demonstrates a lack of awareness or interest in modern technologies like the Blockchain and AI. The paper states that these technologies will play an essential role in the process of automated filtering and should be considered for better 'transparency and accountability'. If the Act is promoting transparency it should foster transparency itself without vague proposals.

Furthermore, the Act provides that AI technologies used for the purpose of general monitoring are acceptable. However, in the case of *Scarlet v SABAM* it was held that general monitoring is prohibited, contradicting the DSA's recommended use of AI technologies for general monitoring.

2.1.3. The need for new governance structures

As mentioned above, the digital sphere is constantly evolving. For this reason, digital services require different governance structures that stimulate innovation and at the same time protect individuals. The DSA has not succeeded in solving this problem, which requires developing a new type of governance structure flexible and future proof.

2.2. The UK Online Harms White Paper

In 2019 the UK Department for Digital, Culture Media and Sport and the Home Office published the Online Harms White Paper (OHWP). This sets out the government's plan to give powers to a regulator to oversee and enforce a regulatory framework (DCMS, 2019). The OHWP introduces a statutory duty of care which requires companies to take active steps and a greater responsibility to tackle online harms and keep users safe. The paper states that this duty of care applies to platforms that involve the sharing of user-generated content or allow users to interact with each other online. However, there is no explanation as to how this 'duty of care' will apply to claimants. For example, there is no mention that users can make a claim against a company that acted negligently and failed to satisfy the online duty of care, as would be the case under the traditional duty of care. As it stands now, duty of care in the OHWP is nothing more than a confusing label and is not suitable in its current form, rather like trying to fit a square peg in a round hole. Any proposals under the OHWP need to be clear and logical to all providers of online content, from small enterprises to the huge online conglomerates since new regulation is likely to have some financial impact in order to be successful. For example, in complying with this new regulation online companies will likely have to bear some added costs, both in the implementation of the new rules such as additional resources and software and also in potential loss of some profitability as a result of any limitations imposed on content.

2.2.1. Codes of practice

In the OHWP the UK government has laid down codes of practice to implement the suggested duty of care. This has created some issues that seem to arise from the way the codes were drafted. The government stated that there will not be a code of practice for each category of harm since this would pose a very unreasonable legal burden on relevant companies. As a result, the only individual codes of practice being proposed will be in relation to terrorists and child sexual exploitation. However, the government has again followed a pattern in the OHWP by failing to clarify what any proposed codes actually will entail, leaving many companies uncertain of how they can ensure compliance with these new codes of practice.

In moving forward, the OHWP adopts a proactive approach towards the various forms of content, not just those which relate to serious harm such as child sexual exploitation and terrorism which are covered by individual codes of practice. This may be somewhat worrying since it could mean implementing AI enabled mechanisms such as upload filtering, general monitoring and take downs. This could present a conflict with numerous, existing fundamental protections, as noted elsewhere in this report.

2.2.2. The regulator and the appointment of powers

After the OHWP’s initial release a consultation commenced on 8 April 2019 and lasted until 1 July 2019. This received well over 2,400 responses from tech companies, academics and charities. In February 2020, the government published a response to the consultation (Article 19, 2019), which clarified that Ofcom would be appointed as the independent regulator in charge of overseeing the online systems and processes of the relevant companies. In this task, Ofcom would be subject to s.149 of the Equality Act which provides that ‘a public authority must, in the exercise of its functions, have due regard to the need to eliminate discrimination, harassment, victimisation and any other conduct that is prohibited under this Act’. Be that as it may, the response did not define the sanctioning powers that would be available to Ofcom.

2.2.3. Lack of urgency

Further progress with the OHWP has been delayed due to Covid-19 issues occupying all available bandwidth. The DCMS has indicated that the legislation will not be available until 2024 (DCMS 2020). However, the Covid-19 pandemic has demonstrated well that the material posted in social media platforms can affect the very fabric of our society and should be dealt with urgently.

3. Self-regulation: An attractive alternative?

3.1. The need to moderate content on social media platforms

The growing concern pertaining to the dissemination of illegal, harmful and misleading content ("fake news") on online media platforms has not been adequately addressed: an outdated set of EU laws and a patchwork of national legislative responses suggests that we still far from a fully-fledged regulatory response to combat an infodemic. Could self-regulation be an attractive alternative, learning from similar examples in the newspaper/media industry?

A self-regulatory framework based on voluntary compliance and principles that reflect a common understanding of the values and ethics that inform practises and professional conduct of digital platforms (Article 19, 2018) has been put forth as an appropriate response to combating fake news.

Examples of self-regulation include setting industry standards, agreeing to best practices and establishing professional codes of ethics as well as setting up independent bodies (with or without disciplinary powers) to make decisions in relation to content. As of 2019, over forty national laws have entered in to force to regulate the dissemination of harmful/illegal content (Funke et al., 2019).

3.2. Some examples of self - regulation

Several self-moderation initiatives have been put forward by social media platforms. These include Facebook's Oversight Board, the Bluesky project by Twitter and Tiktok's Content Advisory Council. So far, Facebook's Oversight Board seems to be the most developed one.

Facebook's Oversight Board is an independent body established by Facebook. The Board hears appeals about the most difficult and important content decisions Facebook makes and explains the reasons for its decisions. The Oversight Board announced its first members in May 2020 and also released bylaws (Facebook, 2020). The bylaws are a significant step as they show how the Board will work and its limitations. The Board will obtain from Facebook the information necessary to decide a case and may receive written statements from the other parties such as the content creator and the complainant/s (Oversight Board Charter, art. 3). It may also gather information from experts or otherwise necessary to provide context (Oversight Board Charter, art. 3). The Board should review the cases on the basis of Facebook's content policies and values while taking into account the human rights standards that protect freedom of expression (Oversight Board Charter, art. 2). Decisions taken by the Oversight Board are binding on Facebook (Oversight Board Charter, art. 4). Finally, the decisions must be made public and clearly justified (Oversight Board Charter, art. 3).

Some commentators have raised some concerns particularly as to the composition and operation of the Oversight Board. Article 19 is concerned that the global level at which the Board operates will make it difficult to understand local contexts (social, political, cultural, historical, linguistic, etc.) and their complexity. (Article 19(1b), 2019). Latonero also points out that the selection of the Oversight Board members by Facebook may undermine the panel members independence, and as the Board's decisions will be based on Facebook's values and content policies, may not be necessarily be in line with international human rights standards. (Latonero, 2020). The Board can deal with a variety of controversial take down decisions. However, as noted by Doucek, one of the most controversial content moderation decisions by Facebook are decisions that concerns leaving the content up, and not taking it down. (Doucek, 2020) In this regard, the bylaws limit the ambit of the Oversight Board.

On December 2019 Twitter launched their BlueSky project. Accordingly, Twitter will fund a small independent team of up to five experts comprising open source architects, engineers and designers to develop an open and decentralised standard for social media. As opposed to the below mentioned initiative of Facebook which consist of reviewing/ assessing online platform's content moderation practices, the Twitter initiative is markedly different. Arguably, this decentralised approach can reduce criticism pertaining to platforms' content moderation practices, as the control of content would no longer be under the monopoly of one dominant company which could also offset issues associated with dominance of major platforms. (European Parliament, 2020) Twitter's project may also be beneficial to tackle the dominance and influence of the major platforms on online expression (Article 19, 2020)

Nevertheless, Twitter's BlueSky project was not without critics. Commentators have suggested that previous efforts to take on proprietary social networks with open standards have failed to reach mainstream success as evidenced in the case of OpenSocial and Jabber. (Fried, 2019) Furthermore, Mastodon, a free open source microblogging service with more than 1 million users across several different serves already exists, which makes Twitter's motives questionable. (Robertson, 2019) A developer suggested, Twitter could have helped existing projects by donating to them instead of competing with them (Robertson, 2019). At time of writing, there is not more information on Twitter's BlueSky project, and it remains to be seen whether the project will be a failure or a success.

Finally, on March 2020, the short video app Tiktok, has also formed a group of outside experts to advise on its content moderation policies to address concerns pertaining to data security and content moderation (Sheng, 2020). Its content advisory council is expected to provide plain and straightforward views" and advice around its content-moderation policies and practices. The seven-member committee were expected to meet at the end of March 2020 to discuss topics around platform integrity, including policies against misinformation and election interference. At the time of writing, no further details have been found on the TikTok's Advisory Council, nor is it not clear how progressed this proposal is.

3.3. Advantages and disadvantages of self-regulation

Overall, self- regulation offers the following advantages:

- It is flexible, quick and responsive. As noted by Douek, semi-independent self-regulatory oversight mechanisms such as the Oversight Board by Facebook offer significant advantages such as flexibility and speedy response which is required to tackle illegal and harmful content. (Douek, 2019) For instance, YouTube has stated that most videos that violate their terms and conditions are flagged and removed within an hour of dissemination (Foxman and Wolf, 2013).
- Avoids politicising tech companies as they are not instructed by government officials and remain independent content moderators. (Samples, 2019).
- Platforms are more likely to adhere to standards created by them as opposed to standards dictated to them by governments.

3.4. Shortcomings of self -regulation

Self- regulation of content moderation entails the following disadvantages:

- It does not necessarily meet high standards of due process and transparency (Douek, 2019)
- It may lack teeth and legitimacy as it is introduced by private actors as opposed to state actors
- It can lead to biased decisions and lead to excessive removal of legitimate content (Urban J et al. (2017); Elkin-Korel & Perel (2020).

Arguably, the current content moderation regime, particularly self-regulation, does not represent an effective approach, as good regulation focuses on achieving outcomes rather than technical compliance (Black, 2008). In this regard, if the intention is to reduce the amount of harmful content, merely deleting content will only offer a short-term solution and will not suffice to change the overall culture of the platform (Common, 2020). More importantly, as it has been evidenced in Facebook Oversight Board incentive, self-regulatory approaches adopted/ pioneered by one dominant platform often offer little, if any, value or insight to another platform so it is not very helpful in finding a collective solution to common problems faced by different platforms.

3.5. Moving forward: Co-regulatory polycentric models

Due to the shortcomings of both regulation and self-regulation, a co-regulatory approach could be a better way to deal with harmful, illegal and misleading content. In this regard, co-regulation can be an effective way to moderate content, as it will mean that platforms can develop – individually or collectively – mechanisms to regulate their own users, which will also need to approved by democratically legitimate state regulators or legislatures, who also monitor their effectiveness (Marsden et al., 2020). Below we offer some recommendations in this respect.

4. Recommendations

A co-regulatory framework with a clear scope on intermediary liability and effective oversight of meta-regulation could provide a much-needed answer to the pressing need to protect free speech in an informational digital chaos. The EU Digital Services Act might be a step towards the right direction for a harmonized approach, however it would not be enough on its own to provide for an effective enforcement mechanism. We suggest that liability legislation should complement self-regulation and provide for legal certainty, transparency, accountability, and oversight. This synergy should be further supported with strategic choices in architecture and market structures to ensure free (mediated) speech. To this end, we adopt Lessig's classic formulation that law (intermediary liability laws), digital architecture (decentralized structures), market mechanisms (competition – media laws for diversity) and societal/community norms (self-regulation in an accountable manner), all are external forces that jointly regulate (or have the potential to regulate) online activity.

4.1. Addressing definitional certainty & scope clarity in intermediary liability laws

The ongoing regulatory shift towards stricter content regulation of harmful content addresses an overlooked issue in intermediary liability, however it warrants closer scrutiny with regards to free speech (art 10 ECHR). Nowadays, digital platforms and social networks act as information gatekeepers, facilitating thereby data flows and enabling users to receive and impart information. The argument that they constitute "mere conduits" (Tamiz v Google Inc Google UK Ltd [2012] EWHC 449 (QB) 02 March 2012) and lack the ability to monitor content authored by users with no contractual nexus to them (LICRA et UEJF v. Yahoo! Inc. et Yahoo Fr., T.G.I. Paris, May 22, 2000) is no longer convincing: Several instances of social media platforms using covert AI enabled manipulative tactics of public discourse (EC – Content and Technology DG for Communication Networks, 2018), either for political (Cadwalladr & Graham-Harrison, 2018) or marketing purposes (Kramer et al, 2014) demonstrate well that digital platforms play a key role in modern democracies (Karanasiou, 2019). This however has often placed an unreasonable burden on digital platforms to monitor illegal (D*elfi AS v. Estonia* (2015) ECtHR 64669/09) or equivalent material (Eva Glawischnig-Piesczek v Facebook Ireland Limited C-18/18). As a result, there is an increasing trend towards imposing hard time limits for "expeditious" take down of illegal content, which are established without any clear definition of what constitutes ”harmful content", "hate speech", "fake news" etc. (Edwards, 2018). The current EU framework (E-Commerce Directive – art 14 ECD), addresses only intermediary liability for "illegal content", whereas other types of potentially harmful content (e.g. disinformation, misinformation) appear to be a gray area, mostly subject to voluntary self-regulation and to non-legally binding Codes of Practice (EU Code of Practice on Disinformation). The recent trend to broaden the scope of content moderation (meta-regulation) together with legal uncertainty over "harmful content", poses significant challenges for free speech as it may often result to over-blocking material that may be deemed problematic. A clear definition of what constitutes "harmful/potentially harmful" content as well as a narrow and clear scope of required moderation (take down orders) is essential so that such types of speech are limited in a proportionate manner and as necessary in a democratic society (art 10 para 2 ECHR).

4.2. Automated content moderation & algorithmic accountability

The over-reliance on algorithmic/technical means of moderating content has been a weapon of choice for most digital platforms. Amidst the Covid-19 crisis, automated content moderation became a much-preferred option for digital platforms, whose employees were in lockdown and thus unable to use office facilities for moderation. We have already argued (co-signees to the open letter for content preservation data, 2020) that outsourcing moderation to AI systems poses significant risks to free speech, open research, democracy, privacy, and -in the Covid-19 infodemic context- health. Algorithmically driven technical solutions that leave human operators out of the loop do not provide transparent and context aware answers (Karanasiou and Pinotsis, 2017). Moreover, due to data retention laws, auditing of such decisions can be difficult. Such practises are used widely, however the law is still lagging behind. We recommend that provisions should be made for AI content moderation to allow for (i) human oversight of fully automated moderation (ii) performing due diligence to notify users on automated filtering/removal of content (iii) a right not to be subject to an automated content moderation, similar to art 22 GDPR (iv) a clear and effective complaint and redress mechanism.

4.3. Free speech by design – Embracing decentralised structures

Architectural choices play a significant role in content regulation: a fact, which often goes unnoticed with under-developed relevant policies. Whilst it is important to acknowledge the power digital platforms have gained as facilitators of communication and to entrust them with significant responsibilities as enablers of digital speech, it is equally important to explore the infrastructures of the online ecosystem that (i) often encourage disinformation and (ii) are responsible for concentration of moderation executed from a handful of large tech companies. Centralised structures are often exploited to allow for "lock-ins" and to limit pluralism, posing thereby significant challenges for online democratic discourse. The rapid centralisation of digital platforms and the high concentration in online intermediaries have made it impossible for the state to guarantee constitutional protection for the right to free speech online. Even though the legislative response has been to involve intermediaries in content moderation, the issue of free speech by proxy in a highly concentrated ecosystem has not been addressed. The distributed decision-making approaches to content moderation from open source "federated" platforms, such as Mastodon, present an interesting alternative that could inform the current debate focussing solely on giant platforms with centralised structures (Masnick, 2019). Twitter's Bluesky initiative might be a good step towards this direction, but it is important that such design choices are included in the DSA ("free speech by design" as a concept similar to "privacy by design"). Taking this point a step further, we posit that maintaining distributed architectures should be perceived not only as a useful means of preserving the Internet's sustainability, but also as a constitutional imperative for the policymaker (Karanasiou, 2016).

4.4. Principle based approach and oversight – Due process

It has often been suggested that digital platforms lack a legitimacy to regulate data flows and impose free speech limitations without a court order (Manila Principles on Intermediary Liability, 2015). This is further exacerbated by the lack of due process and oversight mechanisms in self-regulated meta moderation. Woods and Perrin (Woods and Perrin, 2019) highlight the need for an independent regulator (no links to industry/government), whose role will be to balance between industry and public interests. In the UK, this could potentially mean extending the powers of existing regulatory authorities, such as Advertising Standards Authority (ASA), the British Board of Film Classification (BBFC), the Information Commissioners Office (ICO) the HSE or OFCOM. That said, we side with the view that this regulatory authority should adopt a principle-based approach (rather than enforcing a duty of care), allowing intermediaries flexibility whilst maintaining accountability (Murray 2019). Such an approach would boost a co-regulatory model, as it would suffice to provide intermediaries with some much-needed legitimacy (Black 2008) and guidance, which in turn translates to a heightened level of responsibility and a greater degree of transparency.

4.5. Digital market structural changes

Finally, we suggest that regard is also given to the commercial setting within which harmful speech (esp. fake news) is regulated online. Social media operate on business models that thrive on data flows, which are often triggered as responses to fake news; clicks and shares of misinformation generate traffic and advertising revenues for intermediaries. In such a commercial ecosystem, a hands-off approach to fake news amounts in fact to "a normative choice, one that benefits powerful, moneyed interests who crave power no matter the cost" (Waldman, 20218). It is therefore important to address content regulation from various vantage points: digital platforms are both communicatory platforms as well as media businesses, requiring a holistic approach with principles from human rights, competition, and media law (Drexl, 2019).

Bibliography

Article 19 (2018, March). Self-regulation and hate speech on social media platforms.https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-%E2%80%98hate-speech%E2%80%99-on-social-media-platforms_March2018.pdf

Article 19 (2019, September). Facebook: New oversight board is not sufficient to safeguard freedom of expression online. https://www.article19.org/resources/facebook-new-oversight-board-is-not-sufficient-to-safeguard-freedom-of-expression-online/.

Article 19 (2019, July) Response to the Consultations on the White Paper on Online Harms

Article 19 (2020, March). Why decentralisation of content moderation might be the best way to protect freedom of expression online. https://www.article19.org/resources/why-decentralisa tion-of-content-moderation-might-be-the-best-way-to-protect-freedom-ofexpression-online/

Black, J. (2008). Constructing and Contesting Legitimacy and Accountability in Polycentric Regulatory Regimes. Regulation and Governance 2: 137–164.

Cadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The guardian*, *17*, 22.

Common, MacKenzie F. (2020). Fear the Reaper: how content moderation rules are enforced on social media, International Review of Law, Computers & Technology, 34:2, 126-152.

Crews Jr, C. (2020). The Case against Social Media Content Regulation: Reaffirming Congress' Duty to Protect Online Bias,"Harmful Content," and Dissident Speech from the Administrative State. *Harmful Content," and Dissident Speech from the Administrative State (June 28, 2020)*.

Department for Digital, Culture and Sport and Home Office (2019) Online Harms White Paper

Department for Digital, Culture and Sport and Home Office (2020, February), Government minded to appoint Ofcom as online harms regulator

Doucek, E (2019). Verified Accountability: Self-Regulation of Content Moderation as am Answer to Special Problems of Speech Regulation' Aegis Series Paper No. 1903. https://assets.documentcloud.org/documents/6419386/Evelyn-Douek-Hoover-Aegis-Paper-Verified.pdf

Doucek E (2020, January 28). Facebook's Oversight Board Bylaws: For Once Moving Slowly. https://www.lawfareblog.com/facebooks-oversight-board-bylaws-once-moving-slowly

Drexl, J. (2019). Economic Efficiency vs. Democracy: On the Potential Role of Competition Policy in Regulating Digital Markets in Times of Post-Truth Politics. In *Reconciling Efficiency and Equity-A Global Challenge for Competition Policy* (pp. 242-267). Cambridge University Press.

Elkin-Koren, N & Perel M (2020). Guarding the Guardians: Content Moderation by Online Intermediaries and the Rule of Law: Edited by Giancarlo Frosio Oxford Handbook of Online Intermediary Liability.

Euronews. (2019, January 10). How can Europe Tackle Fake News in the Digital Age? https://www.euronews.com/2019/01/09/how-can-europe-tackle-fake-news-in-the-digital-age

European Commission (2020, June), Consultation on the responsibilities for digital services and on ex ante instruments for gatekeepers

European Commission (2020, June), Shaping Europe's digital future – The Digital Services Act package

European Parliament. (2020, June). Online Platform Moderation of Illegal Content Online Law, Practices and Options for Reform. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652718/IPOL_STU(2020)652718_EN.pdf

European Parliament (2020, May), Collection of Studies for the IMCO Committee – Digital Services Act

Facebook. (2019) Facebook Oversight Board Structure (2019) https://about.fb.com/news/2019/09/oversight-board-structure/

Facebook. (2019) Oversight Board Charter. https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf#

Foxman, A. & Wolf, C. (2013). Viral Hate: Containing Its Spread On The Internet. New York: Palgrave MacMillan, 107.

Fried, I (2019). Twitter's bid for a social network standard draws skepticism. available https://www.axios.com/twitter-social-network-standard-open-source-bluesky-e05a9a38-a995-4583-8955-7ccc16cac802.html

Frosio, G. F. (2017). Reforming intermediary liability in the platform economy: a European digital single market strategy. *Nw. UL Rev. Online*, *112*, 18.

Funke, D & Flamini, D (2019). A guide to misinformation actions around the world' Poyneter Institute. https://www.poynter.org/ifcn/anti-misinformation-actions/

Gillespie, T. (2018). Platforms are not intermediaries. *Georgetown Law Technology Review*, *2*(2), 198-216.

Karanasiou, A. P. (2016). Law encoded: Towards a free speech policy model based on decentralized architectures. *First Monday*, *21*(12).

Karanasiou, A. P., & Pinotsis, D. A. (2017). A study into the layers of automated decision-making: emergent normative and legal aspects of deep learning. *International Review of Law, Computers & Technology*, *31*(2), 170-187.

Karanasiou, A. (2019). Written evidence submitted to the Lord Select Committee on Democracy and Digital Technology. House of Lords, UK Parliament 2019.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788-8790.

Latonero, M. (2020, January 29). Can Facebook's Oversight Board Win People's Trust? Harvard Business Review. https://carrcenter.hks.harvard.edu/publications/can-facebook%E2%80%99s-oversight-board-win-people%E2%80%99s-trust.

Mac Síthigh, D. (2020). The road to responsibilities: new attitudes towards Internet intermediaries. *Information & Communications Technology Law*, *29*(1), 1-21.

Marsden, Ch, Meyer, Tr & Brown, I (2020) Platform values and democratic elections: How can the law regulate digital disinformation. Computer Science and Security Review 36.

Macnick, M. (2019, July 16). Gab, Mastodon And The Challenges Of Content Moderation On A More Distributed Social Network. Techdirt. https://www.techdirt.com/articles/20190715/00244442587/gab-mastodon-challenges-content-moderation-more-distributed-social-network.shtml

Open Letter to Social Media and Content Sharing Platforms for Data Preservation (2020) – https://cdt.org/wp-content/uploads/2020/04/COVID-19-content-moderation-research-letter-English-PDF.pdf

Radu, R. (2020). Covid19: Fighting the 'Infodemic': Legal Responses to COVID-19 Disinformation. *Social Media+ Society*, *6*(3).

Robertson, A (2019, December12) 'Twitter wants to decentralize, but decentralized social network creators don't trust it' https://www.theverge.com/2019/12/12/21012553/twitter-bluesky-decentralized-social-network-developers-reaction-mastodon-activitypub

Samples, J. (2019). Why the government should not regulate content moderation of social media. *Cato Institute Policy Analysis*, (865).

Sheng, W (2020, March 19). TikTok promises to let US experts guide its content moderation. https://technode.com/2020/03/19/tiktok-promises-to-let-us-experts-guide-its-content-moderation/'

UK Parliament (June 2020), Democracy under threat from 'pandemic of misinformation' online – Lords Democracy and Digital Technologies Committee

Urban, J & Karaganis, J & Schofield, B. (2017). Notice and Takedown: Online Provider and Rightsholder Accounts of Everyday Practice64 J. Copyright Soc'y 371.

Waldman, A. (2018). The marketplace of fake news. University of Pennsylvania Journal of Constitutional Law, 20(4), 845-870.

Woods, L & Perrin, W. (2019) Internet harm reduction: an updated proposal, Carnegie UK Trust.

AUTHORS

Dr Argyro Karanasiou* is the Director of LETS Lab and a Senior Lecturer in Law at the University of Greenwich. She also holds visiting affiliations with Yale Law School (ISP), NYU Law (ILI), Harvard Law (CopyX), Complutense Madrid (ITC). She has contributed invited expert insights on a number of occasions, most notably for the Equality and Human Rights Commission (AI in recruitment), the Chatham House, the US Air Force (AI & Augmented Cognition), the Royal Society (Machine Learning), and the Electronic Frontiers Foundation, and has served as a contracted consultant for the Council of Europe (regional South Eastern Europe expert in Media), as an OSCE expert for Online Media, and as a registered European Commission expert for European Research & Innovation.

Dr Aysem Diker Vanberg** is a Senior Lecturer in Law at the University of Greenwich. She holds an LLB (University of Ankara), an LLM (University of Bremen) and a PhD in Competition Law (University of Essex) and has taught at the Universities of Essex and Anglia Ruskin. She has also worked as a lead in-house counsel for multinational companies including MAN Nutzfahzeuge AG and Cimpor Cimentos de Portugal. Dr Diker Vanberg has expertise on EU competition laws regulating digital platforms and her main research interests lie in data protection law, EU and UK competition law for online platforms as well as the interplay between data protection law and competition law.

Mr Charalampos (Harry) Kliaris*** is an affiliate researcher at LETS Lab with interests in content regulation and digital surveillance. He holds an LLB, currently studying for an LPC, and has worked for Royal Tunbridge Wells Citizens Advice Bureau, Clements Solicitors (Ipswich, UK), Incontext Solutions (Teddington, UK) and Harbor Shipping & Trading (Piraeus, Greece).

*Parts 1 and 4
**Part 3
*** Part 2