

Dorina Gyetván

Users v Facebook

A system designed to undermine users

Thesis Book

Supervisor:

Gergely Gosztonyi

Professor of Law

Doctoral School of Law

Eötvös Loránd University (ELTE)

Budapest, 2026

The exercise of freedom of expression has increasingly drifted from traditional public spheres such as town squares to online platforms mainly social media platforms in the past two decades: these platforms are generally available to everyone and anyone with internet access. Online platforms transformed the landscape of public discourse, political campaigns, by providing access to information in real time, across borders without the direct presence of government censorship, however, users are being subjected to multiple actors' control simultaneously and being limited by platform ToSs. As people could become their own broadcasters, connect with others worldwide,¹ the proliferation of UGC poses an immense challenge for conventional methods of content control, which resulted in a flood of questions surrounding the extent to which intermediaries should be or can be held liable for users' comments or other third-party content.

Social media platforms' content moderation decisions, including content removals, account suspensions, and affecting the visibility of certain pieces of content directly affect the realization of freedom of expression² while generally lacking the substantive and procedural safeguards, and they happen usually behind the veil of opaque platform rules embedded in ToSs without adequate transparency that are all warranted in case of state interventions.

The dissertation is based on the assumption that the fundamental right to freedom of expression of individuals as enshrined in human rights treaties and interpreted by various human rights tribunals, may only be restricted by adhering to the strict conditions of permitted restrictions and also assumes that when such human right is interfered with and restricted by private actors on a daily basis, individuals (users) should have the possibility to file for an effective remedy.

Both the Oversight Board and ACE aspire and claim to provide effective remedies for content moderation disputes as meaningful alternatives to long-established national, regional and other supranational redress mechanisms that traditionally require significant financial and human resources in addition to legal expertise that tends to deter most users from filing a claim in the first place.

The focus on the Oversight Board and ODS bodies contributes to evaluating their legitimacy, independence, and influence on freedom of expression. Moreover, it also affects broader issues,

¹ Balkin, Jack M.: Free Speech in the Algorithmic Society: Big Data, Private Governance, And New School Speech Regulation. UC Davis Law Review, 51/3 (2018) 1149-1210. 1151. <https://doi.org/10.2139/ssrn.3038939>

² Balkin, Jack M.: The Fiduciary Model of Privacy. Harvard Law Review, 134/1 (2020) 11-33.

especially concerning the interaction between self-regulation state-led legal systems in ensuring effective remedies for rights infringements.

Research aims and research questions

The aim of the present research is to present a comprehensive picture of the state of European intermediary regulation, the contribution of the Oversight Board and ACE to the long-established judicial, administrative remedy mechanisms of Member States, the EU and the CoE framework, with an emphasis on the aspects of these complementary novel tools that could be improved to better facilitate an effective, transparent and accessible remedy system for social media users that present a real alternative to judicial-administrative procedures.

For the purpose of these aspects, the present dissertation will focus on the following four research directions, some of them containing multiple research questions:

1) Regarding the main factors contributing to social media platforms becoming arbiters:

- a. How are social media platforms regulated in Europe and in the US, with a focus on platform liability rules for UGC and platforms' human rights responsibilities?
- b. What are the main factors that contribute to social media platforms, such as Facebook, moderating content at scale, and to social media being forced into the role of online governors?

2) Whether the Oversight Board is a legitimate remedy mechanism:

How does the Oversight Board compare to other entities, and to what extent does it fulfil the main requirements of an effective remedy mechanism? Could the entity be considered a legitimate actor?

3) How ACE, as an example for ODS bodies, compares to the Oversight Board?

How do ODS bodies, specifically ACE, compare to the Oversight Board appeals process in light of the lessons learnt from the Oversight Board's first five years of operations?

4) Proposals for systematic remedies

What procedural and legislative amendments could contribute to ODS bodies and the Oversight Board providing more effective remedies and how the network of these entities could be organized to better compensate the deficiencies of online content moderation?

Structure of the dissertation

In sum, the research is organized into eleven chapters, which is led by the introduction (Chapter **Error! Reference source not found.**) and consist of seven main chapters. The first part (Chapters **Error! Reference source not found.-Error! Reference source not found.**) constitutes the historical-legislative environment that focuses on the crucial issues of online freedom of expression and intermediary liability. The second part (chapters **Error! Reference source not found.-Error! Reference source not found.**) is fundamentally based on the contractual regulation of freedom of expression but evaluates the industry's contractual approach against a human rights perspective and additionally it also focuses on the most recent extra-judicial remedy mechanisms prevalent in the social media realm. Chapter 11 summarizes the relevant findings and attempts to provide answers to the research questions set out in Chapter **Error! Reference source not found.**, formulates some recommendations for the possible future improvement of alternative dispute resolution mechanisms in relation to social media platforms in a way that better equips users to seek redress.

Methodology

The research situates itself within an interdisciplinary field, critically exploring how legal, institutional, and normative developments can uphold freedom of expression while ensuring effective remedies for violations in the evolving landscape of social media service providers.

For the purpose of answering the research questions, the dissertation primarily employs historical-descriptive and analytical methodology, supplemented by a one-on-one style interview with the CEO of ACE, Thomas Hughes as a qualitative research method, conducted by the author via zoom. The research also explores the legal nature of the Oversight Board as it would have been incorporated in the EU, against the standards and elements that are clear from the legal provisions and case law of the CJEU and the ECtHR.

For the preparation of the present dissertation, besides assembling and analyzing the relevant legal provisions, I also had to gather two groups of datasets:

- 1) Information on the outcomes, types and other relevant data regarding the decisions of the Oversight Board: I have assembled available information from already published decisions, transparency reports (quarterly, semi-annual and yearly) and where information was unavailable, I tried to find those on the Facebook Transparency Center and publications, newsroom of the Oversight Board. Once all relevant data was collected, I used statistical analysis to interpret these findings and prepared tables, graphs and figures for the purpose of visualizing and describing such statistical analysis.
- 2) Information on the technical and operational workings of Facebook, the Oversight Board and ACE. I collected these data from
 - a. the ToS, Community Standards and other policies of Facebook,
 - b. the governing documents, LLC Agreement and Trust Agreement regarding the Oversight Board,
 - c. the Rules of Procedures and the first Transparency Report of ACE,
 - d. different online videos and informational materials of the relevant entities,
 - e. my personal inquiries and experiences I had as a long-time user of social media platforms, specifically Facebook, and
 - f. to fill in the gaps left by the first Transparency Report, the Rules of Procedures and limited information available on ACE, in order to help my understanding of ACE as an entity still in its infancy, I conducted an online interview with the CEO of ACE, Thomas Hughes.

Main research findings

1. Factors contributing to social media platforms becoming dispute settlers

Considering the conclusions and issues identified in this dissertation, I concluded that there are three distinct group of reasons contributing to online social media platforms over moderating content, violating users' freedom of expression and, generally speaking, driving social media platforms to proceed cautiously, to the detriment of freedom of expression: (a) overly vague

legislation that leaves too much room for interpretation and therefore, uncertainty drives over moderation, (b) the technological and financial power possessed by social media platforms that make it practically feasible for them to proceed as quasi-arbitrators and (c) social media providers' economic interests and their contractual freedom that remains eventually untouched despite both national and regional regulatory attempts.

- (a) Overly vague legislation, with certain terms being interpreted more and more extensively by different judicial entities that is gradually more in favor of elaborate voluntary measures and permissive towards restricting users' freedom of expression, mostly due to the uncertain distinction between general and specific monitoring obligations, showing a worrying tendency towards all-encompassing monitoring and filtering that ultimately has a chilling effect of online freedom of expression.

Regarding the neutrality and passivity condition, the DSA adopted the view that hosting service providers only need to remain passive to a certain degree, and may play an active role to some extent, as long as they limit themselves to the mere automatic and technical processing of the information and that role does not necessarily lead to them obtaining knowledge of, or control over, the information. Such approach further encourages monitoring as extensive voluntary measures that previously would disqualify intermediaries from liability exception, now fit into the passive enough condition that still guarantees liability exemption if conditions are met.

- (b) Platforms are in possession of more power and the technological means to take action on hate speech than any state actors: technological capabilities and institutional power vested in social media platforms position them as more effective and responsive actors in addressing hate speech. The global reach, technological means possessed by platforms and their ability to quickly adapt to technological advancements, make platforms uniquely suited to deal with hate speech as the closest actor to user expression and prevent future harm as well for similar content. However, such technological means also carry with them certain risks and vast amount of uncertainties: systems are not only directed at illegal content, but also undesirable content that would be capable of deterring users fearing an unsafe environment and therefore, lessen user-engagement and decrease ad revenues for platforms and by entrusting and encouraging platform

policing, putting constitutional obligations and human right responsibilities on platforms mitigates concerns arise surrounding censorship.

- (c) Even though social media platforms provide a space for public discourse, which unquestionably entails a higher degree of responsibility for respecting fundamental rights, but the essence of the platform-user relationship still remains in the ToSs of social media platforms which is within the realm of freedom of contract, determined unilaterally or predominantly by the platforms. This contractual regulation of online expression coupled with insufficient procedural safeguards surrounding content moderation result in many deficiencies that threaten freedom of expression.

2. Whether the Oversight Board is a legitimate remedy mechanism

I have concluded that the Oversight Board neither provides an effective remedy, nor can we consider it a legitimate element in the overall remedy mechanism system:

- a) Facebook's Community Standards and policies cannot be considered to limit the discretion of Meta itself or the Board as in most cases such standards are to be understood differently from human rights considerations for the pure reason of Meta being a corporation, but no Board decision addresses the basis or reasoning behind such deviation;
- b) The Oversight Board combines and, on a surface level, bears many hallmarks of the courts, arbitral tribunals and international human rights tribunals, it is even somewhat similar to the ODS bodies under Article 21 of the DSA. However, it fails to meet the essential requirements for it to be considered as a(n effective) remedy mechanism:
 - a. the Oversight Board is inaccessible to users, there is no real chance that a user appeal could be 'heard' by the Board,
 - b. the entire appeal procedure places users at a disadvantage by reserving more influence to Facebook, therefore users are at a disadvantage;
 - c. through financial dependence, participation in the appointment and recall of Oversight Board members, exercising control over the substantial and procedural rules that form the basis of decisions, the Oversight Board neither

fulfils the requirements of independence on the institutional nor on the personal level;

- d. users do not have any real influence over the composition of the Oversight Board and the chosen panel for any appeal deliberation, the Board lacks legitimacy, the Oversight Board is undemocratic in this sense;
- e. the Board lacks effective and transparent communication and also being lenient with Facebook when it comes to the accessibility and foreseeability of content moderation rules: turning a blind eye to Facebook, basing its decisions on internal, opaque rules that are not available to users without urging the platform to publish such rules;
- f. although the decisions of the Oversight Board are, in theory, binding, the Board is more preoccupied with recommendations: there are more recommendations published than decisions, even some “binding decisions” are more focused on the advisory opinion part than the actual dispute, evaluating the affected policy against international human rights law standards and not deciding the appeal as a dispute between adversary parties.

3. How ACE, as an ODS body, compares to the Oversight Board?

In my dissertation I opted for the analysis of the procedural rules and procedural realities of a concrete example from the numerous ODS bodies, ACE, in order to better highlight what the advantages of these bodies might be in comparison to the Oversight Board. Surprisingly, I found that the way Meta indirectly finances ACE, still does not hit the threshold of being rejected certification due to the lack of independence.

ACE as an ODS body improved, in terms of the deficiencies identified in case of the Oversight Board, namely accessibility was improved, ACE dealt with ten times more review requests during the first fourteen months, than the Oversight Board did in five years; however, ACE still suffers from two of the most controversial issues that were also the center of attention in case of the Oversight Board: lacking accountability and overall transparency surrounding its operations.

4. Proposals for systematic remedies

The dissertation formulates recommendations and possible amendments in three areas with a view to providing users with a more effective and efficient out-of-court redress system that would give them a realistic opportunity to challenge platform decisions and thus, obtain appropriate redress:

Ad. a) Legislative changes for better freedom of expression protection: Eliminating some of the vague terminologies from legislation that encourages platforms to over moderate UGC and addressing the gaps between constitutional protection of freedom of expression and the low threshold afforded by contractual documents would ensure that less freedom of expression violations occur on social media platforms in the first place.

Ad. b) Amendment of the rules of procedures, specialized ODS system, transparent communication for effectivity and appropriate signposting: Firstly, entities certified to deal with contradictory matters should ensure that their rules of procedures are in line with the requirements of effective remedies Secondly, creating a cooperative network of specialized ODS bodies and other platform-founded mechanisms could better facilitate the needs of users when it comes to appropriate redress. And thirdly, avoiding misleading advertising regarding and incomplete signposting for different redress mechanisms could streamline appeals in the correct direction, ensure timely and appropriate remedy for freedom of expression violations that are free from the platforms' economic interests and consider human rights, avoid unnecessary backlog, mass-refusals.

Ad. c) As the apparent phasing out of the dispute settlement impact of the Oversight Board suggest, the Board and similar entities would be better suited to address policy concern and provide advisory services prior to enacting different policies and their amendments.

Bibliography of the thesis book:

Balkin, Jack M.: Free Speech in the Algorithmic Society: Big Data, Private Governance, And New School Speech Regulation. UC Davis Law Review, 51/3 (2018) 1149-1210.
<https://doi.org/10.2139/ssrn.3038939>

Balkin, Jack M.: The Fiduciary Model of Privacy. Harvard Law Review, 134/1 (2020) 11-33.

List of publications related to the topic of the doctoral dissertation

Gyetzván Dorina: Censorship and Freedom of Expression in the Age of Social Media. *ELTE Law Journal*, 12/2 (2025), 27–40. <https://m2.mtmt.hu/api/publication/35704014>

Gyetzván Dorina – Kovács Andrea: Közösségi média a jogban a nyilvános fórum doktrína és az európai szabályozás tükrében. *Themis*, 21/2 (2025), 6–27. <https://m2.mtmt.hu/api/publication/35667711>

Gosztonyi Gergely – Gyetzván Dorina – Kovács Andrea: Theory and Practice of Social Media's Content Moderation by Artificial Intelligence in Light of European Union's AI Act and Digital Services Act. *European Journal of Law and Political Science*, 4/1 (2025), 33–42. <https://m2.mtmt.hu/api/publication/35756748>

Gosztonyi Gergely – Gyetzván Dorina: The applicability of neomedievalism, technofeudalism and sovereignty in contemporary internet governance. *Ethical Space: The International Journal of Communication Ethics*, 22/1–2 (2025), 31–46. <https://m2.mtmt.hu/api/publication/36098444>

Gosztonyi Gergely – Gyetzván Dorina – Kovács Andrea: A mesterséges intelligencia szólásszabadsága és tartalommoderálási gyakorlata a közösségi médiában a DSA és az MI rendelet vonatkozásában. *Iustum Aequum Salutare*, 21/2 (2025), 61–78. <https://m2.mtmt.hu/api/publication/36393972>

Gosztonyi Gergely – Gyetzván Dorina: Gondolatok a neomedievalizmus, a technofeudalizmus és az internetszuverenitás fogalmainak lehetséges kapcsolódási pontjairól. In: Gosztonyi Gergely – Koltay András – Képes György (szerk.): *A szabadság színeváltozásai. Ünnepi tanulmányok Révész T. Mihály 80. születésnapjára*, (2025), 47–58. <https://m2.mtmt.hu/api/publication/36468139>

Gyetzván Dorina – Gosztonyi Gergely: Garanciális elemek az online platformok véleménynyilvánítást korlátozó moderálási gyakorlatával szemben a digitális szolgáltatásokról szóló rendelet (DSA) előtt és után. *Magyar Jog*, 72/4 (2025), 226–236. <https://m2.mtmt.hu/api/publication/36114588>

Gyetzván Dorina: „A cenzúra rövid története korántsem lenne rövid.” *Jogtörténeti Szemle*, 30/3 (2023), 76–77. <https://m2.mtmt.hu/api/publication/34058285>

Gyetzván Dorina: A Digital Services Act és a Facebook Oversight Board szerepe a jogorvoslat biztosításában a közösségi média vonatkozásában. *In Medias Res*, 12/2 (2023), 190–208. <https://m2.mtmt.hu/api/publication/34526055>

Gyetzván Dorina: A cenzúra privatizációja avagy a cenzúra új virágkora? In: Gosztönyi, Gergely (szerk.): *A velünk élő történelmi cenzúra*, (2022), 57–69. <https://m2.mtmt.hu/api/publication/33127493>

Gyetzván Dorina: Gosztönyi Gergely: Cenzúra Arisztotelésztől a Facebookig: a közösségi média tartalomszabályozási gyakorlatának komplexitása. *In Medias Res*, 11/2 (2022), 184–187. <https://m2.mtmt.hu/api/publication/33551212>

Gyetzván Dorina: Az általános nyomon követési kötelezettség mint a közvetítő szolgáltatók felelősségének jövője? *Jogi Tanulmányok*, (2021), 299–315. <https://m2.mtmt.hu/api/publication/32162771>

Gyetzván Dorina: Recenzió Zódi Zsolt Platformok, robotok és a jog Új szabályozási kihívások az információs társadalomban című könyvéről. *Pro Futuro*, 11/1 (2021), 213–216. <https://m2.mtmt.hu/api/publication/32162775>