

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF LAW AND POLITICAL SCIENCE
DOCTORAL SCHOOL OF LAW

Liza Bukor

**Online Hate Speech in Light of New Technological Challenges,
with Special Regard to Social Media Platforms**

DOCTORAL THESES

Supervisor:

Dr. Zoltán Pozsár-Szentmiklósy, *habilitated associate professor*

Budapest, 2025.

1. Objectives of the Doctoral Dissertation

The rapid development of information technology and the growth of social media platforms have created new avenues for the exercise of freedom of expression.¹ The widespread use of the internet, particularly social media, has created new spaces for communication where freedom of expression increasingly operates in a global, cross-border context. However, the risk of harmful content such as hate speech appearing and spreading has also increased, which can seriously infringe upon other fundamental rights, especially the right to human dignity and equal treatment.²

Online hate speech is one of the most serious and complex challenges to fundamental rights in the digital age, particularly given the changing patterns of communication, the evolving role of social media companies and the use of artificial intelligence-based moderation technologies. The problem is further compounded by the fact that social media platforms, through their own rules and the increasingly widespread use of automated content-moderation systems, assume a decision-making role³ that resembles the classic exercise of state law enforcement.

The aim of this doctoral dissertation is to examine whether online hate speech can be addressed within the framework of the rule of law and the protection of fundamental rights, and, if so, how. A central normative question is whether the legal restriction of online hate speech is permissible, and if so, under what conditions and with what procedural safeguards to ensure a balance between competing fundamental rights is maintained.

The doctoral dissertation aims to address the following research questions:

- What legal and technological frameworks can be used to define online hate speech?

¹ Bissera ZANKOVA; Gergely GOSZTONYI: *Quo vadis, European Union's New Digital Regulation Package?*. Business and Law (2), 2021. pp. 69. DOI: <http://dx.doi.org/10.370.75/BAL.2021.2.03>; Krisztina NAGY; Gábor POLYÁK: *Az internetes forgalomirányító szolgáltatások alapjogi vonatkozásai. (Fundamental Rights Aspects of Internet Traffic Management Services.)* JURA, 24 (1), 2018. pp. 88-99.

² András KOLTAY: *A véleménynyilvánítás szabadsága. (Freedom of Expression.)* In: Lóránt CSINK; Balázs SCHANDA; András VARGA ZS. (eds.): *A magyar közjog alapintézményei. (The Fundamental Institutions of Hungarian Public Law.)* Pázmány Press, Budapest, 2020. pp. 569-597.

³ Jack M. BALKIN: *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation.* 51 UC Davis Law Review, 2018. pp. 1149–1210.

- What role do social media platforms play in managing such content, and what responsibilities do they have?
- Are current platform policies and moderation practices adequate for protecting fundamental rights?
- What are the implications of content moderation based on artificial intelligence and algorithms?
- How can the Oversight Board (FOB), established by Meta, contribute to the oversight of hate speech management based on the rule of law?

Therefore, the objective is not merely to describe and critically interpret current practices, but also to formulate regulatory, institutional and technological proposals that offer long-term, sustainable, rule-of-law-based solutions to the challenges of the digital environment. The aim of these proposals is to promote the transparent, responsible and accountable operation of social media platforms while effectively safeguarding users' fundamental rights.

It is of paramount importance that the balance between fundamental rights, particularly between freedom of expression and the protection of human dignity, is upheld in the digital space. Achieving this balance requires mechanisms that ensure an effective response to harmful content while preventing the unjustified restriction of free speech. Accordingly, the dissertation seeks practical solutions that consider the dynamics of technological development, legal cultural diversity, and the specific characteristics of global platform actors. These solutions will serve as a normative compass for future regulatory and institutional frameworks. The dissertation therefore proposes a hybrid approach that combines the speed and scalability of automated systems with the contextual sensitivity of human moderation. This approach ensures the effective enforcement of procedural safeguards, particularly transparency, the right to a statement of reasons, independent review and the availability of remedies.

2. Structure of the Doctoral Dissertation

The doctoral dissertation is divided into seven main thematic units which guide the reader through a complex analysis of online hate speech. These units are closely aligned with the research questions and are strongly interrelated.

The introductory chapter outlines the importance of the research, its initial questions and hypotheses, and its objectives. It also defines the dissertation's structural and methodological framework.

The second chapter expands upon the dogmatics of freedom of expression, examining the theoretical and legal foundations of hate speech. Particular attention is given to how hate speech is defined and categorised within different legal systems and international human rights instruments. It begins with the acknowledgement that hate speech is not a universally defined legal concept⁴; while some countries permit active criminal prosecution, others allow only limited restrictions. After reviewing these differences, the chapter provides a detailed analysis of the constitutional balance between freedom of expression and efforts to combat hate speech in relation to international human rights standards. It also considers the regulatory practices of social media platforms. These platforms establish their own normative systems through internal policies and community guidelines, which determine whether content qualifies as hate speech. The chapter maps the main characteristics of these normative systems and highlights that they often do not align with formal legal frameworks yet play a decisive role in shaping global public discourse in practice. This normative duality — the tension between formal law and platform regulation — fundamentally defines the possibilities and limitations of addressing online hate speech.

The central thesis of the chapter is that social media platforms perform a legal function, as their decisions directly impact the enforcement of users' fundamental rights, particularly the right to freedom of expression. I demonstrate that the principles of the right to a fair trial, as set out in Article 6 of the European Convention on Human Rights (ECHR), Article XXVIII of the Fundamental Law and the rule of law, require minimum guarantees. These include the right to participate, impartial decision-making, the duty to provide reasons and the right to an effective remedy. Although social media platforms are not public authorities, horizontal effect obliges them to uphold certain fundamental rights requirements whenever their decisions impact users' rights. However, platforms are not bound to comply with constitutional or human rights standards in the same way as state actors, which can lead to serious shortcomings in terms of the transparency of content moderation, the reasoning behind decisions and the availability of remedies. The chapter emphasises that the lack of procedural safeguards,

⁴ Eric HEINZE: *Hate Speech and Democratic Citizenship*. Oxford, 2016., Online edn, Oxford Academic, 21 Apr. 2016. DOI: <https://doi.org/10.1093/acprof:oso/9780198759027.001.0001>

especially given the increasing use of algorithmic moderation, poses a risk of arbitrariness. Ultimately, the chapter concludes that social media platforms should be subject to the requirements of fair procedure, at least to a minimum extent, to ensure transparency, independent review and the availability of effective remedies.

The fourth chapter focuses on artificial intelligence (AI) and algorithm-based content moderation, one of the most widespread — yet also most controversial — technological tools for addressing hate speech today. It aims to explore the theoretical and practical issues arising from automated decision-making by AI systems, especially when such decisions directly impact the exercise of fundamental rights, most notably freedom of expression. It provides a detailed account of how algorithmic moderation is reshaping the responsibility framework of platforms, raising critical questions related to the rule of law, ethics and procedural fairness.

The fifth chapter examines the relationship between law enforcement and the transparency reports published by social media companies. The chapter's central question is the extent to which these documents enable monitoring of platforms' content moderation practices and the role they can play in ensuring transparency, accountability and adherence to norms essential for operations based on the rule of law.

The sixth chapter outlines the institutional operations and decision-making processes of the Oversight Board (FOB), which was established by Meta (formerly Facebook). The chapter focuses particularly on cases concerning hate speech. It outlines the circumstances of the FOB's creation, its legal status and the procedural mechanisms that govern its decision-making. It also analyses how the FOB fulfils its role as a quasi-law-applying and remedial body. The analysis focuses on how the FOB is able to make context-sensitive decisions based on human rights norms, particularly freedom of expression, human dignity and equality, thereby supervising the platform's moderation practices. By examining all FOB decisions concerning hate speech, the chapter illustrates how the Board assesses borderline cases and the interpretative framework it employs. It also explains the considerations on which the Board's jurisprudence is based. Particular attention is given to the influence of the FOB's decisions on Meta's internal regulation and moderation culture. Furthermore, the chapter demonstrates that the FOB fulfils a corrective role within Meta while also functioning as a bridge between private regulation and the international human rights regime, despite its limited jurisdiction and enforcement mechanisms.

Finally, the concluding chapter summarises the dissertation's main findings and makes recommendations regarding the future regulation of platforms, algorithmic decision-making and the protection of human rights. The chapter aims to present the theoretical and practical findings of the dissertation in a coherent framework, identifying areas for further development. Additionally, the concluding chapter emphasises how the relationship between social media platforms and state regulators may evolve within the global digital sphere in the future.

3. Research Methods Applied in the Doctoral Dissertation

This dissertation takes an interdisciplinary approach to the issue of online hate speech, combining legal, technological and social science research methods. The research incorporates qualitative and quantitative elements based on three main methodological pillars:

Through a dogmatic legal analysis, this dissertation examines domestic and international doctrines concerning freedom of expression and the right to a fair trial. Particular emphasis is placed on the dual function of freedom of expression — as both individual self-expression and a safeguard of democratic public discourse — and on the guarantees inherent in fair procedure, such as participation, reasoning, remedy and impartiality. Based on these theoretical principles, the dissertation asserts that hate speech restrictions should be evaluated using the three-step proportionality test, as this offers a methodological framework that ensures compliance with the rule of law and safeguards fundamental rights.

As part of a comparative legal analysis, it primarily examines the norms set out in international human rights documents, such as the International Covenant on Civil and Political Rights, the European Convention on Human Rights and the Charter of Fundamental Rights of the European Union. Building on these, the dissertation analyses how different legal systems⁵ — such as the European model, the constitutional framework of the United States, and various national regulatory efforts — interpret and restrict hate speech. The characteristics of these normative systems are compared with the internal policies, moderation guidelines, and practices of major social media platforms—YouTube, Meta, X (formerly Twitter), and TikTok. The aim is to explore the structural and substantive differences between the normative frameworks developed by non-state actors and traditional, law-based regulatory systems. In

⁵ Robert KAHN: *A gyűlölet szabadsága – amerikai és európai perspektívák. (Freedom to Hate: American and European Perspectives.)* Wolters Kluwer Kft., Budapest, 2016. pp. 7.

addition, the dissertation provides a detailed analysis of how platform rules—and especially the decisions of the Facebook Oversight Board (FOB) established by Meta—align with international human rights standards. It examines the extent to which these decisions reflect the required balance between freedom of expression and human dignity, as well as the ways in which they diverge from these standards.

As part of a quantitative content analysis, I processed the statistical data found in the transparency reports published by the platforms, paying particular attention to the quantitative indicators related to the moderation of hate speech. The analysis covers temporal changes in the volume of content removed, the extent to which artificial intelligence and algorithms are used, and the degree of automation in decision-making processes. I also examine the consistency, transparency and procedural detail of individual platforms' moderation practices. This method allows me to empirically assess the extent to which social media platforms' moderation practices can maintain the balance between freedom of expression and the protection of human dignity, and the degree to which procedural safeguards are effectively ensured.

As part of the qualitative analysis, I adopt a case study approach, focusing on the decisions issued by Meta's Oversight Board. The selected cases are representative examples that effectively illustrate the challenges of platform moderation practices, such as context-sensitive interpretation, navigating the boundaries between hate speech and political expression, and accounting for cultural and linguistic differences.⁶ The analysis focuses on how these decisions contribute to the development of soft law norms and the extent to which they set a precedent for the operation of the platform and the broader regulation of digital public discourse.

Finally, theoretical legal interpretation is an integral part of the research, through which I examine the foundational principles that may justify restricting freedom of expression. This methodological approach allows for a more in-depth, normative analysis of the legal assessment of hate speech and a critical evaluation of the rule of law foundations of various restriction models.

⁶ András KOLTAY: *A gyűlöletbeszéd korlátozása a magyar jogrendszerben. (Restrictions on Hate Speech in The Hungarian Legal System.)* In: András KOLTAY (ed.): *A gyűlöletbeszéd korlátozása Magyarországon. (Restrictions on Hate Speech in Hungary.)* Wolters Kluwer, Budapest, 2013.

4. Findings of the Doctoral Dissertation

The research has yielded novel findings across multiple dimensions, which are relevant from both theoretical and practical perspectives.

1. One of the dissertation's most significant scholarly contributions is its detailed demonstration that the concept of online hate speech is not defined uniformly across different legal systems. In fact, even among democratic states, substantial differences exist regarding the definition of hate speech, the permissibility of restrictions, and the legal consequences.
 - a. While some countries, such as Germany and France, apply a broader, more proactive interpretation and may respond to hate speech with criminal sanctions, others, such as the United States, recognise the legitimacy of restrictions only within a much narrower scope, citing constitutional protections for freedom of speech. These differences are not merely theoretical: they directly impact how global social media platforms handle hate speech.
 - b. By contrast, platforms such as Meta, YouTube, TikTok and X have their own definitions of hate speech, which are often vague or overly broad. These definitions are frequently inconsistent with national laws or international human rights standards. These differences in definition can lead to divergent legal interpretations, inconsistent practices and disproportionate restrictions in moderation decisions by platforms.⁷ This is particularly concerning when platforms fail to provide sufficient transparency about the rationale behind their decisions or when users lack access to effective legal remedies or appeals processes.
 - c. The dissertation therefore proposes developing a context-sensitive and normatively grounded definition of hate speech. This definition would take into

⁷ András KOLTAY: *Az új média és a szólásszabadság: A nyilvánosság alkotmányos alapjainak újragondolása.* (New Media and Freedom of Speech: Rethinking the Constitutional Foundations of Public Discourse.) Wolters Kluwer, Budapest, 2019. pp. 232.

account the specific characteristics of online communication environments,⁸ such as the speed of mass dissemination, linguistic diversity, cultural differences and the mediated nature of technology. The core of the proposed approach lies in its flexibility to adapt to different social contexts while remaining aligned with international human rights standards.

- d. As a scientific contribution, the dissertation demonstrates that the dual individual and collective function of freedom of expression requires the proportionality test to be applied when assessing the restrictability of hate speech. This theoretical foundation not only applies in the context of state law enforcement, but should also, in a normative sense, guide the decisions of private platforms.
2. Another key finding of the research is the critique of the quasi-judicial role assumed by social media companies based on the rule of law. While not formally considered public authorities, these companies' decisions, particularly those involving content removal, restriction or user suspension, have a tangible impact on users' fundamental rights. These decisions significantly influence the nature of online public discourse⁹, determining which opinions and societal conversations are visible or suppressed in the digital space.¹⁰
 - a. The dissertation provides a detailed account of the institutional and procedural issues inherent in platform decision-making processes. Content moderation mechanisms are often opaque: policy references are unclear, justifications for decisions are lacking or automated, and users are often denied access to effective complaint or appeal processes.¹¹ These issues violate key principles of the rule of law, particularly the principles of accountability, predictability and

⁸ Jack M. BALKIN: *How to Regulate (and Not Regulate) Social Media*. 1 Journal of Free Speech Law 1 pp. 71-96. 2021. DOI: <http://dx.doi.org/10.2139/ssrn.3484114>

⁹ Bernát TÖRÖK: *A szólásszabadság a közösségi platformokon és a Digital Services Act. (Freedom of Speech on Social Media Platforms and the Digital Services Act.)* In: Bernát TÖRÖK; Zsolt ZÓDI: *Az internetes platformok kora. (The Age of Internet Platforms.)* NKE Ludovika Egyetemi Kiadó, Budapest, 2022. pp. 195.

¹⁰ János Tamás PAPP: *A közösségi média szabályozása a demokratikus nyilvánosság védelmében. (Regulating Social Media to Protect Democratic Public Discourse.)* Wolters Kluwer, Budapest, 2022. pp. 171.

¹¹ SZIKORA Tamás: *A platformszabályozás egy új iránya Donald Trump Facebookról való kitiltásának példáján keresztül. (A New Direction in Platform Regulation, as Seen in Donald Trump's Ban from Facebook.)* In: Bernát TÖRÖK; Zsolt ZÓDI: *Az internetes platformok kora. (The Age of Internet Platforms.)* NKE Ludovika Egyetemi Kiadó, Budapest, 2022. pp. 211.

the right to remedy, and, over time, they erode trust in the legitimate operation of these platforms.

- b. Building on the dogmatic foundations of the right to a fair trial, the dissertation shows that the minimum requirements of this right – participation, reasoning, transparency and remedy – also apply to platforms under the principle of horizontal effect, given that their decisions affect fundamental rights.
 - c. The dissertation also examines the impact of the Digital Services Act (DSA), which raises the transparency and accountability standards for platforms¹² across Europe and could establish a legal framework for protecting users' rights.
 - d. Although transparency reports published by platforms may allow certain conclusions to be drawn about content moderation¹³, they do not provide sufficiently detailed information about the legality of moderation decisions or the procedural context in which they are made.
 - e. The numbers do not necessarily indicate a more effective and just system; rather, they often reveal increasingly automated mechanisms that exclude user participation. For this reason, it is crucial that social platforms do not view technical developments as the sole solution, but instead empower users' rights, comply with transparency requirements and provide genuine and effective remedies.
3. Systems based on artificial intelligence are playing an increasingly significant role in the content moderation practices of online platforms.
- a. Although AI-powered moderation systems enable the rapid processing of large amounts of content, they lack the capacity to accurately interpret social, cultural and linguistic contexts, especially in borderline cases of hate speech.¹⁴ This

¹² Dorina GYETVÁN; Gergely GOSZTONYI: *Garanciális elemek az online platformok véleménynyilvánítást korlátozó moderálási gyakorlatával szemben a digitális szolgáltatásokról szóló rendelet (DSA) előtt és után. (Safeguards Against Moderation Practices that Restrict Expression on Online Platforms Before and After the Digital Services Act (DSA).)* Magyar Jog, 72 (4), 2025. pp. 226-236. DOI: <https://doi.org/10.59851/mj.72.04.3>

¹³ Amelie HELDT: *Reading between the lines and the numbers: an analysis of the first NetzDG reports.* Internet Policy Review, 8 (2), 2019. pp.10. Elérhető itt: <https://policyreview.info/pdf/policyreview-2019-2-1398.pdf> DOI: <https://doi.org/10.14763/2019.2.1398>;

¹⁴ Emma LLANSÓ; Joris VAN HOBOKEN; Paddy LEERSSEN; Jaron HARAMBAM: *Artificial Intelligence, Content Moderation, and Freedom of Expression.* Transatlantic Working Group, Working Paper 2020. pp. 8. Elérhető itt: <https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf>

absence of contextual sensitivity can result in erroneous decisions, such as the removal of lawful content or the retention of harmful material.¹⁵ Such algorithmic errors violate the principle of legal certainty, particularly when decisions are unsubstantiated and users have no access to effective remedies.

- b. The doctoral dissertation emphasises the crucial role of human decision-making in evaluating sensitive cases, as this is the only way to strike the right balance between fundamental rights.¹⁶ It also calls for greater transparency and accountability in the operation of algorithms¹⁷, arguing that users should be informed of the logic behind automated decisions and have access to effective remedies. Only then can artificial intelligence support, rather than undermine, the functioning of the rule of law.
 - c. The dissertation concludes that solutions based solely on technical developments often exclude user participation, which undermines the requirements of fair procedure and legitimacy in the long term.
 - d. As a novel insight, the dissertation demonstrates that a single change in ownership or strategic direction may fundamentally reshape a platform's commitment to fundamental rights, thereby rendering the enforcement of rule-of-law guarantees in the digital sphere more vulnerable.
4. The Meta Oversight Board plays a model role in the practical application of fundamental rights standards; however, it is also subject to legitimate criticism.
- a. The novelty of the dissertation lies in its comprehensive analysis of all FOB decisions¹⁸ relating to hate speech. This provides the first thorough overview of the Board's practices in this area.

¹⁵ R. GORWA; R. BINNS; C. KATZENBACH: *Algorithmic content moderation: Technical and political challenges in the automation of platform governance*. In: *Big Data & Society*, 7 (1), 2020. pp. 5. DOI: <https://doi.org/10.1177/2053951719897945>

¹⁶ Anna SCHMIDT; Michael WIEGAND: *Survey on Hate Speech Detection using Natural Language Processing*. In: *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, Valencia, Spain. Association for Computational Linguistics. 2017. pp. 1–10. DOI: <http://dx.doi.org/10.18653/v1/W17-1101>

¹⁷ Bernát TÖRÖK: *Az Alkotmányjog horizontális hatálya? (The Horizontal Scope of Constitutional Law?)* In.: Bernát TÖRÖK; Zsolt ZÓDI (eds.): *A mesterséges intelligencia szabályozási kihívásai. (Regulatory Challenges of Artificial Intelligence.)* Ludovika Kiadó; Budapest, 2021. pp. 149.

¹⁸ Gergely LENDVAI: *A Facebook Ellenőrző Bizottság működése és bíraskodása a gyűlöletbeszéd kontextusában. (The Facebook Oversight Board's Operation and Adjudication in the Context of Hate Speech.)* In: *Medias Res* 13, 2024. pp. 195–221. DOI: <https://doi.org/10.59851/imr.13.1.11>

- b. The FOB's decisions are context-sensitive, transparent and well-founded from a human rights perspective. In each case, the Board provides a detailed justification, clearly outlining the human rights standards applied and assessing the legality of content removal in light of these standards.¹⁹ Particular attention is paid to cultural and societal context, the intent and impact of the expression, and the potential risks posed to other fundamental rights, such as human dignity.²⁰ This approach enables the FOB to exercise a deliberative practice that meets theoretical and practical expectations of the rule of law, and addresses the complexity of the hate speech issue.
- c. The decisions of the FOB also carry precedent value for the platforms. The Board's decision-making process is gradually becoming a norm-setting practice within the global digital regulatory landscape, establishing 'soft law'-type principles and interpretive patterns that can guide other technology companies. According to the dissertation, more widespread dissemination and adoption of the FOB's practices could significantly contribute to the development of a legitimate and lawful moderation culture, one that is not only grounded in technical efficiency, but also in fundamental rights principles.
- d. Although concerns have been raised regarding the FOB's independence, particularly in relation to the transparency of the selection process for its decision-makers. In the long term, this lack of transparency could undermine the Board's legitimacy.
- e. The practical enforceability of the FOB's decisions is limited because Meta does not automatically implement all of the Board's recommendations. Consequently, the legal impact of these decisions is often uncertain.

In conclusion, the legal and technological management of online hate speech can only be effective if it considers the balance of fundamental rights, platform responsibility, and the unique characteristics of the new technological environment simultaneously. The doctoral

¹⁹ Joan BARATA: *The Decisions of the Oversight Board from the Perspective of International Human Rights Law*. COLUM. U. GLOBAL FREEDOM OF EXPRESSION 10, 2022. <https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2022/10/The-Decisions-of-the-OSB-from-the-Perspective-of-Intl-Human-Rights-Law-Joan-Barata-.pdf> Utolsó letöltés dátuma: 2025. 05. 15

²⁰ Gergely GOSZTONYI: *Az internetes tartalomszabályozással kapcsolatos új gondolkodási irányok az Amerikai Egyesült Államokban. (New Approaches to Internet Content Regulation in the United States.)* Miskolci Jogi Szemle, 16 (4), 2021. pp. 47. DOI: <https://doi.org/10.32980/MJSz.2021.4.1257>

dissertation aims to explore this balance, developing theoretical and practical frameworks that support the protection of fundamental rights within digital communication. Ultimately, the dissertation concludes that platforms' rule-of-law legitimacy can only be strengthened if they reinforce users' procedural rights, comply with transparency obligations, and ensure the availability of genuine and effective remedies, rather than relying solely on automated rule enforcement.

Bibliography of Theses

Amelie P. HELDT: *Reading between the lines and the numbers: an analysis of the first NetzDG reports.* Internet Policy Review, 8 (2), 2019. pp. 1-18. DOI: <https://doi.org/10.14763/2019.2.1398>

András KOLTAY: *A gyűlöletbeszéd korlátozása a magyar jogrendszerben: A büntetőjogi, polgári jogi és médiaszabályozási normák alkotmányos értelmezése és gyakorlati alkalmazásuk elemzése, európai kitekintéssel. (Restrictions on Hate Speech in the Hungarian Legal System: Constitutional Interpretation of Criminal Law, Civil Law, and Media Regulation Standards and Analysis of Their Practical Application, with a European Perspective.)* In: András KOLTAY (ed.): *A gyűlöletbeszéd korlátozása Magyarországon: Alkotmányos és jogalkalmazói megközelítések, európai kitekintéssel. (Restricting Hate Speech in Hungary: Constitutional and Law Enforcement Approaches, with a European Perspective.)* Wolters Kluwer, Budapest, 2013. pp. 1-123.

András KOLTAY: *A véleménynyilvánítás szabadsága. (Freedom of Expression.)* In: Lóránt CSINK, Balázs SCHANDA, András VARGA ZS. (eds.): *A magyar közjog alapintézményei. (The Fundamental Institutions of Hungarian Public Law.)* Pázmány Press, Budapest, 2020. pp. 569-597.

András KOLTAY: *Az új média és a szólásszabadság - A nyilvánosság alkotmányos alapjainak újragondolása. (New Media and Freedom of Speech: Rethinking the Constitutional Foundations of Public Discourse.)* Wolters Kluwer, Budapest, 2019.

Anna SCHMIDT; Michael WIEGAND: *Survey on Hate Speech Detection using Natural Language Processing.* Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, Valencia, Spain. Association for Computational Linguistics, 2017. pp. 1–10. DOI: <http://dx.doi.org/10.18653/v1/W17-1101>

Bernát TÖRÖK: *A szólásszabadság a közösségi platformokon és a Digital Services Act. (Freedom of Speech on Social Media Platforms and the Digital Services Act.)* In: Bernát TÖRÖK; Zsolt ZÓDI: *Az internetes platformok kora. (The Age of Internet Platforms.)* NKE Ludovika Egyetemi Kiadó, Budapest, 2022. pp. 195-207.

Bernát TÖRÖK: *Az Alkotmányjog horizontális hatálya? (The Horizontal Scope of Constitutional Law?)* In.: Bernát TÖRÖK; Zsolt ZÖDI (eds.): *A mesterséges intelligencia szabályozási kihívásai. (Regulatory Challenges of Artificial Intelligence.)* Budapest, NKE Ludovika Egyetemi Kiadó, 2021. pp. 143-156.

Bissera ZANKOVA; Gergely GOSZTONYI: *Quo vadis, European Union's New Digital Regulation Package?*. Business and Law (2), 2021. pp. 67-90. DOI: <http://dx.doi.org/10.370.75/BAL.2021.2.03>

Dorina GYETVÁN; Gergely GOSZTONYI: *Garanciális elemek az online platformok véleménynyilvánítást korlátozó moderálási gyakorlatával szemben a digitális szolgáltatásokról szóló rendelet (DSA) előtt és után. (Safeguards Against Moderation Practices that Restrict Expression on Online Platforms Before and After the Digital Services Act (DSA).)* Magyar Jog, 72 (4), 2025. pp. 226-236. DOI: <https://doi.org/10.59851/mj.72.04.3>

Emma LLANSÓ; Joris VAN HOBOKEN; Paddy LEERSEN; Jaron HARAMBAM: *Artificial Intelligence, Content Moderation, and Freedom of Expression*. Transatlantic Working Group, Working Paper, 2020.

Eric HEINZE: *Hate Speech and Democratic Citizenship*. Oxford, 2016., Online edn, Oxford Academic, 21 Apr. 2016. DOI: <https://doi.org/10.1093/acprof:oso/9780198759027.001.0001>

Gergely GOSZTONYI: *Az internetes tartalomszabályozással kapcsolatos új gondolkodási irányok az Amerikai Egyesült Államokban. (New Approaches to Internet Content Regulation in the United States.)* Miskolci Jogi Szemle, 16 (4), 2021. pp. 40-54. DOI: <https://doi.org/10.32980/MJSz.2021.4.1257>

Gergely LENDVAI: *A Facebook Ellenőrző Bizottság működése és bíraskodása a gyűlöletbeszéd kontextusában. (The Facebook Oversight Board's Operation and Adjudication in the Context of Hate Speech.)* In Medias Res 13, 2024. pp. 195–221. DOI: <https://doi.org/10.59851/imr.13.1.11>

Jack M. BALKIN: *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*. 51 UC Davis Law Review, 2018. pp. 1149–1210.

Jack M. BALKIN: *How to Regulate (and Not Regulate) Social Media*. 1 Journal of Free Speech Law 1, 2021. pp. 71-96. DOI: <http://dx.doi.org/10.2139/ssrn.3484114>

János Tamás PAPP: *A közösségi média szabályozása a demokratikus nyilvánosság védelmében. (Regulating Social Media to Protect Democratic Public Discourse.)* Wolters Kluwer Kft., Budapest, 2022.

Joan BARATA: *The Decisions of the Oversight Board from the Perspective of International Human Rights Law.* COLUM. U. GLOBAL FREEDOM OF EXPRESSION 10, 2022. <https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2022/10/The-Decisions-of-the-OSB-from-the-Perspective-of-Intl-Human-Rights-Law-Joan-Barata-.pdf> Utolsó letöltés dátuma: 2025. 05. 15

Krisztina NAGY; Gábor POLYÁK: *Az internetes forgalomirányító szolgáltatások alapjogi vonatkozásai. (Fundamental Rights Aspects of Internet Traffic Management Services.)* JURA, 24 (1), 2018. pp. 88-99.

R. GORWA; R. BINNS; C. KATZENBACH: *Algorithmic content moderation: Technical and political challenges in the automation of platform governance.* Big Data & Society, 7(1), 2020. pp. 1-15. DOI: <https://doi.org/10.1177/2053951719897945>

Robert KAHN: *A gyűlölet szabadsága – amerikai és európai perspektívák. (Freedom to Hate: American and European Perspectives.)* Wolters Kluwer Kft., Budapest, 2016

Tamás SZIKORA: *A platformszabályozás egy új iránya Donald Trump Facebookról való kitiltásának példáján keresztül. (A New Direction in Platform Regulation, as Seen in Donald Trump's Ban from Facebook.)* In: Bernát TÖRÖK; Zsolt ZÓDI: *Az internetes platformok kora. (The Age of Internet Platforms.)* NKE Ludovika Egyetemi Kiadó, Budapest, 2022. pp. 209-226.

List of Publications Related to the Topic of the Doctoral Dissertation

Liza BUKOR: *Az online gyűlöletbeszéd és a tárhelyszolgáltatók felelőssége.* (Online Hate Speech and the Responsibility of Hosting Providers.) *Közjogi Tükör* 1, 2019. pp. 4-8.

Liza BUKOR: *A gyermeki jogok védelme a közösségi médiában.* (Protecting Children's Rights in Social Media.) *Jogi Tanulmányok* 22, 2021. pp. 104-118.

Liza BUKOR: *Tartalomszabályozás a közösségi média korában.* (Content Regulation in the Age of Social Media.) *JEL-KÉP: Kommunikáció Közvélemény Média* 43 : 3, 2022. pp. 103-105.

Liza BUKOR: *A Facebook Ellenőrző Bizottságának jogalkalmazó tevékenysége, különös tekintettel az online gyűlöletbeszédre.* (The Law Enforcement Activities of Facebook's Oversight Board, with a Focus on Online Hate Speech.) *Jogi Tanulmányok* 23 : 1, 2022. pp. 9-21.

Liza BUKOR: *A DSA a véleménynyilvánítás szabadságának tükrében.* (The DSA in the Light of Freedom of Expression.) In: András KOLTAY; Tamás SZIKORA; András LAPSÁNSZKY (eds.): *A vadnyugat vége?: Tanulmányok az Európai Unió platformszabályozásáról.* (The End of the Wild West?: Studies on Platform Regulation in the European Union.) ORAC Kiadó Kft., Budapest, 2024. pp. 311-325.

Liza BUKOR; Anikó TRÄGER: *Az online gyűlöletbeszéd detektálásának kérdései, különös tekintettel az algoritmikus rendszerekre.* (Issues in Detecting Online Hate Speech, with Special Regard to Algorithmic Systems.) *MTA LAW WORKING PAPERS* 2, 2024. pp. 1-14.