



The Ethical Consequences of Artificial Intelligence in Countering Cyber Speech: Combining Effectiveness with Maintaining Human Rights

Ashini Hansika Godigamuwa

LLB (Hons) Law, Staffordshire University

Email address of the corresponding author - ashinihansika261@gmail.com

Abstract

Artificial Intelligence (AI) offers tremendous potential and difficult moral dilemmas in the fight against cyber speech, including hate speech, disinformation, and cyberbullying. This study looks at the two requirements that must be met to protect civil rights and successfully combat harmful online speech. By showcasing developments in deep learning algorithms, natural language processing, and automated moderation tools, it explores the potential of AI systems to identify, regulate, and lessen harmful online behavior. The ethical implications of AI in moderating online debate are rigorously examined in this paper, with particular attention paid to issues with biases, privacy, and freedom of speech. AI creates concerns about data exploitation and spying. It may also over-censor or misinterpret context, which puts permissible expression at risk of being unfairly suppressed. Additionally, AI systems have the power to amplify and perpetuate preconceptions, resulting in biased judgments that affect marginalized communities. Through an analysis of case studies and statutes, the study seeks to strike a balance between the need to preserve fundamental rights and AI's ability to make online places safer. It promotes a plan that upholds justice and human dignity by fusing technical advancements with strict moral standards and open governance.

Keywords: Cyber Speech; Artificial Intelligence; Civil Liabilities; Privacy Concerns; International Cooperation; Legislatives

Introduction

The digital age has transformed communication and networking, but it also creates obstacles, particularly in dealing with 'unwanted' cyber speech. A complex strategy that covers legal, technological, ethical, and sociological concerns is essential to develop successful regulatory frameworks. The regulation of undesired cyber speech in the information society, emphasising on its intricacies, global character, technical improvements, and the balance between free expression and harm prevention. It wants to add to the continuing discussion about how to balance fundamental rights with the negative effects of cyber speech.

Literature Review

Studies reveal AI systems are getting better at spotting harmful content like hate speech, disinformation, and cyberbullying, especially when it comes to machine learning and natural language processing through Algorithms as stated by Katzenbach & Ulbricht.¹ However, serious ethical issues still exist. The

enormous data collection necessary for AI moderation gives rise to privacy concerns as mentioned by Zuboff.² Because AI may inadvertently identify acceptable content, resulting in excessive censorship, freedom of expression is an jeopardy as stated by Zuckerman.³ Furthermore, biased AI systems may lead to biased actions that harm underprivileged groups as mentioned by Noble.⁴ These worries highlight the necessity for a well-rounded strategy that incorporates ethical issues, strong legal frameworks, and technological efficacy.

Methodology

The researcher utilized an extensive desk research approach, carefully examining qualitative material taken from primary and secondary legal sources, laws, regulations, and precedents. The ethical ramifications of AI in suppressing cyber speech are investigated in this study using a mixed-methods methodology. Practical examples are given by case studies of cyber security on social media, to comprehend how various governments handle AI's ethical concerns. This multifaceted strategy guarantees both the preservation of civil rights and a sophisticated knowledge of Cyber security.

Research Problem and Results

An intricate research question that lies at the nexus of technological capabilities and ethical responsibility is how to use AI technology to combat cyber speech.

¹ Katzenbach, C. and Ulbricht, L., 'Algorithmic Governance' (2019) 8(4) Internet Policy Review <https://www.econstor.eu/handle/10419/210652> accessed 30 July 2024..

² Zuboff, S., *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (PublicAffairs 2019).

³ Zuckerman, E., 'New Media, New Civics?' (2016) 6 Policy & Internet 151-168.

⁴ Charissa, N., *Extended from What?: Tracing the Construction, Flexible Meaning, and Cultural Discourses of "Extended Vocal Techniques"* (Santa Cruz ProQuest Dissertations &Theses 2019).

Although AI provides effective methods for identifying and removing dangerous internet information, there are serious ethical concerns with its applications. This study aims to address concerns related to the methods that could be employed to safeguard civil rights to privacy and freedom of speech, thereby resulting in online safety and respecting the values of equality, fairness, and human dignity.

Cyber Speech

Cyber Speech refers to any kind of online expression or communication, including text, photos, and videos on platforms like the internet and social media. It includes a wide range of digital activities and discussions and legal and ethical considerations concerning topics such as free expression, privacy, and cyberbullying. Cyberspace allows for free communication, but social media is being abused for online hate speech, due to their race, color, nationality, sexuality, gender, nationality, faith, and political affiliation.⁵

Unwanted Cyber Speech

Unwanted Cyber Speech is controlled to balance free expression with harm prevention, addressing possible harm, rights infringement, and regulatory impediments. Teenage cyberbullying carries threats to their mental health. Hate speech is controlled and spreads discrimination on online forums. Security and well-being are impacted by online abuse, including stalking. False information distorts perceptions and is harmful, particularly regarding health. Anti-discrimination laws forbid cyber-racism. With its provocative material, trolling stifles debate. Pornography that glorifies revenge injures people's feelings of privacy and discomfort, particularly young people. Anonymity makes identity theft easier and fosters harassment. Greed-driven cyber-white-collar crimes that harm society include copyright violations and bribes.

⁵ Castano-Pulgarin S. A., Natalia Suarez-Betancur, Tilano Vega, L. M., Herrera Lopez, 'Aggression and Violent Behavior: Internet, social media and online hate speech. Systematic review' (2021) 58 (10) Science Direct <<https://www.sciencedirect.com/science/article/abs/pii/S1359178921000628>> accessed 6 April 2021.

Challenges

The digital era provides a huge difficulty in harmonizing free speech rights with possible harm, necessitating complicated legal, ethical, and social issues to control online communication platforms.

- **Freedom of Expression vs. Preventing Harm:** Balancing free expression with harm reduction is a difficult topic that necessitates detecting harmful communication and adopting suitable rules without inhibiting it.⁶ Social networking sites such as
- Facebook, Twitter, and YouTube have been chastised for failing to remove potentially hazardous information.⁷ Freedom of expression is essential for democracy, yet it can be limited by the state for national security, public safety, wellness, and ethics, frequently without justification.⁸
- **Global Reach and Cross-Border Issues:** The cross-border nature challenges of internet regulatory attempts. Different legal structures and cultural variables make establishing global norms for dealing with inappropriate online speech challenging.⁹
- **Anonymity and Pseudonymity:** Online hate speech, like offline hate speech, is anonymized, spreads swiftly, and is confusing across countries, making it difficult to identify and keep those responsible accountable.¹⁰
- **Rapid Spread in Information:** Cyber speech can spread quickly, causing serious damage before intervention, and the rapid dissemination of material online undermines standards techniques of response and prevention.¹¹
- **Role of Intermediaries:** Online platforms and social media are critical in shaping and distributing content, screening user-generated content, and resolving concerns such as bias, censorship, and disruptive behavior.¹²
- **Resource Limitation:** Limited resources, varied legal systems, and ongoing innovation by malicious actors impede efficient surveillance and enforcement of online rules, making globally applicable legal structures difficult.
- **Public Education and Awareness:** Raising public knowledge and competency with technology is critical in light of the privacy challenges raised by Edward Snowden's revelations concerning surveillance and data collecting.¹³
- **Complexity of Definitions:** The complexities of cyber speech definitions can lead people to express controversial viewpoints getting legalized, raising problems regarding online expression boundaries and law enforcement's involvement in addressing perceived harms.¹⁴ The United Na-

⁶ RES/38/7 The Human Rights Council of 18 June-6 July 2018 the promotion, protection and enjoyment of human rights in the Internet.

⁷ Hatano, A., 'Regulating Online Hate Speech through the Prism of Human Rights Law: The Potential of Localized Content Moderation' (2023) 41 (1) ResearchGate <<https://www.researchgate.net/publication/375636662>> accessed 23 October 2023.

⁸ Gunatilleke, G., 'Justifying Limitations on the Freedom of Expression' (2021) 22 Human Rights Review 02.

⁹ Casino, F., Pina, P., Lopez-Aguilar, P., Batista, E., Solanas, A., Patsakis, C., 'SoK: cross-border criminal investigations and digital evidence' (2022) 8 (1) Journal of Cybersecurity <<https://academic.oup.com/cybersecurity/article/8/1/typac014/6909060>> accessed 25 May 2022.

¹⁰ Hatano (n 3).

¹¹ Khan, T., Michalas, A., Akhunzada, A., 'Fake news outbreak 2021: Can we stop the viral spread?' (2021) 190 Journal of Network and computer Applications <<https://www.sciencedirect.com/science/article/pii/S1084804521001326>> 15 September 2021.

¹² United Nations Human Rights, Office of the High Commissioner, Special rapporteur on the promotion and protection of the right to freedom of opinion and expression 1993, A/HRC/53/25.

¹³ J Pujol, 'Is this the end of privacy? Snowden and the power of conscience' (2020) 5 (1) Church, Communication and Culture <<https://www.tandfonline.com/doi/full/10.1080/23753234.2020.1713017>> accessed 17 February 2020.

¹⁴ Casciani, D., 'Harry Miller: Legal victory after alleged transphobic tweets' (BBC News, 20 December 2021) <<https://www.bbc.com/news/uk-england-lincolnshire-59727118>> accessed 20 December 2021.

tions (UN) handles cyber dangers internationally through efforts that promote human rights and freedom of expression, but member nations' different perceptions confuse global definitions.¹⁵ The problem of balancing free expression with protection from misleading and damaging claims in the digital domain.¹⁶

- **Danger of Future Generation:** Online Child Sexual Abuse (CSA) includes pornography for both business and personal use, plus anonymous sessions via platforms. Online CSA can suggest ongoing or potential child abuse, or it can contribute to future physical abuse by an abuser.¹⁷
- **Legal Issues:** Cyber white-collar crime is a sophisticated, difficult crime in law, including digital transactions, which frequently results in serious accusations, prolonged investigations, widespread distrust, and economic harm.¹⁸

The United Nations (UN) Enforcement

The UN endorses the Universal Declaration of Human Rights, which permits restrictions on expression if it violates the rights of others. UN Special Rapporteurs look at the advancement of free speech both offline and online. The UN recommends ethical technology use, media development, internet governance, and

¹⁵ Bostoen, F., 'Understanding the Digital Markets Act' (2023) 68 (2) Journal Indexing and Metrics <<https://journals.sagepub.com/doi/10.1177/0003603X231162998>> accessed 12 April 2023.

¹⁶ Defamation Act (DA) 2013.

¹⁷ Ali, S., Haykal, H. A., and Youssef, E. Y. M., 'Child Sexual Abuse and the Internet: A Systematic Review' (2021) 4(1) Human Arenas: Arena of Technologies https://www.researchgate.net/publication/352120820_Child_Sexual_Abuse_and_the_Internet-A_Systematic_Review accessed 11 May 2021.

¹⁸ Payne, B. K. and Pitman, L., 'Technology, White-Collar Cybercrime, and White-Collar Crime: An Analysis of Utah's White-Collar Crime Registry Data' (2022) 4(2) Journal of White Collar and Corporate Crime <https://journals.sagepub.com/doi/10.1177/2631309X221128622> accessed 22 September 2022.

journalist safety. The Internet Society does not restrict expression, even as it encourages cybersecurity cooperation.¹⁹ A global campaign against hate speech was introduced by the UN Secretary-General in June 2019.²⁰ Online safety and cybersecurity are areas in which the UN is active. Accords such as the Budapest Convention deal with cyber conversation about illegal activity.²¹ EU Charter of Fundamental Rights and UN Convention on the Rights of the Child both safeguard rights of children.²²

The European Union (EU) Enforcement

In order to improve cybersecurity, the EU harmonizes cyber speech laws while taking linguistic, legal, and cultural considerations into account.²³ The EU Treaty Article 114²⁴ seeks to advance social and economic progress while limiting the spread of terrorist propa-

¹⁹ United Nations International Telecommunication Union (ITU), Department of Economic and Social Affairs 2022.

²⁰ Hatano (n 3).

²¹ Kleinwachter, W., 'Cybersecurity, Internet Governance, and the Multistakeholder Approach' (2021) Cyberstability Paper series <<https://hcss.nl/report/cybersecurity-internet-governance-and-the-multistakeholder-approach-the-role-of-non-state-actors-in-internet-policy-making/>> accessed 9 December 2021.

²² United Nations Convention on the Rights of the Child, 1989.

²³ Council Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of security network and information systems across the Union (NIS Directive) OJ L 194, 19.7.2016, p.1.

²⁴ Consolidated version of the Treaty on the Functioning of the European Union – Part Three: Union Policies and Internal Actions – Title VII: Common Rules on Competition, Taxation and Approximation of Laws – Chapter 3: Approximation of laws – Article 144 (ex Article 95 TEC).

ganda online.²⁵ A proposed law²⁶ defines harmful information and regulates digital services with an emphasis on online content management.²⁷

The Digital Single Market Directive Article 13²⁸ suggests that ISPs employ content identification technology to stop copyright breaches.²⁹ Large systems governing digital ecosystems and cross-border issues³⁰ present legal challenges for the EU.³¹ To address technological advancements³² and copyright

²⁵ Proposal for a Regulation of the European parliament and of the council on preventing the dissemination of terrorist content online, A contribution from the European Commission to the Leaders' meeting in Salzburg on 19-20 September 2018.

²⁶ Chiarella, L., 'Digital Services Act (DSA) and Digital Markets Act (DMA): New Rules for the EU Digital Environment' (2023) 9 Athens Journal of Law <<https://www.athensjournals.gr/law/2023-9-1-2-Chiarella>> accessed 29 December 2022.

²⁷ Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) PE/30/2022/REV/1 OJ L 277, 27.10.2022.

²⁸ Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market, COM/2016/0593, Article 13.

²⁹ Andrew Murray, *Information Technology Law: The Law and Society* (4th Edn, OUP 2019) 05.

³⁰ Council Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce) OJ L 178, 17.7.2000.

³¹ European Commission, *The Digital Service Act: Ensuring a safe and accountable online environment 2019* <https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en>.

³² David, E 'The EU AI Act passed-now comes the waiting / Delays in implementing the AI Act means nothing changes for now' (The Verge, 15 December 2023) <<https://www.theverge.com/2023/12/14/24001919/eu-ai-act-foundation-models-regulation-data>> accessed 15 December 2023.

concerns, the EU has proposed comprehensive regulation pertaining to AI.³³ When a platform violates the Digital Services Act, the EU Commission has the authority to charge up to 6% of worldwide earnings. Aiming to ensure a safer online environment and improve privacy rights,³⁴ companies such as Amazon oppose the DSA.³⁵ To protect consumers-especially children-and foster greater trust, the EU is modernizing rules,³⁶ regulating personal data, and creating compliance bodies.³⁷

Detection, Moderation and Legitimate Regulation

Creating global listings of hostile content for social media is challenging because of platform and lexicon differences.³⁸ Financial and technological con-

³³ Council Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

³⁴ Regulation (EU) 2016/679 (General Data Protection Regulation) in the current version of the OJ L 119, 04.05.2016; cor. OJ L 127, 23.5.2018.

³⁵ Emma Roth, 'The EU's Digital Services Act goes into effect today: here's what that means/The Digital Services Act forces companies to rethink their policies on advertising, transparency, and moderation' (The Verge, 25 August 2023) <<https://www.theverge.com/23845672/eu-digital-services-act-explained>> accessed 25 August 2023.

³⁶ Council Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provisions of audiovisual media service (Audiovisual Media Services Directive) in view of changing market realities PE/33/2018/REV/1 OJ L 303.

³⁷ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJ L 119, 4.5.2016.

³⁸ Siegel, A., 'Online Hate Speech' in Nathaniel Persily and Joshua A. Tucker (eds), *Social Media and Democracy: The State of the Field, Protects for Reform* (1st edn, Cambridge University Press, 2020) 56.

straints hamper governments' attempts to enforce laws against hate speech online. There are legislative recommendations available in the Rabat Plan to combat hate speech.³⁹ The Facebook oversight board has come under fire for abdicating accountability despite its legally enforceable rulings.⁴⁰ Violence has resulted from Facebook and other platforms' unequal resource allocation in locations like Myanmar.⁴¹ Global social media networks can reduce online harassment by implementing codes of conduct. Legal constraints make it more difficult to prosecute offensive social media posts.⁴² The UK deals with international computer-related crimes and makes disturbing communications illegal. International conventions forbid the promotion of violence and hatred.⁴³

Remedies

In order to safeguard privacy, Google is required by a ruling from the European Court of Justice to remove unnecessary personal information from search results. Libelous link removal efforts by websites such as "Solicitors from Hell" impact the impacted parties'

³⁹ United Nations Human Rights, Office of the High Commissioner, OHCHR and freedom of expression vs incitement to hatred: The Rabat Plan of Action 2023 <<https://www.ohchr.org/en/freedom-of-expression>>.

⁴⁰ Klonick, K., 'The Facebook Oversight Board: creating and Independent Institution to Adjudicate Online Free Expression' (2020) 129 (8) Yale Law Journals <<https://www.yalelawjournal.org/feature/the-facebook-oversight-board>> accessed June 2020.

⁴¹ Allen, C., 'Facebook's Content Moderation Failures in Ethiopia' (Council on Foreign Relations, 19 April 2022) <<https://www.cfr.org/blog/facebooks-content-moderation-failures->> accessed 19 April 2022.

⁴² Yahoo!, Inc. v La Ligue Contre Le Racisme, 169 F. Supp. 2d 1181 (N.D. Cal. 2001). See also, Telnikoff v Matusevitch [1990] 3 All ER 865, [1991] 1 QB 102, Bachchan v India Abroad Publications Inc (Sup. Ct. N. Y. Country 1992).

⁴³ European Union Agency for Criminal Justice Cooperation, Production Order under Article 18 of the Budapest Convention on Cybercrime and Extraterritorial Powers 2022 <<https://www.eurojust.europa.eu/publication/production-orders-under-article-18-budapest-convention-cybercrime-and-extraterritorial>>.

search engine rankings.⁴⁴ A proposal to update journalistic rules places a strong emphasis on maintaining anonymity when sharing public content. The goal of the movement to create new social norms is to strike a balance between social media and internet privacy and freedom of speech. Different interpretations of the Communication Decency Act (CDA) center legal discussions in the United States around ISPs' culpability for encouraging unwanted scrutiny.⁴⁵

Defenses

Lawsuits alleging defamation are filed against original writers, media organizations, users, editors, moderators, and anybody else disseminating false information. While preserving free speech, defamation cases are governed by UK legislation such as the Defamation Act (DA) 1996⁴⁶ and 2013.⁴⁷ There are different degrees of responsibility for user-contributed content across Twitter, Facebook, Google, and ISPs. US corporations are protected under the CDA but must enforce agreements. In the UK, the distributor's defense enables people or companies to assert that they took appropriate measures and were not the main producers of defamatory content. In addition to the privilege for legislative or judicial speech, the DA offers defenses based on truth, innocent dissemination, and honest opinion. Retraction and apology can lessen the harm. Website operators have protections under the DA and the EU E-Commerce Directive⁴⁸ from liability for user content. Focusing on defendants who possess significant financial resources, the 'deeper pocket' strategy has the potential to impact court decisions, particularly when it comes to major settlements.⁴⁹

⁴⁴ The Law Society v Kordowski [2011] EWHC 3185 (QB), Par. 43.

⁴⁵ The Communications Decency Act 1996, s 230.

⁴⁶ Defamation Act (DA) 1996, s 1.

⁴⁷ DA (n 12) 2013.

⁴⁸ Sparas, D., 'EU regulatory framework for e-commerce' (WTO, 17 June 2013) <<https://www.wto.org/english/tratop e/serv e/wkshop june13 e/sparas e.pdf>> accessed 18 June 2013.

⁴⁹ Cappelletti, M., Justifying Strict Liability: A Comparative

Jurisdictional Issues

According to the *Berezovsky v Forbes case*, the injury caused by a defamatory statement published abroad can reach a non-UK resident.⁵⁰ The term “Slander tourism” describes the international dissemination of discredit. This is limited in some ways by the Defamation Act.⁵¹ England and Wales must be considered as proper venues by courts based on the ‘appropriateness of jurisdiction.’ Claimants are granted preference in the UK, as demonstrated by the *Berezovsky case*. Online defamation in the UK happens when content is accessible rather than just preserved. Even if the publishing occurs elsewhere, a claimant who suffers significant harm in the UK may bring a lawsuit there. Even when the defendant’s content does not seem defamatory, defamation trials

The Importance of Privacy, Open Speech and Balancing Regulations

In contrast to isolation, privacy is the state of being anonymous. Instances of privacy violations include controlling access to someone’s data and controlling someone’s access them. The ECHR Article 10⁵² right for the public and press to freely transmit or get data for public benefit is being evaluated against the plaintiff’s Article 8 right to privacy.⁵³ The *Food Lion v ABC case*⁵⁴ highlights the detrimental effects of disclosing personal information and calls for stricter regulations, more robust privacy laws, digital literacy, the protection of whistleblowers, and international collaboration.⁵⁵

Analysis in Legal Reasoning (1st Edn, OUP 2022) 155.

⁵⁰ *Berezovsky v Forbes Inc* (No. 1) [2000] 1 WLR 1004.

⁵¹ DA 2013 (n 12), s 9.

⁵² Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) article 10.

⁵³ *ibid*, Article 8.

⁵⁴ 964 F. Supp. 956 [1997].

⁵⁵ Kindermann, D., ‘Against Hate Speech’ (2023) 40 (5) *Journal of Applied Philosophy* <<https://onlinelibrary.wiley.com/doi/10.1111/japp.12648>> accessed 6 April 2023.

Conclusion

Comprehensive regulations are necessary to combat hate speech online. Clearly defined terms that differentiate hate speech from legitimate speech should content moderation regulations mandating that platforms quickly remove any unlawful content. While transparency and reporting standards would guarantee that platforms reveal actions taken against harmful speech, duty of care regulations can hold them responsible for preventing harm. In addition to intermediary responsibility rules that define platforms’ responsibilities, cross-border collaboration is required to harmonize international standards. Remedy methods, like payments or legal action, ought to be available to victims of hate speech. To help the public identify and combat hate speech, governments can also encourage digital literacy. It is important to criminalize severe hate speech and impose suitable punishments on it. Additionally, independent regulatory bodies should be set up to keep an eye on compliance and impose penalties as needed. AI-based detection techniques should be used with stringent privacy protections, and supervision and appeals procedures would ensure equitable content moderation. Moreover, civil fines might enable plaintiffs to seek compensation, guaranteeing a more secure and civilized online space.

References

Statutes

1. Consolidated version of the Treaty on the Functioning of the European Union – Part Three: Union Policies and Internal Actions – Title VII: Common Rules on Competition, Taxation and Approximation of Laws – Chapter 3: Approximation of laws – Article 144 (ex Article 95 TEC).
2. Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) article 10.
3. Council Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016

concerning measures for a high common level of security of security network and information systems across the Union (NIS Directive) OJ L 194, 19.7.2016, p.1.

4. Council Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provisions of audiovisual media service (Audiovisual Media Services Directive) in view of changing market realities PE/33/2018/REV/1 OJ L 303.
5. Council Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.
6. Council Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce) OJ L 178, 17.7.2000.
7. Defamation Act (DA) 2013.
8. Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market, COM/2016/0593, Article 13.
9. Proposal for a Regulation of the European parliament and of the council on preventing the dissemination of terrorist content online, A contribution from the European Commission to the Leaders' meeting in Salzburg on 19-20 September 2018.
10. Regulation (EU) 2016/679 (General Data Protection Regulation) in the current version of the OJ L 119, 04.05.2016; cor. OJ L 127, 23.5.2018.
11. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJ L 119, 4.5.2016.
12. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) PE/30/2022/REV/1 OJ L 277, 27.10.2022.
13. RES/38/7 The Human Rights Council of 18 June-6 July 2018 the promotion, protection and enjoyment of human rights in the Internet.
14. The Communications Decency Act 1996, s 230.
15. United Nations Convention on the Rights of the Child, 1989.
16. United Nations International Telecommunication Union (ITU), Department of Economic and Social Affairs 2022.

Case Law

Berezovsky v Forbes Inc (No. 1) [2000] 1 WLR 1004.

1. Jameel v Dow Jones Inc [2005] EWCA Civ 75.
2. McAlpine v Bercow [2013] EWHC 1342 (QB).
3. Monroe v Hopkins [2017] EWHC 433 (QB).
4. The Law Society v Kordowski [2011] EWHC 3185 (QB), Par. 43.

Journals and Websites

- Ali, S., Haykal, H. A., and Youssef, E. Y. M., 'Child Sexual Abuse and the Internet: A Systematic Review' (2021) 4(1) *Human Arenas: Arena of Technologies* https://www.researchgate.net/publication/352120820_Child_Sexual_Abuse_and_the_Internet-A_Systematic_Review accessed 11 May 2021.
- Allen, C., 'Facebook's Content Moderation Failures in Ethiopia' (Council on Foreign Relations, 19 April 2022) <https://www.cfr.org/blog/facebooks-content-moderation-failures-> accessed 19 April 2022.
- Bostoan, F., 'Understanding the Digital Markets Act' (2023) 68(2) *Journal Indexing and Metrics* <https://journals.sagepub.com/doi/10.1177/0003603X231162998> accessed 12 April 2024.
- Cappelletti, M., *Justifying Strict Liability: A Comparative Analysis in Legal Reasoning* (1st edn, OUP 2022) 155.
- Casciani, D., 'Harry Miller: Legal Victory after Alleged Transphobic Tweets' (BBC News, 20 December 2021) <https://www.bbc.com/news/uk-england-lincolnshire-59727118> accessed 20 December 2021.
- Casino, F. et al., 'SoK: Cross-Border Criminal Investigations and Digital Evidence' (2022) 8(1) *Journal of Cybersecurity* <https://academic.oup.com/cybersecurity/article/8/1/typac014/6909060> accessed 25 May 2022.
- Castano-Pulgarin, S. A., Suarez-Betancur, N., Tilano Vega, L. M., and Herrera Lopez, H. M., 'Aggression and Violent Behavior: Internet, Social Media and Online Hate Speech. Systematic Review' (2021) 58(10) *Science Direct* <https://www.sciencedirect.com/science/article/abs/pii/S1359178921000628> accessed 6 April 2021.
- Chiarella, M. L., 'Digital Services Act (DSA) and Digital Markets Act (DMA): New Rules for the EU Digital Environment' (2023) 9 *Athens Journal of Law* <https://www.athensjournals.gr/law/2023-9-1-2-Chiarella>. accessed 29 December 2022.
- Council Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market (Directive on Electronic Commerce) OJ L 178, 17.7.2000.
- David, E., 'The EU AI Act Passed—Now Comes the Waiting: Delays in Implementing the AI Act Means Nothing Changes for Now' (The Verge, 15 December 2023) <https://www.theverge.com/2023/12/14/24001919/eu-ai-act-foundation-models-regulation-data> accessed 15 December 2023.
- European Union Agency for Criminal Justice Cooperation, Production Order under Article 18 of the Budapest Convention on Cybercrime and Extraterritorial Powers 2022 <https://www.eurojust.europa.eu/publication/production-orders-under-article-18-budapest-convention-cybercrime-and-extraterritorial>.
- Gunatilleke, G., 'Justifying Limitations on the Freedom of Expression' (2021) 22 *Human Rights Review* 02.
- Hatano, A., 'Regulating Online Hate Speech through the Prism of Human Rights Law: The Potential of Localized Content Moderation' (2023) 41(1) *ResearchGate* <https://www>.

- researchgate.net/publication/375636662 accessed 23 October 2023.
- Katzenbach, C. and Ulbricht, L., 'Algorithmic Governance' (2019) 8(4) *Internet Policy Review* <https://www.econstor.eu/handle/10419/210652> accessed 30 July 2024.
- Khan, T. et al., 'Fake News Outbreak 2021: Can We Stop the Viral Spread?' (2021) 190 *Journal of Network and Computer Applications* <https://www.sciencedirect.com/science/article/pii/S1084804521001326> accessed 15 September 2021.
- Kindermann, D., 'Against Hate Speech' (2023) 40(5) *Journal of Applied Philosophy* <https://onlinelibrary.wiley.com/doi/10.1111/japp.12648> accessed 6 April 2023.
- Kleinwachter, W., 'Cybersecurity, Internet Governance, and the Multistakeholder Approach' (2021) *Cyberstability Paper Series* <https://hcss.nl/report/cybersecurity-internet-governance-and-the-multistakeholder-approach-the-role-of-non-state-actors-in-internet-policy-making/> accessed 9 December 2021.
- Klonick, K., 'The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression' (2020) 129(8) *Yale Law Journal* <https://www.yalelawjournal.org/feature/the-facebook-oversight-board> accessed June 2020.
- Murray, A., *Information Technology Law: The Law and Society* (4th edn, OUP 2019) 05.
- Noble, C., *Extended from What?: Tracing the Construction, Flexible Meaning, and Cultural Discourses of "Extended Vocal Techniques"* (Santa Cruz ProQuest Dissertations & Theses 2019).
- Payne, B. K. and Pitman, L., 'Technology, White-Collar Cybercrime, and White-Collar Crime: An Analysis of Utah's White-Collar Crime Registry Data' (2022) 4(2) *Journal of White Collar and Corporate Crime* <https://journals.sagepub.com/doi/10.1177/2631309X221128622> accessed 22 September 2022.
- Pujol, J., 'Is This the End of Privacy? Snowden and the Power of Conscience' (2020) 5(1) *Church, Communication and Culture* <https://www.tandfonline.com/doi/full/10.1080/23753234.2020.1713017> accessed 17 February 2020.
- Roth, E., 'The EU's Digital Services Act Goes into Effect Today: Here's What That Means' (The Verge, 25 August 2023) <https://www.theverge.com/23845672/eu-digital-services-act-explained> accessed 25 August 2023.
- Siegel, A. A., 'Online Hate Speech' in N. Persily and J. A. Tucker (eds), *Social Media and Democracy: The State of the Field, Projects for Reform* (1st edn, Cambridge University Press 2020) 56.
- Sparas, D., 'EU Regulatory Framework for E-Commerce' (WTO, 17 June 2013) https://www.wto.org/english/tratop_e/serv_e/wkshop_june13_e/sparas_e.pdf accessed 18 June 2013.
- United Nations Human Rights, Office of the High Commissioner, OHCHR and Freedom of Expression vs Incitement to Hatred: The Rabat Plan of Action 2023 <https://www.ohchr.org/en/freedom-of-expression>.
- Yahoo!, *Inc. v La Ligue Contre Le Racisme*, 169 F. Supp. 2d 1181 (N.D. Cal. 2001). See also *Telnikoff v Matusевич* [1990] 3 All ER 865, [1991] 1 QB 102; *Bachchan v India Abroad Publications Inc* (Sup. Ct. N. Y. County 1992).

Zuboff, S., *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (PublicAffairs 2019).

Zuckerman, E., 'New Media, New Civics?' (2016) 6 *Policy & Internet* 151-168.