

Review

Not peer-reviewed version

---

# The Arrest of Telegram CEO Pavel Durov: A Crossroad between Free Speech and Content Moderation through the Lens of the Harm Principle

---

[SAMUEL DURAIVEL](#) \*

Posted Date: 28 August 2024

doi: 10.20944/preprints202408.2053.v1

Keywords: Harm Principle; Free Speech; Content Moderation; Digital Ethics; Telegram



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

# The Arrest of Telegram CEO Pavel Durov: A Crossroad Between Free Speech and Content Moderation Through the Lens of the Harm Principle

SAMUEL DURAIVEL

Loyola Institute of Social Research (LISor), India; duraivelsamuel@gmail.com

**Abstract:** The arrest of Pavel Durov, the CEO of Telegram, represents a critical moment in the ongoing discourse surrounding free speech and content moderation. This event exemplifies the tension between upholding individual liberties and addressing the potential societal harms caused by unregulated digital communication. This paper explores the implications of Durov's arrest and Telegram's stance on free speech through the lens of John Stuart Mill's Harm Principle. The Harm Principle, as articulated in Mill's seminal work *On Liberty* (1859), provides a valuable theoretical framework for examining the balance between individual freedom and societal protection. This paper argues that the application of the Harm Principle can offer a nuanced perspective on the ethical responsibilities of digital platforms like Telegram in the age of global communication.

**Keywords:** Harm Principle; free speech; content moderation; digital ethics; Telegram

## Introduction

The arrest of Pavel Durov, the CEO of Telegram, represents a critical moment in the ongoing discourse surrounding free speech and content moderation. This event exemplifies the tension between upholding individual liberties and addressing the potential societal harms caused by unregulated digital communication. This paper explores the implications of Durov's arrest and Telegram's stance on free speech through the lens of John Stuart Mill's Harm Principle. The Harm Principle, as articulated in Mill's seminal work *On Liberty* (1859), provides a valuable theoretical framework for examining the balance between individual freedom and societal protection. This paper argues that the application of the Harm Principle can offer a nuanced perspective on the ethical responsibilities of digital platforms like Telegram in the age of global communication.

## Methodology

This paper employs a literature review approach, guided by Mill's Harm Principle, to analyze the intersection of free speech, content moderation, and digital governance. Academic databases such as Google Scholar, JSTOR, and Taylor & Francis were used to source relevant literature. Keywords such as "Harm Principle," "free speech," "content moderation," "Telegram," and "Pavel Durov" were employed to select 25 peer-reviewed articles and scholarly books. These sources were analyzed to explore how the Harm Principle can be applied to the current challenges faced by digital platforms.

## Literature Review

### *Theoretical Framework: The Harm Principle*

John Stuart Mill's Harm Principle is a foundational concept in liberal thought, asserting that the actions of individuals should only be limited to prevent harm to others. Mill's argument for maximizing individual liberty is grounded in the belief that free expression is essential for personal development and societal progress. However, this freedom is not absolute; it must be curtailed when it causes harm to others (Mill, 1859). The principle provides a basis for evaluating the limits of free

speech, particularly in cases where speech may incite violence, spread misinformation, or otherwise harm individuals or society.

### *Historical Context: Telegram and Pavel Durov*

Telegram was founded by Pavel Durov in 2013, following his departure from VKontakte, Russia's largest social network, which he created in 2006. Durov's experience with VKontakte, where he resisted government demands for user data, deeply influenced the development of Telegram. The platform was designed with a strong emphasis on privacy, security, and resistance to censorship (Marechal, 2018). Telegram's encryption features and commitment to free speech quickly attracted a global user base, including activists, journalists, and dissidents. However, the same features that protect user privacy have also made Telegram a platform for illegal activities, raising ethical concerns.

### *The Harm Principle and Free Speech on Telegram*

Applying Mill's Harm Principle to the case of Telegram reveals the complexities of balancing free speech with the prevention of harm. On one hand, Telegram's commitment to free speech aligns with Mill's advocacy for individual liberty and resistance to censorship. The platform's encryption features provide a safe space for users to express their views without fear of government surveillance, which is particularly important in authoritarian contexts (Parida, 2021).

However, the unregulated nature of Telegram also allows for the spread of harmful content. Extremist groups, for instance, have used the platform to coordinate activities and disseminate propaganda (Looney et al., 2022). According to the Harm Principle, such content justifies intervention, as it poses a clear risk of harm to others. Mill's principle supports the idea that while free speech is valuable, it should not extend to actions or expressions that cause direct harm to individuals or society.

The challenge for Telegram, therefore, is to reconcile its commitment to free speech with the need to prevent harm. From a Harm Principle perspective, the platform has an ethical obligation to implement content moderation practices that mitigate the risks associated with harmful content, even if it means limiting some forms of expression (Mill, 1859).

### *Content Moderation: A Necessary Intervention?*

Content moderation on digital platforms like Telegram can be understood as a practical application of the Harm Principle. While Durov's commitment to privacy and free speech has been unwavering, the rise of harmful content on Telegram highlights the need for a more balanced approach. The Harm Principle provides a rationale for content moderation as a necessary intervention to prevent harm. However, the challenge lies in implementing moderation without infringing on the legitimate rights of users to free expression (Etzioni, 1993).

Telegram's approach to content moderation has been largely decentralized, placing responsibility on users and channel administrators. This method aligns with the platform's philosophy of minimal interference, but it has also allowed harmful content to proliferate (Tuck et al., 2023). The Harm Principle suggests that Telegram should adopt more robust moderation practices to prevent harm, especially in cases where speech incites violence or spreads misinformation. Such moderation does not contradict the principle of free speech but rather ensures that speech does not result in harm to others.

### *The Role of Governments and Global Governance*

The conflict between Telegram and various governments, particularly Russia, illustrates the tension between state sovereignty and the autonomy of digital platforms. Governments often invoke national security concerns to justify demands for data access and content control, while platforms like Telegram resist these demands in the name of free speech and privacy (Wijermars & Lokot, 2022). The Harm Principle can offer a framework for evaluating these conflicts, suggesting that state intervention is justified only when it seeks to prevent harm rather than suppress dissent.

However, the application of the Harm Principle in a global context is challenging. Different governments have different interpretations of what constitutes harm, and these differences complicate the development of a consistent approach to content moderation and free speech on global platforms (Zaliznyak, 2024). While the Harm Principle supports the idea of limiting speech to prevent harm, it also raises questions about who gets to define what harm is and how it should be mitigated.

#### *Telegram's Ethical Dilemma: Privacy vs. Harm*

Telegram's commitment to privacy and free speech creates an ethical dilemma when it comes to preventing harm. On one hand, the platform's encryption features protect users from government surveillance, aligning with the Harm Principle's emphasis on protecting individual liberty. On the other hand, these same features can be exploited by malicious actors, leading to harm that the platform is ill-equipped to address (Knockel et al., 2023).

From a Harm Principle perspective, Telegram has a moral responsibility to balance its commitment to privacy with the need to prevent harm. This might involve adopting more proactive moderation practices, particularly in cases where speech crosses the line into incitement or direct harm (Mill, 1859). However, any intervention must be carefully calibrated to avoid unnecessary infringement on legitimate free speech.

#### *The Future of Free Speech and Content Moderation on Telegram*

The future of free speech and content moderation on Telegram will likely involve navigating the tension between protecting individual liberties and preventing harm. As digital platforms continue to evolve, the Harm Principle provides a valuable framework for evaluating the ethical responsibilities of these platforms. Telegram's challenge will be to develop moderation practices that align with the Harm Principle while maintaining its commitment to free speech and privacy (Conway et al., 2023).

### **Conclusion**

The arrest of Pavel Durov and the ongoing debate over Telegram's role in free speech and content moderation highlight the relevance of John Stuart Mill's Harm Principle in the digital age. While Telegram's commitment to free speech is commendable, the platform must also address the potential harms that arise from unregulated communication. By applying the Harm Principle, Telegram can find a balance between protecting individual liberties and preventing societal harm. As digital platforms continue to shape global communication, the Harm Principle will remain a crucial tool for guiding ethical decision-making in the realm of free speech and content moderation.

### **References**

1. Conway, M., Mattheis, A. A., & McCafferty, S. (2023). *Violent extremism and terrorism online in 2023*. VoxPol. Retrieved from [https://voxpol.eu/wp-content/uploads/2024/07/YIR-2023\\_FINAL.pdf](https://voxpol.eu/wp-content/uploads/2024/07/YIR-2023_FINAL.pdf)
2. Etzioni, A. (1993). *The spirit of community: Rights, responsibilities, and the communitarian agenda*. Crown Publishers.
3. Knockel, J., Dalek, J., Meletti, L., & Ermoshina, K. (2023). *Not OK on VK: An analysis of in-platform censorship on Russia's VKontakte*. Tspace Library, University of Toronto. Retrieved from <https://tspace.library.utoronto.ca/handle/1807/129345>
4. Looney, S., Conway, M., & Watkin, A. L. (2022). *Violent extremism and terrorism online in 2021: The year in review*. Plymouth Research Commons. Retrieved from <https://plymouth.researchcommons.org/cgi/viewcontent.cgi?article=1235&context=sc-research>
5. Marechal, N. (2018). *From Russia with crypto: A political history of Telegram*. USENIX. Retrieved from <https://www.usenix.org/system/files/conference/foci18/foci18-paper-marchal.pdf>
6. Mill, J. S. (1859). *On liberty*. J.W. Parker and Son.
7. Parida, S. K. (2021). *Telegram revolution—an analysis of political instability of Belarus in 2020*. Vital Library. Retrieved from <https://vital.lib.tsu.ru/vital/access/services/Download/koha:000723248/SOURCE1>

8. Ternov, N., Nurtazina, R., & Serikzhanova, A. (2024). *The Sociopolitical January 2022 Protests in Kazakhstan's Telegram Channels: Agenda Interception*. Global Perspectives, University of California Press.
9. Tuck, H., Guhl, J., Smirnova, J., & Gerster, L. (2023). *Researching the evolving online ecosystem: Telegram, Discord and Odysee*. Casm Consulting. Retrieved from <https://files.casmconsulting.co.uk/researching-the-evolving-online-ecosystem-telegram-discord-odysee.pdf>
10. Wijermars, M., & Lokot, T. (2022). *Is Telegram a "harbinger of freedom"? The performance, practices, and perception of platforms as political actors in authoritarian states*. *Post-Soviet Affairs*, 38(4), 267-287.
11. Zaliznyak, Y. B. (2024). *Freedom of the press and national interests: Russian information aggression in Ukrainian information space*. *Global Journalism in Comparative Perspective*. Taylor & Francis.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.