The Problem with Automated Content Moderation

20 January 2023

Zeerak Talat
Floating
zeerak_talat@sfu.ca | @zeeraktalat

Through the darkness of future past, the magician longs to see, one chants out between two worlds, fire walk with me!

Leland Palmer, Twin Peaks (1990)

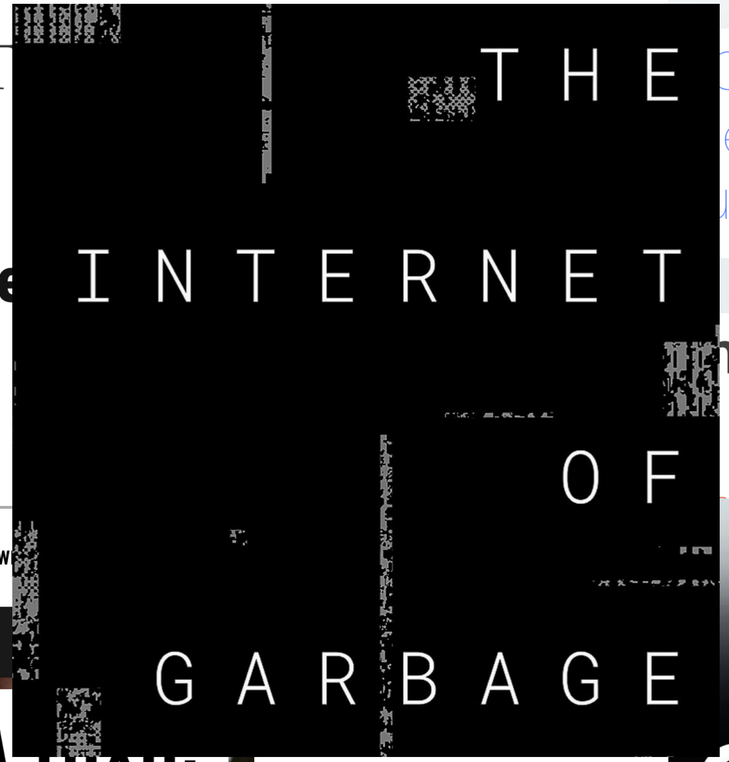
Clean up the Internet

Clean up the internet is an independent, UK-based organisation concerned about the degradation in online discours its implications for democracy. We campaign for evidence-based action to increase civility and respect online, and to online bullying, trolling, intimidation, and misinformation.

Defund Hate Speech: T
Clean Is Upon Us

Content on Digital Platforms is on
e. How Can Brands Rebuild
umer Trust?

Abuse, racism and hate spe
the comments online



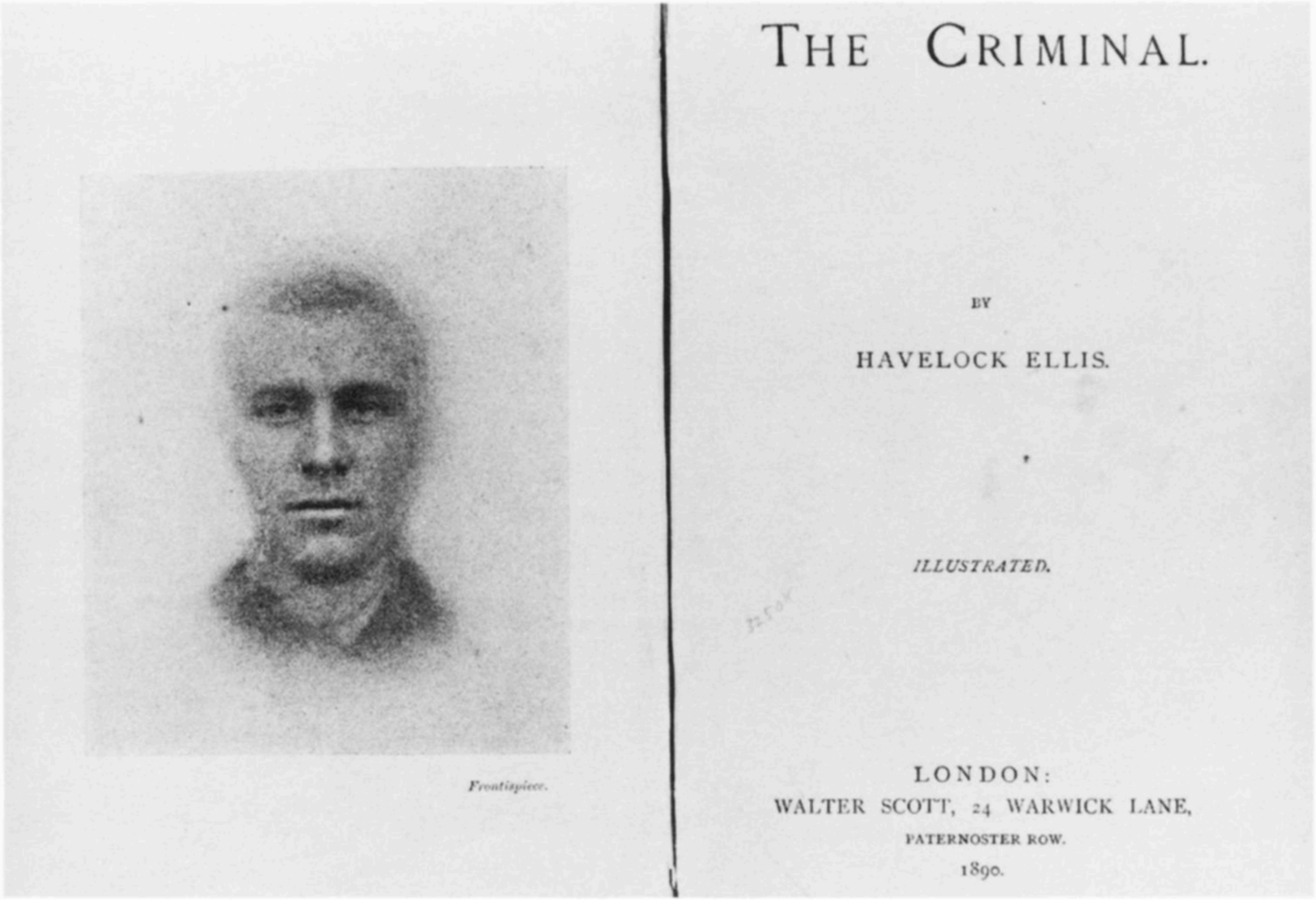
hilever warns social media
clean up "toxic" content

[I]deas about separating, purifying, demarcating and punishing transgressions have as their main function to impose system on an inherently untidy experience.

Mary Douglas, Purity & Danger (1978)

“Respectability politics upholds the idea that the supposed worthiness of a marginalized group should be evaluated—that is, by comparing the traits and actions of the marginalized group to the values of respectability set solely by the dominant group.”

Studio ATAO (n.d.)



Frontispiece.

THE CRIMINAL.

BY

HAVELOCK ELLIS.

ILLUSTRATED.

LONDON:
WALTER SCOTT, 24 WARWICK LANE,
PATERNOSTER ROW.
1890.

Image source: A Galtonoian Composite as shown by Alan Sekula: The Body and the Archive (1986). October. MIT Press

“Imperialism leaves behind germs of rot which we must clinically detect and remove from our land but from our minds as well.”

Franz Fanon, The Wretched of the Earth (19963)

References

1. Costanza-Chock, Sasha. “Design Justice, A.I., and Escape from the Matrix of Domination.” *Journal of Design and Science*, 2018.
2. Dias Oliva, Thiago, Dennys Marcelo Antonialli, and Alessandra Gomes. “Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online.” *Sexuality & Culture* 25, no. 2, 2021.
3. Douglas, Mary. *Purity and Danger: An Analysis of the Concepts of Pollution and Taboo*. Repr. London: Routledge, 1978.
4. Dunn, Jonathan. “Mapping Languages: The Corpus of Global Language Use.” *Language Resources and Evaluation* 54, no. 4, 2020.
5. Fanon, Frantz. *The Wretched of the Earth*. New York: Grove Press, 2002.
6. Foucault, Michel. *Archaeology of Knowledge*. Routledge Classics. London; New York: Routledge, 1969.
7. Fraser, Nancy. “Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy.” *Social Text*, no. 25/26, 1990.
8. Hall, Stuart. “Race, the Floating Signifier.” Lecture, Media Education Foundation, 1997
9. Hall, Stuart. “The Spectacle of the Other.” In Representation: Cultural Representations and Signifying Practices, Vol. 7. Sage London, 1997.
10. Kalluri, Pratyusha. “Don’t Ask If Artificial Intelligence Is Good or Fair, Ask How It Shifts Power.” *Nature* 583, no. 7815, 2020
11. Sekula, Allan. “The Body and the Archive.” *October* 39, 1986.
12. Studio ATAO. “Understanding Respectability Politics.” Studio ATAO. Accessed June 13, 2022.
13. Talat, Zeerak, and Anne Lauscher. “Back to the Future: On Potential Histories in NLP.” arXiv, 2022.