

recommendations



trafficking
victims



minors &
vulnerable



sex workers &
marginalized

sw users

platforms

legislation

peer policing

educate users on recognizing
trafficking vs consensual

advocacy

educate users on recognizing
trafficking vs consensual

resource referral

provide support to users who
may have been exploited online

collaborative policy

invite advocacy groups to
design moderation policies

regulate ai

enforce best practices and
implement safeguards for abuse

fund research

finance the exploration of
trafficking interventions

safe harbor protections

provisions for sex workers who
report trafficking crimes

decriminalization

target perpetrators, not
consensual workers

opt-in settings

require users to indicate they
agree to view explicit content

parental consent

prohibit children from creating
accounts without permission

teen-only accounts

create accounts with limited
exposure and features

age verification

include user birthdate in sign-up
to confirm appropriateness

flagger review

penalize abuse from users
targeting sex workers or minors

safety education

create digital literacy programs
focusing on trafficking risks

transparency & accountability

mandate reports from platforms
on moderation actions

community inquiry

encourage users to contact
platforms re: wrongful moderation

appeals process

establish independent oversight
boards for moderation appeals

content warning labels

clearly mark content made for
adult consumption only

lower false-positive rate

adopt human-in-the-loop
systems and balanced datasets

dedicated networks

develop spaces where adult
activity can be monitored

restorative justice

ensure violations of rights are
corrected and prevented

statutory liability

develop spaces where adult
activity can be monitored

repeal fosta/sesta

develop spaces where adult
activity can be monitored

feasible



ambitious



references

Blegen, E. N. (2023). Sex, Drugs, and Rock & Roll: Effectively and Equitably Moderating Vice and Illegal Content Online. *Vanderbilt Law Review*, 76:4, 1265–1298. <https://heinonline.org/HOL/License>

Decoding Content Moderation: Analyzing Policy Variations Across Top Online Platforms. (n.d.). <https://cs.uchicago.edu/news/decoding-content-moderation-analyzing-policy-variations-across-top-online-platforms/>

Douek, E. (2022). Content Moderation as Systems Thinking’ (2022) 136(2) *Harvard Law Review* 526 *MLA 9th ed.* Douek, Evelyn. In *Harvard Law Review* (Vol. 136, Issue 2). <https://heinonline.org/HOL/License>

Gillespie, T., Aufderheide, P., Carmi, E., Gerrard, Y., Gorwa, R., Matamoros-Fernández, A., Roberts, S. T., Sinnreich, A., & West, S. M. (2020). Expanding the debate about content moderation: Scholarly research agendas for the coming policy debates. *Internet Policy Review*, 9(4), 1–29. <https://doi.org/10.14763/2020.4.1512>

Inserra, D. (2024). A Guide to Content Moderation for Policymakers.

Monea, A. (2022). The Digital Closet 4. Pornotopia Bound License: Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC-BY-NC-ND 4.0).

Morrow, G., Swire-Thompson, B., Polny, J. M., Kopec, M., & Wihbey, J. P. (2022). The emerging science of content labeling: Contextualizing social media content moderation. In *Journal of the Association for Information Science and Technology* (Vol. 73, Issue 10, pp. 1365–1386). John Wiley and Sons Inc.

Peterst, J. (2020). Sexual Content and Social Media Moderation’ (2020) 59(3) *Washburn Law Journal* 469 *MLA 9th ed.* Peters, Jonathan. In *Washburn Law Journal* (Vol. 59, Issue 3). <https://perma.cc/4PZ8-7BXY>;

Schaffner, B., Bhagoji, A. N., Cheng, S., Mei, J., Shen, J. L., Wang, G., Chetty, M., Feamster, N., Lakier, G., & Tan, C. (2024, May 11). “Community Guidelines Make this the Best Party on the Internet”: An In-Depth Study of Online Platforms’ Content Moderation Policies. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3613904.3642333>

Seering, J. (2020). Reconsidering Community Self-Moderation: The Role of Research in Supporting Community-Based Models for Online Content Moderation. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2). <https://doi.org/10.1145/3415178>

The, B., & Malena Dailey, B. (n.d.). What Content Social Media Removes And Why.