SPI 594K: Issues in Content Moderation and Platform Governance

Spring 2021

Thursday 7-10 p.m. ET, Zoom

SPIA Policy Seminar
Princeton University

Instructor

Prof. Andy Guess
Fisher 213, aguess@princeton.edu
Student hours: Fridays (schedule on WASE)
Zoom: https://princeton.zoom.us/my/aguess

Course Overview

How should democratic societies respond to the amplification of propaganda, disinformation, and hate speech on digital forums designed to promote free expression? Lacking crucial evidence and facing political constraints, governments and technology companies have struggled to tailor their approaches. This half-semester course will cover practical, legal, and normative debates surrounding content moderation and governance on social platforms. Topics will include legal foundations such as CDA Section 230 ("platform immunity"); the role of algorithms and curation in ranking content; the promise of labeling, fact-checking, and other interventions designed to counter misinformation; and case studies, such as Facebook's Oversight Board.

Book

Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media.* New Haven: Yale University Press.

Course Components and Grading

- **Participation.** You will be expected to carefully complete the readings before class and engage thoughtfully in discussion. (30%)

- **Group mini-presentation.** At the beginning of one session, you will co-lead a group presentation (15-20 minutes long) exploring a case study that focuses on a policy by a government or a moderation dilemma involving a social platform that is *not* one of the "big four" (Facebook, Instagram, Twitter, YouTube). You will form approximately three-person groups in the first week. (15%)

- **Reaction pieces.** You will submit two short reaction pieces (2 to 3 pages, double-spaced) that engage with at least two of the readings from a given week. While you can spend a few sentences summarizing the main points, these pieces should primarily analyze or critique the arguments, identify tensions between them, and suggest constructive ways to synthesize or build on these works. These pieces are due the Wednesday before class at 6 p.m. ET. (20%)

- **Policy memo.** The final written assignment will take the form of a policy memo outlining and justifying a specific policy proposal or content moderation regime to relevant decision-makers. (35%)

Schedule

February 4: Introduction and Fundamentals

- Klonick, Kate. 2018. "The New Governors: The People, Rules, and Processes Governing Online Speech." *Harvard Law Review* 131(6), p. 1598.
- Gillespie, ch. 1
- Keller, Daphne and Leerssen, Paddy. 2020. "Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation." In Persily, Nathaniel and Tucker, Joshua (eds.), *Social Media and Democracy: The State of the Field, Prospects for Reform.* Cambridge: Cambridge University Press.

February 11: Speech

- Kosseff, Jeff. 2019. *The Twenty-Six Words that Created the Internet.* Ithaca: Cornell University Press. Introduction and pp. 57–78.

- King, Gary, Pan, Jennifer and Roberts, Margaret. 2014. "Reverse-engineering censorship in China: Randomized experimentation and participant observation." *Science* 345(6199).
- Read or watch Mark Zuckerberg's 2019 speech at Georgetown University.
- Bazelon, Emily. 2020. "The First Amendment in the age of disinformation." *The New York Times Magazine*.
- Newton, Casey. 2020. "Snap takes a stand." The Verge.

February 18: Technolog(ies)

- Gillespie, ch. 7
- Jiang, S., Robertson, R.E. and Wilson, C. 2020. "Reasoning about political bias in content moderation." In *Proceedings of the AAAI Conference on Artificial Intelligence* 34(9), pp. 13669–13672.
- Faddoul, Marc. 2020. "COVID-19 is triggering a massive experiment in algorithmic content moderation." Brookings Institution Tech Stream.
- Gorwa, R., Binns, R. and Katzenbach, C. 2020. "Algorithmic content moderation: Technical and political challenges in the automation of platform governance." *Big Data & Society*.
- Newton, Casey. 2019. "The Trauma Floor: The secret lives of Facebook moderators in America." The Verge.
- Oremus, Will. 2016. "Who Controls Your Facebook Feed." *Slate*.

February 25: Disinformation and Propaganda

- Nyhan, Brendan. 2020. "Facts and Myths About Misperceptions." *Journal of Economic Perspectives* 34(3): 220–236.
- Barari, Soubhik, Lucas, Christopher and Munger, Kevin. 2021. "Political Deepfake Videos Misinform the Public, But No More than Other Fake Media." Available at OSF: `https://osf.io/cdfh3/`
- Ternovski, John, Kalla, Joshua and Aronow, Peter. 2021. "Deepfake Warnings for Political Videos Increase Disbelief but Do Not Improve Discernment: Evidence from Two Experiments." Available at OSF: `https://osf.io/dta97/`
- Guess, A., Lerner, M., Lyons, B., Montgomery, J., Nyhan, B., Reifler, J., and Sircar, N. 2020. "A digital media literacy intervention increases discernment between mainstream and false news in the United States and India," *Proceedings of the National Academy of Sciences*.

March 4: Rights and Harms

- Müller, Karsten and Schwarz, Carlo. 2020. "Fanning the Flames of Hate: Social Media and Hate Crime." Available at SSRN: https://ssrn.com/abstract=3082972
- Laub, Zachary. 2019. Hate Speech on Social Media: Global Comparisons. Council on Foreign Relations.
- Special Rapporteur's 2018 report to the United Nations Human Rights Council.
- Sylvain, Olivier. 2018. "Discriminatory Designs on User Data." Knight First Amendment Institute.
- Douek, Evelyn. 2020. "COVID-19 and Social Media Content Moderation." Lawfare.

March 11: Governance and Democracy

- Gillespie, ch. 8
- Podcast: "Post no evil redux." 2020. Radiolab.
- Matias, J. Nathan. 2019. "Preventing harassment and increasing group participation through social norms in 2,190 online science discussions." *Proceedings of the National Academy of Sciences* 116(20), pp. 9785–9789.
- Citron, Danielle and Jurecic, Quinta. 2018. "Platform Justice: Content Moderation at an Inflection Point." The Hoover Institution.
- The Santa Clara Principles, https://santaclaraprinciples.org
- Oversight Board:
  - Charter: https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf
  - Douek, Evelyn. 2021. "Facebook Has Referred Trump's Suspension to Its Oversight Board. Now What?" Lawfare.
  - Case Decision 2020-005-FB-UA