History of Data: How Data Became Big (HIST 30705)

It is hard to find a realm of our life that hasn't been exposed to data-empowered algorithms. While Big Data and Machine Leaning became dominant approaches to artificial intelligence only in 1986, their histories have much deeper roots, which are entangled with the histories of statecraft, social engineering, and computing. While situating the intellectual origins of Big Data within these contexts, we will explore the following questions: How do people decide what information should be numerically recorded? How do they choose how to interpret numbers? How do social and cultural contexts play into their decisions? Do numerical methods of governance always guarantee just society? What are the intellectual and technological mechanisms that enable this technology to amply existing asymmetries in social power across the axis of gender, race, and nationality?

Learning Outcomes:

1) Understand a variety of historical contexts that have shaped how we use, collect, and analyze data in the 21st century;
2) Assess historical arguments about the place of data in the growth of state bureaucracies, national security, warfare, and scientific research made in scholarly secondary sources;
3) Use primary sources to make arguments about the historical contingency of the ways practitioners have been using data to solve state problems;

FREEDOM OF EXPRESSION POLICY

In this class, students are encouraged to exercise their right to free inquiry and expression. You are welcome to express any view on the subject matter introduced by the instructor or other class members within the structure of the course. While you are responsible for learning the content of this course, you remain free to take a reasoned exception to the views presented and to reserve judgment about matters of conscience, controversy, or opinion. When you encounter ideas that you find offensive, immoral, or unwise, you are encouraged to engage them with reasons, evidence, and arguments. Your course grade will be based on your academic performance, not on the opinions you express. Our commitment to freedom of expression means that no relevant ideas or positions are out of bounds, but disruptive or disorderly behavior, threats, or harassment are strictly prohibited and will be reported to the Office of the Dean of Students.

See the University's "Commitment to Freedom of Expression" and "Bill of Student Rights" in the University Policies and Statements module on Brightspace.

Grade Distribution:

1. Primary Source Analysis I: 100
2. Report on Experiential Exercise: 100
3. Primary Source Analysis II: 100
4. Midterm: 200 points
5. Op-Ed: 200
6. Responses to any ten secondary sources on Brightspace: 200 points (each response is 20 points)
7. Participation: 100 points

Reading List:

1. Adolphe Quatelet, "Preface" and "Introductory," A Treatise on Man (1842).
2. Francis Galton, "Regression Toward Mediocrity in Hereditary Stature," The Journal of the Anthropological Institute of Great Britain and Ireland Vol. 15 (1886), pp. 246-263
3. Emmanuel Didier, "The America of the Reporters," *Quantification, Democracy, and the Birth of National Statistics,* The MIT Press (2020), 19-85.
4. Marie Hicks, "War Machines: Women's Computing Work and the Underpinnings of the Data-Driven State, 1930-1940" in *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing,* 19-57 (MIT, 2017)
5. Galison, Peter. 1994. "The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision." *Critical Inquiry* 21 (1): 228–66.
6. Alan Turing, "Computing Machinery and Intelligence," Mind 59, no. 236 (1950)
7. Allen Newell, Herbert Simon, "Elements of the Theory of Human Problem Solving," in Psychological Review, Vol. 65, No. 3 (1958).
8. Andrew Lea, "MYCIN Explains Itself" in *Digitizing Diagnosis: Medicine, Minds, and Machines in the Twentieth-Century America.* John Hopkins University Press, 2023, 119-14.
9. Edward Feigenbaum, "Knowledge Engineering. The Applied Side of Artificial Intelligence," Stanford Heuristic Programming Project, September 1980.
10. Li, Xiaochang. ""There's No Data Like More Data": Automating Speech Recognition and the Making of Algorithmic Culture," *Osiris* (2023): 165-182.
11. Janet Abbate, J. (2001). Government, Business, and the Making of the Internet. *Business History Review,* 75(1), 147–176.

12. Benjamin Peters, “Staging the OGAS, 1962 to 1969” in *How Not o Network a Nation, 107-157* (MIT, 2016)
13. Dan Bouk, “Stories in the Data” in *Democracy’s Data: the Hidden Stories in the U.S. Census and how to Read Them*.
14. Sarah Igo, “The Record Prison” in *The Known Citizen. The History of Privacy in Modern America* (2018), 221-263
15. Matthew L. Jones, “Decision Trees, Random Forests, and the Genealogy of the Black Box” in *Algorithmic Modernity: Mechanizing Thought and Action, 1500-2000,* pp. 190-215. (Oxford, 2023)
16. The Great A. I. Awakening, the New York Times, 2016 https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html
17. Chris Wiggins and Matthew Jones ,“The Science of Data” in *How Data Happened*, 196-229 (Norton and Company, 2023).
18. Nick Seaver, “Captivating Algorithms: Recommender Systems as Traps,” *Journal of Material Culture*, Vol. 24, Issue 4, 2018.**
19. Vidan, G., & Lehdonvirta, V. (2018). Mine the gap: Bitcoin and the maintenance of trustless ness. *New Media & Society*, *21*(1), 42-59. https://doi.org/10.1177/1461444818786220
20. Michael Kearns and Aaron Roth, “Ethical Algorithm Design Should Guide Technology Regulation,” Brookings, January 13, 2020 https://www.brookings.edu/articles/ethical-algorithm-design-should-guide-technology-regulation/
21. boyd, danah, & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679.

FREEDOM OF EXPRESSION

In this class, students are encouraged to exercise their right to free inquiry and expression. You are welcome to express any view on the subject matter introduced by the instructor or other class members within the structure of the course. While you are responsible for learning the content of this course, you remain free to take a reasoned exception to the views presented and to reserve judgment about matters of conscience, controversy, or opinion. When you encounter ideas that you find offensive, unwise, immoral, or wrong-headed, you are encouraged to engage them with reasons, evidence, and arguments. Your course grade will be based on your academic performance, not on the opinions you express. Our commitment to freedom of expression means that no relevant ideas or positions are out of bounds, but disruptive or disorderly behavior, threats, or harassment are strictly prohibited and will be reported to the Office of the Dean of Students.