

Computational Social Science with Images and Audio

Elliott Ash, **Philine Widmer**

8 December 2023

Today, we move from “What is?” to “What ought to be?”.

- ▶ Most of the content discussed so far concerns positive questions
 - ▶ How does a given model work?
 - ▶ What are we measuring? Etc.
- ▶ Today, we will discuss the ethics of image- and audio-data related questions
 - ▶ For example, how should models be built from a social viewpoint?
 - ▶ How should we (not) use them? Etc.
 - ▶ How should visual or audio data (not) represent our society?
- ▶ Ethics discussion are necessarily influenced by moral considerations
 - ▶ What is good or bad? Just or unjust?

Scientists do not operate outside of society.

- ▶ Positive science is interested in discovery and facts (the “truth” about the world)
- ▶ However, positive science interacts with normative considerations in many ways, for example:
 - ▶ Which questions do we deem important to research?
 - ▶ What impact will our discovery have on society?
 - ▶ What cost are we willing to carry for discovery (not just monetary)?
 - ▶ What responsibilities do we (not) have as scientists?
 - ▶ Further examples?

Social science is connected to policies in important ways.

- ▶ Subjects of study typically relate to policies rather directly
- ▶ Policies are often linked to social constructs
 - ▶ → Social constructs: Products of human agreements and perceptions
 - ▶ For better intuition: what is the difference between the Matterhorn and Switzerland? What about money?
- ▶ Values and other context-specific characteristics interact with the entire social science pipeline:
 - ▶ Research question → (data collection) → theoretical or empirical analysis → interpretation
 - ▶ Examples?

Let us discuss some case studies. First up: census data.

A government wants to establish a comprehensive population registration system for administrative and statistical purposes. The system should simplify administration and offer perspectives for social research.

- ▶ Do you have any ethical concerns?
- ▶ Would you be more/less concerned if it is a company/researchers/an other actor constructing the comprehensive database? Why?

History has seen horrible examples of unanticipated secondary use.

- ▶ During the Second World War: Dutch census data used to facilitate the genocide against Jews and others:¹
- ▶ Data collection during peaceful times
- ▶ Trusted by citizens
- ▶ Later misuse by the Nazis
- ▶ Extremely high death rate of Dutch Jews: 73%

¹Seltzer, W., & Anderson, M. (2001). The Dark Side of Numbers: The Role of Population Data Systems in Human Rights Abuses. Social Research, 481-513.

Second case study: secret experiments and the like.

► Example 1: Encore²

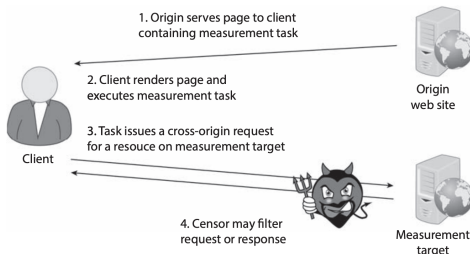


Figure: *Encore* by Burnett and Feamster (2015)

► What could be possible concerns?³

²Burnett, S., & Feamster, N. (2015). Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests. In Proceedings of the 2015 ACM conference on Special Interest Group on Data Communication (pp. 653-667).

³Further reading: Salganik, M. J. (2019). Bit by Bit: Social Research in the Digital Age. Princeton University Press. (In particular pp. 286-7)

Let us discuss further case studies in groups.

- ▶ Surveillance technology
 - ▶ Should governments use face recognition technology? If so, for which applications?⁴
 - ▶ How does this question relate to computational social science?
- ▶ Representation of different groups (e.g., by ethnicity, gender, age)
 - ▶ From a normative viewpoint, what is the “right” amount and kind of representation of different societal groups in the media?⁵
 - ▶ How does this question relate to computational social science?
- ▶ Accents and labor market outcomes
 - ▶ (Under what circumstances) should employers be allowed to consider a worker's accent in their hiring decision?⁶
 - ▶ How does this question relate to computational social science?

⁴Possible reading: Beraja, M., Kao, A., Yang, D. Y., & Yuchtman, N. (2023). Exporting the Surveillance State via Trade in AI. National Bureau of Economic Research Working Paper # 31676.

⁵For example, recall Elliott's presentation: Ash, E., Durante, R., Grebenschikova, M., & Schwarz, C. (2021). Visual Representation and Stereotypes in News Media.

⁶Possible reading: Grogger, J., Steinmayr, A., & Winter, J. (2020). The Wage Penalty of Regional Accents. National Bureau of Economic Research Working Paper # 26719.