

SEMINAR ON TOPICS IN POLITICAL ECONOMY:

SHAPING AI FOR SOCIETY

“A computer would deserve to be called intelligent if it could deceive a human into believing that it was human” – Alan Turing, 1950

I. TOPIC DESCRIPTION

The pervasion of AI-based applications continues to grow at a staggering pace. Already today, they affect many aspects of our lives. Thus, we are confronted with the timely question of how to design AI in ways that are optimal depending on context, users' and society's preferences.

AI promises tremendous opportunities for progress. How to harness these opportunities for the best of society, however, remains an open question. What are potential trade-offs? How can designs look like that work well according to specific measures of welfare? This is the topic of this seminar. Participants develop their own research idea on how to design AI to society's best interests, for example, with one specific and politically relevant context in mind. The goal is to come up with an experimental design (lab or field experiment) that enlightens us about some specific research question within the topic. To find such a context, an entry point may be one of the following.

Digitalization has dramatically reduced search, information and tracking costs, thus making it easier than ever for consumers to compare prices and product features. On the other hand, this development may contribute to information overload and 'attention wars'. What does that imply for market design?

Or, consider the effect of negligible tracking costs such that firms can design highly targeted advertising or implement first-order price discrimination easily. (E.g., consumers using an Apple device may be offered higher prices than others.) What should we think about this from an economic design perspective? Should we regulate these markets, and if so, how?

AI may help people to make better decisions, e.g. by giving nudges that promote healthier lifestyle choices and/or support long-term goals. But how should they become designed in optimal ways? Are there trade-offs? How should we measure welfare? Answers may be quite context-specific, of course, and it may make sense to focus on one specific one for a research question, e.g., how to nudge people to save money, eat healthy, drink less... And, do people enjoy such nudging in the short run? In the long run? Are there differences depending on personal characteristics?

Of course, AI might also give consumers back control over their data, e.g. via apps scanning the usually too long to read Terms & Conditions agreements and provide

consumers with the essential information. How could this be designed in a good way?

Moreover, some people fear that the rise of AI could replace many aspects of human work and thereby potentially cause mass unemployment. If so, how could markets and economics be designed such that societies prosper (nevertheless)?

Furthermore, algorithms can create filter bubbles and political echo chambers. Do we have something to say to this from an information design point of view?

Finally, people fear that AI could contribute to the emergence of totalitarian regimes.

“What Orwell prophesied in 1984, where technology was being used to monitor, control, dictate, or what Huxley imagined we may do just by distracting ourselves without any meaning or purpose. Neither of these futures is something that we want. [...] There are unintended consequences of technology. It is up to us to ensure that some of the more dystopian scenarios don’t come true.”

says Microsoft’s CEO Satya Nadella. Preventing an Orwellian dystopia from becoming reality will crucially depend on the strength and role of political institutions. E.g., consider the case of San Francisco which has recently banned facial recognition technology in public places. How could AI foster security and nevertheless respect aspects of privacy? Suggestions need to be developed and empirically evaluated.

II. SCOPE OF THE SEMINAR

Participants develop their own research question within the spectrum of the seminar. Students design an economic lab or field experiment that aims to answer their research question. Students will work in groups of two or three. If the process of group formation is unsuccessful, participants will be assigned.

The seminar starts with an **introductory meeting on Wednesday, October 16 at 10:30 in building 10.50, room 604**. Ideas for experiments or field studies will be presented in blocked events on January 13, 2020. Each presentation should last max. 20 minutes. Full attendance in all meetings is required for successful participation in the seminar. Papers of 8–10 pages are to be handed in before February 20, 2020. Grades will be based on the quality of presentations (50%), the seminar paper (50%). Students can improve their grades by 0.3 for good and constructive discussion contributions or by 0.7 for excellent and constructive contributions. Application will be possible via the centralized platform until July 8, 2019.

Please note further that we will ask you for a brief description of the topic and research idea you are interested in, in order to ensure a well-balanced variety of the seminar. Prior attendance of the courses “Economics and Behavior” and/or “Auction and Mechanism Design” is recommended but not mandatory. For further questions, please contact David Ehrlich (David.Ehrlich@kit.edu).

III. SOURCES OF INSPIRATION

A. Consumer empowerment

- Chernev, A., Böckenholt, U., & Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, 25(2), 333-358.
- Costa, E., & Halpern, D. (2019). The behavioural science of online harm and manipulation, and what to do about it. Retrieved from: <https://www.bi.team/publications/the-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it/>
- Gabaix, X. & Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *The Quarterly Journal of Economics*, 121(2), 505-540.
- Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, 57(1), 3-43.
- Mariotti, T., Schweizer, N., Szech, N., & von Wangenheim, J. (2018). Information nudges and self-control. [working paper, https://polit.econ.kit.edu/downloads/papers/MSSW_Information_Nudges.pdf]
- OECD. (2018). Personalised Pricing in the Digital Era. DAF/COMP(2018)13. Retrieved from: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DAF/COMP\(2018\)13&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DAF/COMP(2018)13&docLanguage=En)
- Schweizer, N., & Szech, N. (2018). Optimal revelation of life-changing information. *Management Science*, 64(11), 5250-5262.
- Soman, D. (2015). *The Last Mile: Creating Social and Economic Value from Behavioural Insights*. Toronto, Canada: University of Toronto Press.

B. The influence of algorithms on political debates & biases

- Braun, S., Dwenger, N., Kübler, D., Westkamp, A. (2014): Implementing quotas in university admissions: An experimental analysis. *Games and Economic Behavior*, 85, 232–251
- Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016). How we analysed the COMPAS recidivism algorithm. ProPublica. 23 May. Retrieved from: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>; Dressel, J. & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1), eaao5580.
- Schweizer, N., & Szech, N. (2017). Revenues and welfare in auctions with information release. *Journal of Economic Theory*, 170, 86-111.
- Staley, O. (2017). Harvard economist Iris Bohnet on how to design a hiring process that's fair for everyone. Quartz at Work. 17 October. Retrieved from: <https://qz.com/work/1080530/harvard-economist-iris-bohnet-says-to-eliminate-bias-its-easier-to-change-systems-than-change-people/>

- Sunstein, C. R. (2001). *Echo chambers: Bush v. Gore, impeachment, and beyond*. Princeton, NJ: Princeton University Press.
- Szech, N. (2011). Optimal advertising of auctions. *Journal of Economic Theory*, 146(6), 2596-2607.
- West, S.M., Whittaker, M. & Crawford, K. (2019). *Discriminating Systems: Gender, Race and Power in AI*. AI Now Institute. Retrieved from <https://ainowinstitute.org/discriminatingsystems.html>

C. Privacy/security

- Conger, K., Fausset, R., & Kovalski, S.F. (2019). San Francisco Bans Facial Recognition Software. *New York Times*, May 14, 2019. Retrieved from: <https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html>
- Orwell, G. (1949). *Nineteen eighty-four: A novel*. London: Secker & Warburg.