Literaturliste zum Seminar "Ethik und Wissenschaftstheorie des maschinellen Lernens" (Thomas Grote, Philipp Berens, Ulrike von Luxburg, Wintersemester 2017)

17.10. Session 1 Organisatorisches und Amuse Gueule.

Zur Einstimmung lesen Sie bitte die folgenden Zeitungsartikel:

- Zum Trolley-Problem: https://www.heise.de/newsticker/meldung/Ethik-bei-autonomen-Autos-und-das-Trolley-Problem-Was-tut-der-Weichensteller-3766885.html
- Digitalisierung von Arbeitsplätzen: http://www.zeit.de/karriere/2017-06/digitalisierung-arbeitsplaetze-maschinen-verantwortung-politik
- Bias-Thematik: http://www.sueddeutsche.de/digital/kuenstliche-intelligenz-wie-algorithmen-hass-und-vorurteile-zementieren-1.3620668

24.10. Session 2 Was ist Maschinelles Lernen?

Verpflichtende Lektüre:

- M. Jordan, T. Mitchel, 2015: Machine learning: Trends, perspectives, and prospects.
- Wheeler, Gregory (201): Machine Epistemology and Big Data. In L. McIntire und A. Rosenberg (Hrsg.): The Routledge Companion to Philosophy of Social Science. Routledge pdf.

Freiwillige Lektüre und weitere links:

- LeCun, Bengio, Hinton: Deep Learning. A review. Nature, 2015

31.10. Kein Seminar, Feiertag :-)

7.11. Session 3 Ansätze aus der Ethik

Verpflichtende Lektüre:

- Kurzer Public Science Artikel aus dem Guardian : Why are we reluctant to trust robots?
- Sinnott-Armstrong, Walter, "Consequentialism", *The Stanford Encyclopedia of Philosophy* (2015)
- Alexander, Larry and Moore, Michael, "Deontological Ethics", *The Stanford Encyclopedia of Philosophy* (2016)

Freiwillige Lektüre und weitere links: nichts bisher

14.11. Session 4 Wenn Algorithmen Informationen filtern: Algorithmenethik

Verpflichtende Lektüre:

- J. Burell: How the machine 'thinks': Understanding opacity in machine learning algorithms. 2016
- Wang, Kosinski: Deep neural networks can detect sexual orientation from faces. 2017.
- Diskussion über die Schlussfolgerung in einem blog

Freiwillige Lektüre:

- Mittelstadt et al. (2016): The Ethics of Algorithms: Mapping the Debate, in Big Data & Society
- Report "The Ethics of Algorithms: from radical content to self-driving cars". Report by Center for internet and human rights, 2015.
- Predictive policing:
    - Pro: Mara Hvistendahl: Can 'predictive policing' prevent crime before it happens? Science online, 2016
    - Contra: Blog by Emily Thomas: Why Oakland Police Turned Down Predictive Policing

---

21.11. Session 5 Einfluss von künstlicher Intelligenz auf die Gesellschaft

Verpflichtende Lektüre:

- Kleinberg, Ludwig, Mullainathan: A Guide to Solving Social Problems with Machine Learning, 2016.
- Lakkaraju, Rudin: Learning cost-effective and interpretable treatment regimes for judicial bail decisions. AISTATS 2017. (zum Überfliegen)
- Cath, Wachter, Mittelstadt, Taddeo, Floridi: Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. Science and Engineering Ethics, 2017. (zum Überfliegen)

Freiwillige Lektüre und weitere links:

- Surden: Machine Learning and Law. 2014.
- Link: ICML 2016 workshop "data for good"
- Link to some positive promises of AI for society can be found on the AI for good summit

---

28.11. Session 6 Fairness, Diskriminierung und das Recht auf Erklärung

Verpflichtende Lektüre:

- Equality of Opportunity in Supervised Learning NIPS 2016, Moritz Hardt, Eric Price, Nati Srebro.
- Bryce Goodman, Seth Flaxman: European Union regulations on algorithmic decision-making and a "right to explanation". https://arxiv.org/abs/1606.08813

Freiwillige Lektüre:

- Link to the EU regulation (in any language you want). The critical part starts on page 46 of the english version (Article 22).
- Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. NIPS 2016 link to the paper SZ-Artikel dazu: hier
- Es gab alle möglichen workshops zu diesen themen:
    - Link to icml 2016 workshop on interpretability
    - Link to nips 2016 workshop for interpretable ML
    - Link to icml 2016 workshop on reliable ML
    - Link to nips 2016 workshop on reliable ML

5.12. Seminar fällt aus!

12.12. Session 7 Wenn Algorithmen Entscheidungen treffen: Moral machines

Verpflichtende Lektüre:

- Susan Leigh Anderson (2008): Asimov`s "three laws of robotics" and machine metaethics, in: AI and Society 22 (4), S. 477-93.
- Moral Decision Making Frameworks for Artificial Intelligence.
- Bericht der Ethik-Kommission "Automatisiertes und vernetztes Fahren", Juni 2017. (zum Überfliegen)

Freiwillige Lektüre und links:

- MIT applet for moral machines: link

19.12. Session 8 Einführung in die Wissenschaftstheorie

Verfplichtende Lektüre:

- Woodward, James, "Scientific Explanation", *The Stanford Encyclopedia of Philosophy, 2017. Edition.*

Freiwillige Lektüre und weitere links: nichts

9.1. Session 9 Beispiele für ML in den Wissenschaften

Verpflichtende Lektüre:

- **Personalized medicine:**

    - Esteva et al. Dermatologist-level classification of skin cancer using DNNs (link)
    - Leachman & Merlino: The final frontier in cancer diagnosis (commentary on Esteva et al.) (link)

- **Social Science:**
    - Jean, Burke, Xie, Davis, Lobell, Ermon: Combining satellite imagery and machine learning to predict poverty. Science, 2016.
    - Cynthia Rudin: Can Machine Learning Be Useful for Social Science? 2015.

- **Psychology**: Smartphone Based Recognition of States and State Changes in Bipolar Disorder Patients. *IEEE Journal of Biomedical and Health Informatics, 2015.*

Freiwillige Lektüre:

- **Physics:** Machine learning phases of matter. Nature Physics, 2017. Link
- **Linguistics:** Word embeddings, eg described in this blog (have seen them before, in the paper "Man is to Computer Programmer as Woman is to Homemaker")

16.1. Session 10 Die Implikationen von ML für die Wissenschaften

Verpflichtende Lektüre:

- Anderson: The End of Theory. Wired, 2008.
- Kritisch hierzu Pietsch: Aspects of Theory-Ladenness in Data-Intensive Research. Meeting of the Philosophy of Science Association, 2014.

Freiwillige Lektüre und weitere links: nichts

23.1. Session 11 Modelle

Verpflichtende Lektüre:

- Alkistis, Elliot Graves u. Michael Weisberg (2014): Idealization, in: Philosophy Compass 9(3): 176-185
- Winsberg: A Tale of Two Methods , 2009 (login "machine", passwd "learning")
- Healy: Fuck Nuance**.** Sociological Theory, 2017.

Freiwillige Lektüre und weitere links: nichts

30.1.. Session 12 Verstehen und die Erkenntnisgrenzen von ML

Verpflichtend:

- Lynch: Understanding and the Digital Human, Kapitel 8 in The Internet of Us (2016). (login "machine", passwd "learning")
- Elgin (2007): Understanding and the Facts . Philosophical Studies, 2007.

Freiwillig:

- Dreyfus (2002):Why Heideggerian AI Failed and How Fixing it Would Require Making it More Heideggerian *Artificial Intelligence* (2007)

6.2. Session 13 Privatheit und ML

Verpflichtend:

- Barocas und Nissenbaum: Computing Ethics Big Data's End Run Around Procedural Privacy Protections. Recognizing the inherent limitations of consent and anonymity.
- Chaudhuri et al: Signal Processing and Machine Learning with Differential Privacy: Algorithms and Challenges for Continuous Data. 2013.
