FDH3004 Transparency in Technical and Social Systems 7.5 credits

Transparens i tekniska och sociala system

This is a translation of the Swedish, legally binding, course syllabus.

Establishment

Course syllabus for FDH3004 valid from Autumn 2019

Grading scale

P, F

Education cycle

Third cycle

Specific prerequisites

The course can be taken by PhD students from all research disciplines but aims first and foremost at PhD-students in EE and CS.

Language of instruction

The language of instruction is specified in the course offering information in the course catalogue.

Intended learning outcomes

After completing the course, the PhD-student will:

- Be acquainted with notions of knowledge and explanation, including their limitations, in different fields of study.
- Be able to analyze and critically examine different notions of transparency, suitable for different fields of study.
- Be able to analyze and critically examine what can and cannot reasonably be achieved through transparency in different contexts.
- Be acquainted with issues of trust and accountability related to transparency.
- Be able to plan research on transparency within their own field.

Course contents

Today, automated decision-making pervades our daily lives. Algorithms detect spam e-mail, recommend books, assign credit-scores, drive vehicles, and diagnose disease. This has prompted a vigorous public debate on the pros and cons of algorithmic "black boxes" as well as an interest in how they can be made more transparent. For example, the General Data Protection Regulation (GDPR) grants people the right, in certain circumstances, to get "meaningful information about the logic involved" in "automated decision-making" based on their personal data. However, achieving transparency can be difficult, and the consequences are not always easy to foresee.
The course starts out with examining notions of knowledge and explanation with a particular emphasis on similarities and differences between social and technological sciences. We then proceed to examine how we, as humans, are ourselves transparent and opaque and contrast this with technological transparency and opacity. We then consider how it is possible to make use of opaque systems (in spite or because of their opacity) and examine how transparency or the lack thereof affects incentives and markets. The next step is to acquaint ourselves with the issues of trust and accountability that relate to transparency, before the course concludes with examining cases when transparency might be misguided.
Throughout the course, participants will work on small research proposals for transparency research in their own fields, and in the end, participants will act as discussants of each other's proposals.

Course structure

Three course activities will be interleaved:
1. Literature seminars, in total 7 seminars.
2. Guest lectures (1-2) by external lecturers.
3. Opposition seminar, discussing research proposals.

Course literature

- Hollis, Martin (2002). 'Introduction: problems of structure and action' (pp. 1—22) in The philosophy of social science. An introduction. Revised edition. Cambridge University Press.
- Simon, Herbert A. (1996). 'Understanding the Natural and Artificial Worlds' (pp. 1—24) in The Sciences of the Artificial, MIT Press.

- Jeffrey, Richard C. (1969). 'Statistical Explanation vs. Statistical Inference' (pp. 104—113) in Rescher, Nicholas (ed.) Essays in Honor of Carl G. Hempel. Synthese Library, vol 24. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-1466-2_6
- Tversky, Amos and Kahneman, Daniel (1974). Judgment under uncertainty: Heuristics and biases. Science 185.4157, pp. 1124-1131. https://doi.org/10.1126/science.185.4157.1124
- Sorensen, Roy (2004). 'Paradoxes of rationality'. In Mele, Alfred. (ed.) Oxford Handbook of Rationality, Oxford University Press, Oxford, pp. 257-77.
- Anita Avramides (2010). 'Skepticism About Knowledge of Other Minds' in Bernecker, Sven and Pritchard, Duncan (eds.) The Routledge Companion to Epistemology, Routledge. https://doi.org/10.4324/9780203839065.ch40
- Fleischmann, Kenneth R. and Wallace, William A. (2005). A covenant with transparency: Opening the black box of models, Communications of the ACM, May, 2005, Vol.48(5), pp. 93—97. https://doi.org/10.1145/1060710.1060715
- Guidotti, Riccardo; Monreale, Anna; Ruggieri, Salvatore; Turini, Franco; Giannotti, Fosca and Pedreschi, Dino (2018). A Survey of Methods for Explaining Black Box Models. ACM Comput. Surv. 51, 5, Article 93 (August 2018), 42 pages. https://doi.org/10.1145/3236009
- Walach, Harald (2012). 'Double-Blind Procedure' (pp. 387—389) in Salkind, Neil. J. (ed.) Encyclopedia of research design Thousand Oaks, CA: SAGE Publications, Inc. https://doi.org/10.4135/9781412961288
- Arnold, Frances H. (1998). When blind is better: protein design by evolution. Nature biotechnology 16.7 : pp. 617—618. https://doi.org/10.1038/nbt0798-617
- Foyer, Pernilla (2015). 'General Discussion' (pp. 37—42) in Early Experience, Maternal Care and Behavioural Test Design*!* Effects on the Temperament of Military Working Dogs (PhD dissertation). Linköping University Electronic Press, Linköping. https://doi.org/10.3384/diss.diva-122260
- Anderson, Ross (2007) 'Open and Closed Systems Are Equivalent (That Is, in an Ideal World)' (pp. 127—142) in Feller, Joseph; Fitzgerald, Brian; Hissam, Scott A. and Huff, Karim R. (eds.) Perspectives on Free and Open Source Software, MIT Press. https://ieeexplore.ieee.org/document/6277068
- Akerlof, George A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism, The Quarterly Journal of Economics, vol. 84, No. 3 (Aug., 1970), pp. 488-500. https://doi.org/10.2307/1879431
- Bushman, Robert and Landsman, Wayne (2010). The pros and cons of regulating corporate reporting: A critical review of the arguments, Accounting and Business Research, Vol.40(3), pp.259-273. https://doi.org/10.1080/00014788.2010.9663400
- O'Neil, Cathy (2018). 'Introduction' (pp. 1-14), Weapons of Math Destruction, Broadway Books.
- Zerilli, John; Knott, Alistair; Maclaurin, James and Gavaghan, Colin (2018). Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? Philos. Technol. https://doi.org/10.1007/s13347-018-0330-6
- de Laat, Paul B. (2018). Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability? Philos. Technol. 31: pp. 525—541. https://doi.org/10.1007/s13347-017-0293-z
- Lessig, Lawrence (2009). Against transparency, The New Republic, Oct 21, 2009, Vol.240(19), pp. 37—44.

- Prat, Andrea (2005). The Wrong Kind of Transparency, The American Economic Review, Vol. 95, No. 3 (Jun., 2005), pp. 862—877. https://www.jstor.org/stable/4132745
- Schneier, Bruce (2019). There's No Good Reason to Trust Blockchain Technology, WIRED, https://www.wired.com/story/theres-no-good-reason-to-trust-blockchain-technolog

Examination

- EXA1 - Examination, 7.5 credits, grading scale: P, F

Based on recommendation from KTH's coordinator for disabilities, the examiner will decide how to adapt an examination for students with documented disability.

The examiner may apply another examination format when re-examining individual students.

If the course is discontinued, students may request to be examined during the following two academic years.

To complete the course, participants must:

- Read the literature and actively participate in the seminars (missed seminars can be compensated by written literature summaries).
- Hand in smaller assignments preparing for the final research proposals.
- Hand in small research proposals for transparency research in their own fields.
- Act as discussants of each other's proposals in a final seminar.

Other requirements for final grade

- Passed seminar participation.
- Passed assignments.
- Passed research proposal.
- Passed opposition.

Ethical approach

- All members of a group are responsible for the group's work.
- In any assessment, every student shall honestly disclose any help received and sources used.
- In an oral assessment, every student shall be able to present and answer questions about the entire assignment and solution.