FDH3004 Transparens i tekniska och sociala system 7,5 hp

Transparency in Technical and Social Systems

Fastställande

Vice forskarutbildningsansvarig vid EECS-skolan har 2019-10-16 beslutat att fastställa denna kursplan att gälla från och med HT 2019 (diarienummer J-2019-2710).

Betygsskala

P, F

Utbildningsnivå

Forskarnivå

Särskild behörighet

Doktorander inom alla forskningsdiscipliner är välkomna, men den huvudsakliga målgruppen är doktorander inom elektroteknik och datavetenskap.

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Efter genomgången kurs ska doktoranden:

- Vara bekant med begreppen kunskap och förklaring, inklusive deras begränsningar, inom olika vetenskapsområden.
- Kunna analysera och kritiskt granska olika transparensbegrepp som är lämpliga inom olika vetenskapsområden.
- Kunna analysera och kritiskt granska vad som rimligen kan och inte kan uppnås med transparens i olika kontexter.
- Vara bekant med frågor om tillit och ansvar relaterade till transparens.
- Kunna planera forskning om transparens inom sitt eget vetenskapsområde.

Kursinnehåll

Idag är vi ständigt omgivna av automatiserat beslutsfattande. Algoritmer hittar skräppost, ger bokrekommendationer, sätter kreditbetyg, kör fordon och diagnosticerar sjukdomar. Detta har lett till en livlig offentlig debatt om för- och nackdelar med algoritmer som liknar ”svarta lådor” och till ett stort intresse för hur de kan bli mer transparenta. Den europeiska dataskyddsförordningen (GDPR) ger exempelvis människor en rätt, under vissa omständigheter, att få ”meningsfull information om logiken bakom” ”automatiserat beslutsfattande” som bygger på deras personuppgifter. Men att uppnå transparens kan vara svårt och konsekvenserna är inte alltid lätta att förutse.
Kursen börjar med att granska begreppen kunskap och förklaring med särskild tonvikt på skillnader mellan samhälls- och teknikvetenskap. Vi fortsätter sedan med att granska hur vi själva, som människor, är mer eller mindre transparenta och jämför detta med tekniska system. Därefter studerar vi hur man kan dra nytta av icke-transparenta system (trots eller tack vare deras brist på transparens) och granskar hur transparens eller dess motsats påverkar incitament och marknader. Nästa steg är att bekanta oss med frågor om tillit och ansvar relaterade till transparens innan kursen avslutas med att granska fall där transparens kanske är missriktad.
Under kursens gång utarbetar deltagarna en liten forskningsplan för en transparensstudie inom sitt eget forskningsområde och opponerar i slutet på varandras planer.

Kursupplägg

Kursen kommer varva tre aktiviteter:
1. Litteraturseminarier, totalt 7 seminarier.
2. Gästföreläsningar (1-2) av externa föreläsare.
3. Opponeringsseminarium på forskningsplaner.

**Kurslitteratur
**

- 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 - Tentamen, 7,5 hp, betygsskala: P, F

Examinator beslutar, baserat på rekommendation från KTH:s handläggare av stöd till studenter med funktionsnedsättning, om eventuell anpassad examination för studenter med dokumenterad, varaktig funktionsnedsättning.

Examinator får medge annan examinationsform vid omexamination av enstaka studenter.

När kurs inte längre ges har student möjlighet att examineras under ytterligare två läsår.

För godkänt betyg måste deltagarna

- Läsa litteraturen och aktivt delta på seminarierna (missade seminarier ersätts med skriftliga sammanfattningar av litteraturen).
- Lämna in mindre inlämningsuppgifter som steg på vägen mot en slutlig forskningsplan.
- Lämna in en liten forskningsplan för en transparensstudie inom sitt eget forskningsområde.
- Opponera på varandras planer vid ett slutseminarium.

Övriga krav för slutbetyg

- Godkänt seminariedeltagande.
- Godkända inlämningsuppgifter.
- Godkänd forskningsplan.
- Godkänd opponering.

Etiskt förhållningssätt

- Vid grupparbete har alla i gruppen ansvar för gruppens arbete.
- Vid examination ska varje student ärligt redovisa hjälp som erhållits och källor som använts.
- Vid muntlig examination ska varje student kunna redogöra för hela uppgiften och hela lösningen.