Finance & Philosophy: The ethics and governance of algorithmic decision making

Type: Blockseminar
Number: 50130
Contact eMail: mail@carstenjung.net (Carsten), mm822@cam.ac.uk (Marco), h.m.veluwenkamp@rug.nl (Herman)
Participants: P&E Bachelor, P&E Master, BA IWB, BA Economics, MA IWB, MA Economics
Lecturers: Marco Meyer, Carsten Jung, Herman Veluwenkamp
Dates: 12-14 May 2017
Start-time: 12 May, 12.00 s.t.
Rooms: TBA

Deadline for submitting reading questions: 9 May.
Deadline for workshop sign up: 16 April. (We have 3 parallel workshops on Saturday afternoon)

Seminar description
Human decision makers are increasingly replaced by self-learning computer algorithms in economic decision making. Such algorithms use personal data (ranging from one's post code to medical history and Facebook friends) to score and rank people. Such rankings, in turn, can determine who gets offered a job, who gets offered a loan and under what conditions people can insure themselves against illness. Algorithms hold the promise of overcoming human bias and make smarter and more informed decisions. But can we really leave a large number of economic decisions to algorithms? How can we make sure they make good decisions? And how do we ensure they make fair decisions?

In this seminar, we investigate some novel ethical questions raised by algorithms. Our aim is to develop a framework to think through the merits and dangers of using algorithms in decision making, and to decide how the use of algorithms should be governed. We start by developing a model of human decision making, differentiating between the phases of information gathering, belief-formation, and judgement. In a second step, we investigate how self-learning computer algorithms function, and compare our model of human decision making to the way algorithms process information. We will investigate in detail how scoring algorithms in finance, health care, and human resources already inform critical decisions. This allows us to weed out ethical concerns based on misunderstandings of how algorithms work, and to identify truly new problems raised by algorithmic decision making, in contrast to problems that are raised by any kind of decision maker.

Lecturers
Marco Meyer
Marco has a PhD in philosophy from the University of Cambridge and is currently reading for a PhD in economics at the University of Groningen. He received a Master's degree in philosophy from Oxford University, a BA degree in Philosophy & Economics from Bayreuth university, and a

BA in European History from Bayreuth University. His research interests include financial ethics, applied epistemology and global justice. He has taught students at the universities of Oxford and Cambridge, and has given seminars on philosophy and economics.

Carsten Jung
Carsten is an economist at the Bank of England's International Directorate. He currently works on policy issues concerning the global financial vulnerabilities and systemic risks from climate change. He holds a BA in Philosophy & Economics from Bayreuth, an MSc in Economics from Warwick University and an MSc in Economic Sociology from LSE. He was also a Carlo-Schmid intern at the International Monetary Fund.

Herman Veluwenkamp
Herman is a PhD candidate at the University of Groningen. He is also is working at a IT company that specialises in developing software for mobile operators. He received a Master's degree in philosophy, a BA in Philosophy of Computer Science and a BSc in Computer Science. He has taught a wide range of seminars to students at the University of Groningen and is currently teaching Ethics to non-philosophers at the University College in Groningen. His research interests include meta-ethics and the logic of belief-revision.

Learning Goals

- Understand what algorithms are, and how machine learning works.
- Investigate how artificial intelligence processes information and makes decisions.
- Explore to what extent algorithms shape the life prospects of all of us, particularly in finance, health, social media, and the job market.
- Compare artificial intelligence decision making to human decision making.
- Reflect on whether decisions reached with the support of artificial intelligence can be challenged and whether algorithms can be held accountable.
- Investigate risks for morally problematic decisions by algorithms.
- Reflect about whether and how algorithms can incorporate ethical guidelines or be assessed according to ethical criteria.

Seminar Plan, Readings, and Reading Questions

Friday

12.00 (we start s.t.)	**Welcome**, Expectations, Intro to Seminar Topics Reading: Bostrom et al., The Ethics of Artificial Intelligence, only sections 1 and 2 (pp. 1-5) *Q: What are some of the new ethical challenges posed by artificial intelligence?*
13.00	**Algorithms in action (3 cases):** 1. Credit Scoring 2. Criminal Justice 3. Autonomous cars Reading: Diakopoulos - Algorithmic Accountability, pp. 1-15 *Q: What are some similarities and differences between the way laws and algorithms exercise power?*
14.30	**Introduction to AI & machine learning, Part I** Reading 1: Russell & Norvig (2016) - AI as the science of agent design (pp. 34- 59) *Q: In your opinion, which characteristics of artificial agents, described in the chapter, most closely resemble those of humans? Explain.* Reading 2: Casella et al (2015a): Introduction to statistical learning [read entire chapter] *Q: What is the problem with overfitting data - isn't it great that the training MSE is low? (especially pp. 31-32)* *Q: Can you think of a practical application (say, done by a company) of linear regression [parametric] approach and one of the KNN classifier [non-parametric] approach? Describe. (especially pp.37-42)*
16.00	**Introduction to AI & machine learning, Part II** Reading 1: Casella et al (2015b):Linear model selection (pp. 203-214) *Q: Describe one of the model selection algorithms in the chapter and describe what dangers you see, if any, iuf this process is completely automated.* Reading 2: Goodfellow et al (2016): Introduction to Deep Learning

	Q: What is deep learning and why are artificial neural networks helpful for it?
18.00	**End**

Saturday

09.00 (we start s.t.)	**Guarding against Algorithm Bias: The problem of Discrimination** Reading 1: Altman - Discrimination (SEP), only sections 1-5 (pp. 1-16), Reading 2: ProPublica - How We Analyzed the COMPAS Recidivism Algorithm *Q: When is discrimination wrongful, and is the way propublica measures discrimination a case in point?*
11.15	**How to teach ethics to algorithms?** Reading 1: Anderson - The Unacceptability of Asimov's Three Laws of Robotics as a Basis for Machine Ethics *Q: Why are Asimov's 3 Laws insufficient for machine ethics* Reading 2: Abel et al. - Reinforcement Learning as a Framework for Ethical Decision Making *Q: Try to formulate advantages reinforcement learning has over rival frameworks (such as rule-based systems)*
12.45	**Lunch**
14.00	**Key themes in AI and philosophy - workshops (split into 3 groups):** ● Detecting discrimination in algorithmic decision making ● Value learning and wireheading ● Can algorithms make good decisions?
17.00	**Workshop presentations:**
18.00	**Official end**
20.00	**Optional: screening of the film 'Her' + Pizza**

Sunday

09.00 (we start s.t.)	**The difficulty of regulating blackbox algorithms** Reading 1: Pasquale (2015): Search, Transparency and Judgement [7 page excerpt from “The Black Box Society"] Reading 2: Selbst & Baroquas (2017): Regulating inscrutable systems, p. 16-36 Q: *Do you think the firms’ public responses to accusations (described by Pasquale) are in line with the spirit of ‘giving explanations’ (as described, for other areas of law, by Selbst & Barocas)? What did you think their public responses?*
11.00	**The outlook - 3 short perspectives on the role of algorithms in the economy** Readings: Pasquale (forthcoming) [only pp. 4-10], Ohm and Reid (2016) [only pp.1689 - 1695] and Scholz (2016) [only pp 55-61] Q: Summarise very briefly the three different approaches the authors propose to regulate algorithms.
12.00	**Lunch**
13.00	Governing algorithms: Revisit the cases from criminal justice, finance and autonomous cars
15.00	**Feedback**
16.00	**End**

Workshops (Saturday afternoon)
On Saturday afternoon, we will split into three workshops exploring different key themes on the philosophy of artificial intelligence.

A. The Workshop on **Detecting discrimination in algorithmic decision making** looks at how algorithms can be screened for discrimination and discusses a philosophical account of what makes discrimination wrong.
B. The Workshop on **Value learning and wireheading** explores different techniques to get an AI to *learn* the values that humans want it to pursue and the potential problems with this.
C. In the workshop **Can algorithms make good decisions?** we (i) discuss whether we can equip algorithms with a decision theory that is reliably good and (ii) what makes algorithms different from human decision makers, including the role of ‘having reasons’.

Please sign up for one of the workshops here (deadline: 16 April 2017).

Assessment
P&E Bachelor: V, P4, P6, P7, P9; PE Master: Elective

Attendance-only "Schein" (PE: 2 ECTS): You need to (1) do the readings, (2) do the problem set, (3) participate actively in the seminar

Essay "Schein" (PE: 6/8 ECTS): You need to (1) do the readings, (2) do the problem set, (3) write an essay (4.000-5.000 words). If you give a presentation, this will feed into your grade as well.

Readings & Reading Questions

- There will be very few presentations in the seminar, and there won't be presentations to recap readings, so it is crucial that you do the readings.
- All Readings come with reading questions. Please submit your reading questions via this form**.**
- **Deadline for Submitting Reading Questions: 9 May**
- Some of the readings are hard. Rather than giving you watered-down readings, we want to encourage you to develop a reading skill that will become very handy in your academic life: extracting relevant information and arguments from very difficult texts.
- To do this, always keep the respective reading question in mind, and focus on getting an answer to that question out of the text.
- Do not despair if there are parts of the readings that you don't understand.

Presentations
We have three student presentations, all in the session 'Algorithms in action'. The three topics are:

1. Credit Scoring
2. Criminal Justice
3. Autonomous cars

For good presentations, we give an essay grading bonus of one grade-step (e.g. from 1.7 to 1.3). Alternatively, presenters can get a discount on essay length. If you would like to do one of the presentations, please send an eMail to Herman (deadline: 16 April).

General guidelines for presentations:

- Stick to the presentation topic
- Stick to the time limit: 10 min
- A projector will be available
- Presentations should be in English, but don't be intimidated — we'll have an atmosphere which is quite tolerant of less than perfect English; what matters is that you can get your point across.

Essays

Topics: We'll run a short session on possible seminar topics during the Blockseminar, and point out possible seminar topics during the seminar. You are encouraged to come up with your own topics.

General Guidelines

- Between 4,000 and 5,000 words
- English or German
- Before you start writing, look at this essay writing guide: http://www.phil.cam.ac.uk/curr-students/IA/curr-students/writing-skils/
- Re-read and revise before submitting. Ask a friend to read your paper.
- We set a deadline for the essay with students individually to fit their needs.

General Structure

- Define and answer a precise research question (check with us!), give a clear answer to the question which you state in the introduction (your thesis), and use your essay to defend that answer.
- Structure essay according to thesis: No paragraph and indeed no sentence that does not help investigating the thesis belongs in the essay.
- When developing an argument, state premises and conclusion clearly.
    - Make sure the argument is valid
    - Anticipate objections to your premises and discuss in the essay
- When using empirical data, explain and scrutinize it
    - Make sure the data is relevant to your argument and explain how that is
    - Indicate the source
    - Do not only report results, but also show awareness of the methodology used
- Summarize your argument in a brief conclusion

We encourage you to look for additional literature. You are not expected to read everything there is on your topic, however. In depth analysis and development of your own sustained argument matter much more. So, read a few things, but then start writing quickly.