Student Manual

AI for an Open Society

2025-2026

Universiteit Utrecht

Course Coordinators:	Dr. Leendert van Maanen &	Dr. Daan van Rooij
Contact:	l.vanmaanen@uu.nl	d.vanrooij@uu.nl

Course description

Artificial intelligence (AI) models, technology, and methods are heavily influencing society. This goes both through informal and formal institutions. For example, for informal institutions, algorithms estimate what people 'like' and dynamically adapt what posts they see on social media and even news websites. This can create 'filter bubbles' and a societal divide. On the formal side, models are one of the instruments to inform institutional and governmental policy and to create behavior change. For example, cognitive models of human attention can inform policy on how frequently to display speed information above the Dutch Highways, and agent-based models of social interaction can help understand how quickly diseases spread through a network and how this can be slowed down. In essence, AI models impact society at various levels. To shape tomorrow's society, students need to understand how the different types of models work, and how they can impact human behavior, and thereby society.

Learning goals

At the end of this course, students:

1. *Understand* how different disciplines, models, and theories contribute to AI for an open society. Particular emphasis will be placed on theories from the social and behavioral sciences, humanities, and governance.
2. Can *analyze* and *evaluate* concrete problems related to formal or informal institutions with an eye for how the aforementioned disciplines, models, and theories can contribute to it, and where opportunities are for multi- and interdisciplinary research.
3. Can *create* a research proposal to investigate a societal relevant problem. This proposal will require the application of AI techniques through a multi- and interdisciplinary perspective, as well as articulating its relevance to formal and/or informal institutions.

The intended *level* of the course is that students gain a sufficient understanding of the literature from the domains at hand that they can further explore particular elements in other courses, or by taking part in seminars organized by the Human-centered AI focus area of Institutions for an Open Society strategic theme.

For the grant proposal (see below), the intended *level* is that it can form the basis for continuation in an individual BSc project outside of the course, if desired and if feasible within the students' bachelor program.

Teaching structure

This course is taught by researchers and teachers from the Utrecht Focus Area Human-centered AI (esp. the special interest groups on Social and Cognitive Modeling and on AI, Government, and Behavior) and the Utrecht Strategic Theme Institutions for Open Societies (IOS). The aim is to show students how multi- and interdisciplinary research at the intersection of Human-centered AI and Institutions for Open Societies can benefit research and society. Particular emphasis is placed on contributions that are made from the social and behavioral sciences (Faculty FSW), humanities (Faculty GW), and governance (Faculty REBO).

The course set-up is as follows. In ***methodological*** lectures, students gain an understanding of the available tools from social sciences and humanities to study the impact of AI on IOS themes. In ***thematic*** lectures, students are provided with examples of how AI solutions impact society, and how these impacts can be mitigated. In ***practical*** working groups, these models and theories can be applied in a grant proposal. The examination consists of a midterm exam (testing concepts; individually), an oral pitch of an initial idea (group work), and a final grant proposal (group work).

The teaching staff and the students will ensure an interdisciplinary environment. Developing social AI models requires knowledge of both engineering science and social science. Moreover, for the connection to IOS, social science and humanities principles need to be applied in a governance context. The teachers will bring this integration. In addition, in the group work, students from different faculties will work together on an assignment. Course completion will require integration of knowledge, but also developing even wider interdisciplinary competencies (e.g., cooperating in an interdisciplinary team, shared leadership, formulating common ground, reflection).

Examination

1. Learning goals 1 and 2 will be examined through a mid-term, written exam (individual).
2. Learning goal 2 will also be tested through the development of an oral pitch of a research idea (group work).
3. Learning goal 3 will be examined through writing up a grant proposal (group work).

For examination elements 2 and 3, students will work together in groups. The lecturers aim to put together students that come from different bachelor programs, so students can learn from each other's discipline.

To pass the course, all three elements need to be passed. Students that fail the mid-term exam are offered a resit; students that fail the pitch or the grant proposal are offered an individual written assignment. The final grade will be the weighted average of the grades for the three examination elements: 60% mid-term, 5% pitch, 35% grant proposal.

Mid-term exam

The mid-term exam consist of 9 essay questions, covering the lectures as well as the obligatory literature. The mid-term exam is scheduled for **Friday May 29th, 11:00**. The allotted time will be 2 hours, unless you have an exemption. (Double)check mytimetable.uu.nl for the time and location as the date approaches!

Oral Pitch

The pitch consists of a short presentation of your group in which you pitch the idea you have for the grant proposal. The pitch is graded, but above all an opportunity to get valuable feedback on your idea.

Grant proposal

The grant proposal is the final assignment of this course. In the grant proposal, you (mock) pitch your research plan to an investor. The grant proposal is a group assignment, the groups are predetermined by the lecturers. Further details can be found in the document ***Grant Proposal assignment*** on Teams***.***

The deadline for handing in the grant proposal is **Sunday June 28th** end of day**!**

Schedule

The color coding refers to the lecture type: Methodological, thematic, or practical.

Lecture					Literature	Professor
	Date	Time	Week			
1	Wed 22-04	13:15 – 15:00	17	Introduction & IOS	*1*	Van Rooij
2	Fri 24-04	11:00 – 12:45	17	Modeling individual decision making	*2*	Van Maanen
3	Wed 29-04	13:15 – 15:00	18	Autonomy, psychology of interaction with AI systems	*3*	Hortensius
4	Fri 01-05	11:00 – 12:45	18	Emergence of collective patterns in relation to norms	*4*	Klein
5	Wed 06-05	13:15 – 15:00	19	Attitude & Linguistic Models	*5*	Van der Vegt
6	Fr 08-05	11:00 – 12:45	19	Trust in AI	*6*	Grimmelikhuijsen + Liefooghe
7	Wed 13-05	13:15 – 15:00	20	Medical AI / Digital Twins	*7*	Van Rooij + Bontje
8	Wed 20-05	13:15 – 15:00	21	Putting everything together and future outlook		Van Rooij
	Fri 29-05	**11:00 – 13:00**	**22**	**Mid-term exam**		
8	Fri 05-06	11:00 – 15:00	23	Workgroup		Van Rooij, Van Maanen, Liefooghe, v/d Vegt, Klein
9	Fri 12-06	11:00 – 15:00	24	**Pitches +** Workgroup		Van Rooij, Van Maanen
10	Fri 19-06	11:00 – 15:00	25	Workgroup		Van Rooij, Van Maanen
	Fri 26-06		26	Workgroup		Homework group, no location
	Sun 28-06		**26**	**Deadline proposal 23:59**		

Obligatory literature

1. Elliott, K., Price, R., Shaw, P. et al. (2021) Towards an equitable digital society: Artificial Intelligence (AI) and Corporate Digital Responsibility (CDR). *Society, 58*, 179–188. doi: 10.1007/s12115-021-00594-8
2. Palada, H., Neal, A., Vuckovic, A., Russell, M., Samuels, K. & Heathcote, A. (2016). Evidence accumulation in a complex task: Making choices about concurrent multiattribute stimuli under time pressure. *Journal of Experimental Psychology: Applied, 22* (1), 1-23. doi: 10.1037/xap0000074
3. Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., ... Wellman, M. (2019). *Machine behaviour. Nature*, 568, 477–486. doi:10.1038/s41586-019- 1138-y
4. Douven, I. & Hegselmann, R. (2021), Mis- and disinformation in a bounded confidence model, *Artificial Intelligence* 291, 103415, doi:10.1016/j.artint.2020.103415
5. van der Vegt, I., Kleinberg, B., & Gill, P. (2023). Proceed with caution: on the use of computational linguistics in threat assessment. *Journal of Policing, Intelligence and Counter Terrorism,* 18(2), 231-239.
6. Grimmelikhuijsen, S., & Meijer, A. (2022). Legitimacy of algorithmic decision-making: Six threats and the need for a calibrated institutional response. Perspectives on Public Management and Governance, 5(3), 232-242.
7. Wang, W., He, F., Li, Y., Tang, S., Li, X., Xia, J., & Lv, Z. (2023). Data information processing of traffic digital twins in smart cities using edge intelligent federation learning. *Information Processing & Management*, 60(1), 103171