STRUCTURES: Applying Responsible AI to Reduce Inequality

Note: This course manual may be updated during the course. Updates will be communicated via Canvas as well. Make sure you always use the most recent version of this manual. See final page of this manual for a list of updates.

Title of Course
STRUCTURES: Applying Responsible AI to Reduce Inequality

Course Catalogue Number
738200002Y

Credits
30 EC

Examiners
Prof. dr. S. (Sennay) Ghebreab (s.ghebreab@uva.nl)
Prof. T. (Tobias) Blanke (t.blanke@uva.nl)

Entry Requirements
The entry requirements for the course are specified in the Course Catalogue, available online at
https://studiegids.uva.nl/xmlpages/page/2023-2024/zoek-vak/vak/108120

Instruction Language
English

Time Period(s)
Academic year 2023-2024, semester 2

Location
Please see the UvA schedule webpage https://rooster.uva.nl/

Video Recording of Lectures
None

Core Lecturers
dr. Roland Adorjani
dr. Lukas Ansteeg
dr. Roberto Cerina
dr. Kunhe Li
dr. Kirils Makarovs

dr. Regina Nockerts (semester coordinator)
dr. Steve Pickering
dr. Petter Tornberg

Course Content

You will study political science theories to assess structural issues, building on the knowledge that was acquired in the Foundation, Building Blocks, and Connections courses. You will further deepen your knowledge of qualitative and quantitative research methods and acquire skills in machine learning and prediction.

A major part of the semester is organised around a group project (theme: social inequality and AI). Students will work in small project groups to address a real-world problem identified by CSSci staff and using real-world data. In the course of this project, students will:
- Design a viable research question that could be approached by the means of qualitative and quantitative research methods;
- Work with a given dataset to create robust predictions;
- Assess that prediction from a critical perspective and implement techniques to reduce bias;
- Create professional presentations of their project and its deliverables;
- Present the intervention plan to the CSSci community.

Additionally, there will be five individual assignments to reinforce key aims of the course:
1. Essay on AI and Social Inequality;
2. Qualitative Prediction Framework;
3. Ethics Position Statement;
4. Coding Portfolio and Explainer Session;
5. Individual contribution to Group Project V.

Please note that college week 40 (June 03 - 07) is reserved for collaborative group work. Although there may not be formal classes scheduled during this period, students shall be present, prepared, and participating in group work during this time. **Absence during non-designated vacation periods is incompatible with the program's essence** (with exception for extenuating circumstances, see the Force Majeur policy)."

Teaching Methods/Learning Formats

Teaching and learning take place through a set of formats, including lectures, seminars, discussion meetings, and group work. You can find additional information on Canvas.

Course Evaluations & Adjustments of the Course

This is a new course.

Professional Study Attitude

Computational Social Science is organised along a "just in time learning" approach, which means that the weekly content is directly applicable to the project work at that time. This means that it is crucial to participate in all the sessions (check-in, workshops, lectures, and so on) and all scheduled meetings and events with external partners. Your well-prepared, on-site participation is expected for all learning activities and inherently reflected in the programme's exit qualifications. In the event of a learning activity that you have missed, it is your responsibility to make up for the content that was covered during your absence. In case of absence, you must always inform your Core Lecturer beforehand or as soon as possible.

Requirements

In the Computational Social Science programme, students work together, in student groups and with external partners. You must therefore demonstrate a professional study attitude (the four P's):

1. Present
2. Prepared
3. Participating
4. Proper and clear communication

After all, absence or low-quality participation from one team member will impair the ability of other group members to excel and perform well. Additionally, a student's own learning is inhibited by an insufficient professional study attitude.

Consequences

General lack of participation

If a student cannot take part in a learning activity, it is their responsibility to catch up on the contents that were covered and always inform their Core Lecturer beforehand or as soon as possible.

If the Core Lecturer notices patterns of absence or low-quality participation inhibiting a student's learning process, the Core Lecturer will request an explanation from the student. In some cases, the Core Lecturer may refer the student to the programme's Study Adviser to seek guidance.

In case the Core Lecturer does not observe any improvement in performance, they will schedule a meeting with the student to determine:

- If the student can still meet the participation requirement;
- Which pattern of professional study attitude the student needs to develop;
- In which timeframe this improvement in performance needs to be sufficiently demonstrated.

The Core Lecturer will make a reasonable attempt to schedule this meeting with the student; should the student be unresponsive or a time and location cannot be agreed upon in a timely manner, the Core Lecture will instead issue a written warning covering the same points.

A summary of this meeting will be shared with the student via email and counts as a first written warning. In case the student has not sufficiently demonstrated improvement in performance during the set timeframe, a second written warning is issued. This email is also shared with the examiners of the semester course and the programme's Examinations Board.

Lack of adequate further improvement of performance or accountability from the student will lead to an 'Unacceptable' result for the individual assignment 'Individual Contribution to Group Project'. As a consequence, the student will not be able to complete the semester course.

Group work

If, in the informed judgement of the Core Lecturer, the (lack of) participation of a student in a group project is such that the student has not contributed to the project at all stages of its development, that student alone may receive an 'Insufficient' result for the group assignment.

Failure to fulfil the participation requirements, failure to demonstrate professional study attitudes, or continuing and serious disruptive behaviour may also result in a student being removed from their project group or not assigned to new project groups. This student receives an 'Unacceptable'

result for the individual assignment 'Individual Contribution to Group Project' and will not be able to complete the semester course.

<u>Lateness Policy</u>
Teaching and learning and students and staff participating in these activities, are disrupted and distracted when persons arrive after the starting time (late). In case a teaching and learning activity has commenced (e.g. if the door is closed and/or if a person is given instructions or a presentation), a student arriving late should not enter the room or activity. In this case the student will need to wait until the break before entering the room.

Force Majeur

When students are unable to fully participate because of reasons beyond their control, they should notify the Study Adviser and their Core Lecturer as soon as possible. Notification should be given ahead of time if the absence is foreseeable, or as soon as possible in case of unforeseen circumstances such as illness. The programme will make serious efforts to facilitate these students and prevent unnecessary delays. If students are not able to participate in graded assessments due to reasons beyond their control; or they wish to ask for a deadline extension or permission to participate in the resit/repair opportunity without having submitted a serious attempt first, they need to file a documented request with the Examinations Board (see the "Deadline Extension Request Procedure for Graded Assignments" section, below).

Assessment

In the Computational Social Science programme, assessment is geared towards learning. Most assignments are formative, which means that the students receive progress check reports to further improve and learn, but do not receive a grade. In addition, there are summative and therefore graded assignments. These assignments receive a descriptive grade but – in principle - no feedback beyond an explanation of the grade.

Based on the UvA descriptive scheme, Computational Social Science uses the following grading scheme: Excellent, Good, Satisfactory, Sufficient, Insufficient, and Unacceptable. See: https://student.uva.nl/en/topics/assessment-of-your-grade-point-average

Grades will be determined in no longer than 15 working days.

The main part of each semester is the group work. In addition, each student is assessed on the four learning trajectories Social Sciences and Humanities Expertise, Change Making Expertise, Research Expertise, and Digital Expertise. **All students are required to score at least sufficient for these individual assessments to pass the course**.

Students are also individually graded for their contribution to group projects. This assessment is based on roles and responsibilities as specified in a Team Charter, and the student's active and professional participation in group activities.

In case a group receives an insufficient final grade for the project(s), students can compensate this with their individual assignments. This is calculated using the grade points attached to the UvA descriptive grades (see link above) and the weights mentioned in the table below.

All assignments need to be handed in through Canvas. Inspection of the graded work will be possible, by appointment with the relevant Core Lecturer in the week following the release of grades.

Graded Assessments	Date / Deadline	Weight (%)	Sufficient Required (Yes/No)
Individual assignment			
DE Coding Portfolio and Explainer Session	Tuesday, April 2, 2024 (5 p.m)	10	Yes
SSH Essay on Algorithms and Social Inequality	Wednesday, May 15, 2024 (5 p.m)	10	Yes
RE Qualitative Prediction Framework	Monday, June 3, 2024 (5 p.m.)	10	Yes
CME Ethics Position Statement	Monday, June 3, 2024 (5 p.m)	10	Yes
Individual contribution to Group Project V	Thursday, June 6, 2024 (5 p.m)	10	Yes
Group assignment			
Group Project V: The Unequal Machines Grand Challenge	Thursday, June 6, 2024 (5 p.m)	50	No

Rules Regarding Resits

In the case of an "Insufficient" for a graded assessment following a serious attempt, one resit or repair possibility will be offered before the end of the semester. If no serious attempt has been made by a student, the grade will be registered as "Unacceptable," and no resit or repair opportunity will be possible.

Note that there are two weeks reserved before the end of the semester for repairs or resits. **The grade for all repair or resit assignments is capped at "Sufficient."**

Eligible Graded Assessments	Repair or Resit?	Date / Deadline
DE Coding Portfolio and Explainer Session	Repair	Code Portfolio: Friday, June 12, 2024 (5 p.m) Explainer Session: Friday, June 14, 2024 (by appointment)
SSH Essay on Algorithms and Social Inequality	Repair	Friday, June 24, 2024 (5 p.m)
RE Qualitative Prediction Framework	Repair	Wednesday, June 24, 2024 (5 p.m)
CME Ethics Position Statement	Repair	Wednesday, June 24, 2024 (5 p.m)
Individual contribution to Group Project V	Resit: Reflection for Personal Improvement; stretch contribution to Group Project V.	Friday, June 24, 2024 (5 p.m)

Description of Graded Assignments

Group Project V: The Unequal Machines Grand Challenge

The Unequal Machines Grand Challenge project is designed for you to engage simultaneously with

computational social science and inequality research. You are encouraged to study the complex dynamics of inequality through the lens of rich datasets and computational methodologies. This description aims to provide an understanding of the structure and requirements of this project.

The final project deliverable will consist of:
- Demonstration of a socially responsible predictive algorithm;
- Written audit report demonstrating potential biases in the analysis and the measures that were taken to address those biases;
- Code and data submission.

Individual Contribution to Group Project

When participating in a group project, it is important for each individual student to contribute effectively to the success of the project. This requires a professional attitude towards one's colleagues in the group, challenge partners and lecturers, as well as commitment to fulfilling the expectations outlined in the Team Charter. The Team Charter is a document that describes the goals, responsibilities, and expectations of each group member, as well as defines the agreed mode of work in terms of communication and conflict resolution. The Team Charter will be revisited on a regular basis during the check-in meetings to evaluate the performance of the groups; the contents of the Team Charter could be updated during the Semester, if necessary. By behaving professionally and complying to the specifications of the Team Charter, the student will demonstrate their commitment to the success of the group project and help ensure that it is completed to a high standard. At the end of the Semester, the grade for the Individual Contribution to Group Project will be determined by the Core Lecturer by evaluating the student's performance over the Semester against the specifications of the Team Charter; the Core Lecturer will also take into account the reflections provided by other group members and own observations regarding the performance of the student during the Semester.

Individual Assignment CME: Ethics Position Statement

This assignment involves developing a comprehensive ethics position statement that addresses the issue of inequality in Artificial Intelligence (AI). Your task is to explore and critically analyse how AI can perpetuate or mitigate social inequalities, and propose theoretical solutions based on your findings. You should take and defend a contestable ethical position.

This exercise aims to deepen your understanding of the ethical implications of AI in society and encourages you to apply theoretical perspectives to real-world issues. This assignment consists of two main components: Perusall Annotation Tasks (45% of the grade) and an Ethics Position Statement (55% of the grade). It aims to deepen your understanding of AI's ethical implications and encourage applying theoretical perspectives to real-world issues.

Individual Assignment SSH: Essay on Algorithms and Social Inequality

Students are asked to explore an issue of their choosing in which algorithmic decision-making has substantial impact on social inequality. You will assess your topic using one of the two traditional political science theories, Realism or Liberalism (Part A), and then critique that assessment from a critical theory perspective (Part B). The critique should look at how algorithmic decision-making reproduces ideational and material structures and how stakeholders may confront this. Students may draw on concepts from lectures and workshops but are expected to do substantial literature reviews independently.

Students may choose to focus on any topic that is of significant societal concern. Students are

encouraged to choose the same topic for the RE Qualitative Prediction Framework and the SSH Essay on AI and Social Inequality assignments. However, they are **not** required to choose a topic that is closely related to their group project. Every student must choose a distinct societal issue as the topic for their essay.

Individual Assignment RE: Qualitative Prediction Framework

Students will create a framework for modelling and predicting a complex societal phenomenon that is associated with a high degree of uncertainty and interdependency and which relies on qualitative reasoning. They will create a long-range forecast of social inequality as it relates to their topic. The predictive framework will be based on probabilistic reasoning and will consist of identifying the factors affecting the phenomenon of their choice; making an informed guess of their estimates and variability; and outlining the relationship between the elements of the framework. Student will use desk research to justify their decisions and ensure the reliability of their modelling.

The prediction framework will be accompanied by the cover note explaining the rationale for the prediction, justifying the components of the prediction framework, interpreting the results of the modelling. The cover note should also reflect upon the limitations of the modelling, including critically evaluating potential biases in the sources used.

Students may choose to focus on any topic that is of significant societal concern. Students are encouraged to choose the same topic for the RE Qualitative Prediction Framework and the SSH Essay on AI and Social Inequality assignments. However, they are **not** required to choose a topic that is closely related to their group project. Every student must choose a distinct societal issue as the topic for their prediction framework.

Individual Assignment DE: Coding Portfolio and Explainer Session

After each DE practical from weeks 2 to 7, students will be provided with a coding assignment. At the end of the assignment, a minimal set of solutions with a standard dataset will be provided for students to take inspiration in crafting their own solutions. Your task is twofold: 1) to complete, and keep a well-annotated record of, the code and generated results from these assignments using a different dataset; 2) to familiarise yourself with the code and the method enough such that if you are questioned on any part of the coding or modelling choices, you are able to defend your code. At the end of each week, you will be exposed to a mock oral defence, where TAs / CLs will test your ability to defend coding and modelling choices informally, and give you feedback on how to better explain yourself verbally.

In week 9 you will hand in a copy of your coding portfolio for assessment. Then you will undergo your individual oral explainer session. The defence will last around 15 minutes, and two examiners from the CSSci staff will be present. Your job will be to:
1. Answer a series of questions related to your code, and/or
2. Provide accurate justifications for the chosen methodology's compatibility with your data, and/or
3. Provide accurate justifications for the chosen methodology's compatibility with the goals of the analysis, and/or
4. Comment on the results of your DE effort.

Assignment Deadlines

Assignment deadlines are strict. **Canvas settings have been changed: starting with the Spring 2024 semester, Canvas will no longer accept submissions after the deadline**. If you miss the deadline for a graded assignment, you will not receive a grade (NAP), you are not eligible for a repair opportunity, and you cannot complete the course. Your Core Lecturer cannot change this. If you are unable to submit your assignment in a timely manner due to extenuating circumstances, contact the Study Advisor immediately.

When submitting an assignment, please allow sufficient time to deal with potential minor technical issues (problems with your computer or internet connection, time required to upload large files, etc.) as they are not considered a valid reason for missing a deadline.

Please be aware that Canvas does allow multiple submissions of your non-quiz assignments. If you are concerned about the deadline, we recommend that you first submit your deliverable as soon as you have what you consider to be a serious attempt at the assignment. You may then later – but before the deadline! - resubmit the finished version of your deliverable. The most recent submission of the assignment that is made before the deadline will be graded.

Deadline Extension Request Procedure for Graded Assignments

Deadline extensions are given only in exceptional, unforeseen circumstances. To request an extension of the deadline from the Examinations Board due to extenuating circumstances, such as illness, **please submit your request as soon as you are experiencing the circumstance** (e.g., at the start of the illness). **Do not wait until after you have missed the deadline**.

If the request is sent to the Examinations Board in the five days leading up to the assignment deadline, the board cannot guarantee that your request will be processed before the deadline. If the outcome of your request is still pending on the day of the deadline, please make sure to upload your (incomplete) assignment to ensure that you are eligible for participation in the repair opportunity should your request be denied, and in case of an Insufficient result. In case of extenuating circumstances, it is crucial to communicate clearly and timely about your situation with both the study adviser and your Core Lecturer, and if possible, document your circumstances with the study adviser.

In case of sudden, short and unforeseen emergencies on the day of the deadline, or the day before it, that require an extension of a maximum of three days, students should directly contact the study adviser for an extension request without submitting a request to the Examinations Board. Students can be asked for documentation to substantiate their request. **Again, please make sure to upload your (incomplete) assignment to ensure that you are eligible for participation in the repair opportunity should your request be denied, and in case of an Insufficient result.**

Academic Integrity

Rules Regarding Fraud and Plagiarism

The provisions of the Regulations Governing Fraud and Plagiarism for UvA Students apply in full. Access this regulation at https://student.uva.nl/en/content/az/plagiarism-and-fraud/plagiarism-and-fraud.html.

Plagiarism can take many forms, including (but not limited to!):
- making use of or reproducing another person's texts, data or ideas without complete and correct acknowledgement of the sources;

- presenting the structure or central body of ideas taken from third-party sources as one's own work or ideas, even if a reference to other authors is included;
- submitting a text that has previously been submitted, or is similar to a text that has previously been submitted, in the context of assignments for other courses;
- reproducing the work of fellow students and passing it off as one's own;

If you have not read these regulations before, please do so! Note that plagiarism and fraud is extremely serious. In case of alleged plagiarism and/or fraud your Core Lecturer is obliged to inform the Examinations Board immediately.

The Examinations Board may decide on a range of actions, including dismissing the assessment attempt. In that case the grade will be registered as 'Unacceptable'.

It is important to realize that plagiarism applies to all of your academic output, **including programming code**.

In general: DON'T EVEN THINK ABOUT IT.

Rules Regarding the Use of Artificial Intelligence Software

As we continue to advance in the digital era, we wholeheartedly embrace the positive contributions that artificial intelligence offers to your learning journey. An example of this are Large Language Models (LLMs), like ChatGPT. These can help adequately phrase ideas that you feed them, summarise texts they are explicitly fed, or help you when you encounter clearly defined issues in coding. However, we must emphasize that LLMs are no substitutes for your intellectual efforts and critical thinking. Prudentially, they do not rely on any external criterion for truth - this means they may present made-up information as real (even inventing sources), remain vague in their answers and neglect key aspects, or produce buggy code. Ethically, they raise issues of claiming authorship for ideas that are not your own and may undermine responsibility for your own output. Hence, the development of original work remains an essential part of our educational mandate. Remember that LLMs are here to assist, not to replace your creative process.
Please also get acquainted with the UvA general information on the use of ChatGPT and AI tools in your studies: https://student.uva.nl/en/topics/chatgpt-and-your-studies

If a student has made use of ChatGPT or a similar tool for an assignment, this should always be mentioned by the student and the tool should be cited as a source in the bibliography.

Open Research, Data Integrity and Storage

CSSci supports open research. Unless otherwise specified, the results of research conducted with external Challenge Partners will be shared with those partners and may be further distributed by them. CSSci may choose to share student results, reports, code, data visualizations, or other products for research or educational purposes. If a student objects, they must contact their Core Lecturer as soon as possible to explain why the research should not be made available.

It is therefore of the highest importance that research by students should be conducted responsibly, neatly, and archived for a period of two years. Both during ongoing research and for archiving purposes, all data should be stored with UvA storage facilities (OneDrive or another option approved by your Core Lecturer) and appropriately secured. You **may not** use 3rd party options (Google Drive or Dropbox, for example) to store any data which may contain personal information as per the GDPR.

For the rules, see the "Ethical Guidelines for Students" on the GSSS page (especially: Section V. Data storage): https://gsss.uva.nl/about/policy-documents/policy-documents.html

Personal Safety and Security

We want students to be assured they have a positive environment and a safe basis for their studies. We hope you never experience an unsafe situation or undesirable behaviour at the UvA, but should this occur there are different individuals and institutions you can turn to (such as the study adviser and the programme director).
Please see https://student.uva.nl/en/content/az/social-safety/social-safety.html

Literature and Materials

Readings and additional materials are announced through Canvas.

Date of Final Grade

No later than 15 working days after the last graded assessment.

Programme / Weekly Planning and Deadlines

Weekly schedules can be found in Canvas and rooster.uva.nl.

Learning Objectives

The activities of this semester together work towards the following learning trajectory objectives:

Social Sciences and Humanities Expertise
Students are able to:

- Use relevant theories to recognize and explain structural inequality, bias and stratification.
- Use relevant theories to identify potential sources of bias and power relations and values to contribute to the development of socially responsible AI solutions.
- Discuss how design choices in AI reproduce ideational and material structures and how stakeholders may confront this.
- Analyze the ethical, legal and societal impact of AI/algorithms.

Digital Expertise
Students are able to:

- Demonstrate a thorough understanding of machine learning and other digital techniques for prediction.
- Understand how social biases manifest in digital data and predictive algorithms.
- Detect social biases in digital data and predictive algorithms.
- Implement appropriate social bias mitigation techniques.
- Apply algorithm audits to identify potential data-driven fairness issues.
- Discuss the ethical and liability implications of the choices made when implementing digital analyses.

Research Expertise
Students are able to:

- Make predictions incorporating digital data using predictive statistics.
- Make predictions incorporating qualitative data using probabilistic reasoning.
- Assess and interpret the reliability of predictions using the adequate techniques for the relevant method.
- Explain the social issues related to predictive (recommendation) systems.
- Critically reflect on the ethical, legal and social aspects of research designs involving the collection of digital behaviour and/or communication data as well as of their underlying

assumptions and biases.

Change Making Expertise
Students are able to:
- Design and implement projects that are able to take power and values into account.
- Explain socially responsible AI as a change making method.
- Evaluate vis-a-vis ethical aspects of AI/algorithm-based interventions.
- Reflect on collaboration with all kinds of stakeholders.
- Suggest scenarios on how new digital tools may influence individuals and society with probable desirable and undesirable effects.
- Develop long range future scenarios.

Updates to Course Manual

Version	Date	Changes
1.0	19 January 2024	none
1.1	2 February 2024	Course Content - Updated to include information on periods reserved for collaborative group work. Professional Study Attitude > General lack of participation - Updated to include information for unresponsive students. Assessment - Updated to include reference to "Unacceptable." - Repair or Resit? table updated to include "Resit: Reparations, Reflection, and Plan for Personal Improvement" AND to update the "Date / Deadline" column. Rules Regarding Resits - Added "and no resit or repair opportunity will be possible." Repair or Resit? Table - Updated with correct resit/repair deadlines Assignment Deadlines - Added information about the consequences of missing a deadline for submitting a graded assignment. - Updated the information about multiple submissions to be clearer. Open Research, Data Integrity and Storage - Includes "code" in the list of deliverables that may be shared.
1.2	9 February 2024	Graded Assessment Overviews > SSH Essay Updated with corresponding assignment title. Essay on AI and Social Inequality > Essay on Algorithms and Social Inequality. Professional Study Attitude > Consequences - Updated to include lateness policy

= END OF DOCUMENT =