SOSC 4300 / SOSC 5500: Computational Social Science

Spring 2026

Lecture Time: Wednesday 18:30 - 21:20   **Tutorial** Time: Friday 16:30 – 17:20

Lecture Room: LG3008   **Tutorial** Room: 6602

Course homepage: Canvas

Instructor

WEI, Jinlin, Room 2363, Academic Building, jinlinwei@ust.hk, Office Hour: Wednesday 15:00 – 16:00

Teaching Assistant

LI, Jiajun, Room 3001, Academic Building, jliko@connect.ust.hk, Office Hour: TBA

Course Description

This course focuses on two main topics, machine learning and text analysis. This course aims to introduce these fields to senior undergraduate students and postgraduate students. Students will be expected to learn to apply basic machine learning and text analysis techniques to social science phenomena and focus on specific topics that they choose to address social science questions of interest to them. The course employs a range of assessment methods, including coding assignments and group projects, to foster both collaborative learning and personal growth.

Prerequisites

• Students are expected to be familiar with the material covered in basic statistics (e.g., SOSC 2400 for UG students and SOSC 5090 for PG students). Students who have equivalent statistics knowledge but do not meet the formal prerequisites may seek the instructor's approval to enroll.

• Students should also have basic proficiency in at least one statistical programming language. We will use R and Python in tutorials. You can also use other programming languages such as Matlab, Julia, etc., provided they can complete course assignments and projects using those languages.

Reading materials

All materials are suggestive. The two reference books listed make it easier to follow the lectures. Optional readings are for students who are interested in reading more on the topic. Search engines and Generative AIs are very useful if you would like to master specific topics.

Reference Books

1. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R* (R. Tibshirani, T. Hastie, & D. Witten, Eds.; 1st ed., Vol. 103). Springer Nature. https://doi.org/10.1007/978-1-4614-7138-7
2. Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

Intended Learning Outcomes (ILOs)

By the end of the course, students will be able to:

1. Describe the opportunities and challenges of social research in the age of big data.

2. Evaluate research on social phenomena from different fields, including the social sciences and computer science/data science.

3. Apply essential techniques for analyzing social big data.

4. Propose research questions suitable for examination using computational methods and big data.

5. Design a poster applying computational social science to a real-world problem, or write a research article that uses computational social science methods to address a social science question.

Schedule

The class schedule may be adjusted over time.

Week	Date	Topic	Events
1	Feb. 4th	Introduction	
2	Feb. 11th	Big Data; Prediction: LASSO	Fix Grouping by Feb. 20th
3	Feb. 25th	Prediction: LASSO	

4	March 4th	Prediction: Decision Tree	Assignment 1 announced
5	March 11th	Prediction: Supporting Vector Machine	
6	March 18th	Text Analysis: Dictionary Method	Assignment 1 due
7	March 25th	Text Analysis: The Vector Space Model	Optional: Send the proposal for the final poster/paper for feedback
8	April 1st	Mid-term Exam	Assignment 2 announced
9	April 15th	Text Analysis: Word Embedding	
10	April 22th	Text Analysis: Topic Models	Assignment 2 due
11	April 29th	Presentation	
12	May 6th	Advanced topics using LLM	
13	May 24th	Final Submission	

Assessment and Grading

- SOSC4300 and SOSC5500 are assessed separately. Students' final scores will be based on the following components:

i) Participation: 10%.

ii) Assignments: two assignments, each worth 10%. Due dates: March 18th, April 22nd.

iii) Mid-term exam: 25%. Date: April 1st.

iv) Presentation: 10%. Date: April 29th.

v) Peer poster evaluation: 10%. Due date: May 24th.

vi) Final Poster + Demo or Paper Write-up: 25%. Due date: May 24th.

1. Solo work

i. Participation in class activities

- Participation is based on involvement in lectures.
- Students are expected to keep up with course content by completing in-class exercises and participating in discussions.

ii. Mid-term exam

- Multiple-choice Questions
- The exam tests understanding of the course material, especially the methods introduced.

iii. Peer poster evaluation

- Students will vote for their three preferred posters.
- Each student will also be randomly assigned a poster and will write a constructive review of it (maximum 500 words).

2. Group work

- Max group size is 2.
- Finalize group membership by February 22nd. Students who have not formed a group or chosen to work alone by that date will be randomly assigned to groups.
- Both group members will receive the same score for group assessments unless otherwise noted.

i. Assignments

- There will be 2 assignments to assess your ability to apply and evaluate basic computational social science algorithms.
- Discussion with other students is allowed, but assignments must be completed only with your group member.
- Proper usage of generative AI is allowed and encouraged. If you use AI to generate code, you must include the prompts and identify the specific AI model(s) used.

ii. Presentation

Every group needs to submit a presentation by April 29th. Some groups will present in person (a Zoom recording will be made). Others will submit a pre-recorded video presentation.

• If your group is presenting a poster, you have 10 mins for presentation and 5 mins for Q&A. The recording will be distributed to peers for evaluation: 5% of the course grade will be based on peer evaluation and 5% on the teaching team's evaluation.

• If you are presenting a paper, you have 15 mins for presentation and 5 mins for Q&A. Follow a standard academic-talk format. This presentation is worth 10% of the course grade, based on the teaching team's evaluation.

Assessment of the presentation and the final poster/paper

• Your presentations will be assessed based on the following criteria. See the Assessment Criteria provided on Canvas for more details.

• Note that the audience of presentations and the readers of posters/papers engage differently and on different time scales. Tailor your presentation to your materials to attract and retain the audience's attention.

- Attracting the audience: State the question to answer and explain why it is interesting or important.
- Literature Review (applies only to research papers): Summarize relevant work and explain your contribution.
- What is done: Describe how you collected (if applied) and analyzed the data.
- The answers: Explain the results and their implications.
- Presentation style: Clear, engaging, and easy to follow.

iii. Final poster/paper

SOSC4300 students may choose to create a poster that analyzes a "real-world" social science problem or write a research final paper. The difference between the two options lies in the intended audience: the poster is aimed at a lay audience, while the final research paper is intended for researchers. We recommend discussing your ideas with the instructor during the early weeks of the course, either in office hours or by email.

Optional: To solicit feedback on your project idea, you may submit a proposal via Canvas by March 25.

Poster

Focus on a real-world case and aim to communicate social-science ideas to a general audience using compelling data analysis and visualization. Examples:

- https://mdi.georgetown.edu/posters/billy-mcgloin-25-mdi-summer-2024-research-showcase-poster/
- https://www.youtube.com/watch?v=waMvm-S1PGE (includes both a poster and a presentation)
- https://michaelminn.net/tutorials/posters/2025-cracker-barrel-poster.pdf

Provide a short written report covering details that cannot be presented on the poster, including—but not limited to—the importance of your research question and the rationale for your chosen data and methods. Limit: 5 pages, 12-point font, double-spaced, including tables and figures (references excluded).UG students can choose to write a poster.

Research paper

Choose a research topic and write a research paper using computational social science methods or digital data. The paper should follow the standard research-article format: introduction, literature review, methods and data, results, and conclusions. Use examples from class readings and the reading list as models.

Limit the final paper to no more than 20 pages, 12-point font, double-spaced, including tables and figures (references excluded).

SOSC4300 students may choose either option. SOSC5500 students must submit a research paper.

Grading policies for the final poster/paper

- You may submit multiple drafts before the due date. We encourage submitting an initial version and revising it.
- Final submissions will be checked with plagiarism-detection software. Students must avoid plagiarism; confirmed cases may result in severe penalties, including a failing grade.

Late Submission Policy

- All assessments will have a clear deadline by which they are due. Late submissions, unless you have valid reasons supported by a doctor's note, will get partial credit, depending on lateness:
- 1) Submission within 24 hours of the deadline: a minimum of 75% of the assigned score (-2.5% every hour)
- 2) Submission between 24 and 72 hours after the deadline: a minimum of 50% of the assigned score (-2.5% every hour)
- 3) Submission more than 72 hours after the deadline: no credit

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
Participation	ILO1, ILO2, ILO4	This task assesses students' understandings of (ILO1) and ability to evaluate (ILO2) the application of big data in social science research and in solving real-world problems. It also assesses students' ability to propose potential solutions to social science problems. (ILO4).

Assignments	ILO1, ILO3	This task assesses students' understandings (ILO1) and their ability to apply computational and statistical tools (ILO3) to problems involving big data in social science research.
Mid-term exam	ILO1, ILO2, ILO3	This task assesses students' understanding of (ILO1) and their ability to evaluate (ILO2) the application of big data in social science research and in solving real-world problems. It also assesses students' ability to come up with appropriate solutions to social science problems in given scenarios (ILO3).
Evaluation of Peers' Posters	ILO2, ILO5	This task assesses students' ability to evaluate (ILO2) the application of big data in social science research and in solving real-world problems. It also assesses students' ability to reflect on their research projects when reviewing peer students' (ILO5).
Group Paper/ Project	ILO1, ILO2, ILO3, ILO4, ILO5	This task assesses students' understanding (ILO1) of applying big data in social science research in specific topics in which students have interests. It assesses their ability to evaluate the importance of specific research in corresponding topics (ILO2) and propose a meaningful research question or a practical social problem to solve (ILO4). It assesses students' ability to collect data, conduct analyses (ILO3), and their ability to communicate findings in a paper or poster/project (ILO5).

Final Grade Descriptors

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Demonstrates a deep understanding of fundamental concepts and methods in machine learning and text analysis. Exhibits exceptional critical thinking skills in evaluating and designing social science research using big data, consistently applies appropriate computational social science methods to enhance problem-solving, and produces insightful analyses. Effectively communicates complex research ideas clearly and persuasively.

B	Good Performance	Shows a solid grasp of the fundamental concepts and the application of methods in machine learning and text analysis. Demonstrates sound critical thinking skills in evaluating and designing social science research using big data, generally applies appropriate computational social science methods to support problem-solving, and produces accurate analyses. Clearly communicates research ideas.
C	Satisfactory Performance	Shows an adequate understanding of the fundamental concepts and the application of methods in machine learning and text analysis. Displays basic critical thinking skills in evaluating and designing social science research using big data, applies computational methods sufficiently to address problems but with limited depth or sophistication, and produces correct but sometimes shallow analyses. Communicates the basic frameworks of research ideas.
D	Marginal Pass	Possesses only basic knowledge of the fundamental concepts and the application of methods in machine learning and text analysis. Shows limited critical thinking skills in evaluating and designing social science research using big data, applies computational methods with minimal effectiveness, and produces incomplete or weak analyses. Communication of ideas is often unclear or superficial.
F	Fail	Demonstrate insufficient understanding of the fundamental concepts and the application of methods in machine learning and text analysis. Lacks critical thinking skills in evaluating and designing social science research using big data and the ability to apply computational methods effectively. Fails to communicate or justify analyses adequately.

Communication and Feedback

Assessment marks for individual assessed tasks will be posted on Canvas within 10 working days of submission. Feedback on assignments will include comments on strengths and areas for improvement.

Students who have further questions about the feedback including marks should consult the instructor within 5 working days after the feedback is received.

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest

standards of academic integrity. The University has zero tolerance for academic misconduct. Please refer to Academic Integrity | HKUST -Academic Registry for the University's definition of plagiarism and ways to avoid cheating and plagiarism.