

Leveraging Large Language Models for Public Opinion Analysis

Facilitator

Dr. Christina Haag, Digital & Mobile Health Group, Institute for Implementation Science in Health Care, University of Zurich and Epidemiology, Biostatistics and Prevention Institute, University of Zurich

Description

Large-scale text data, such as social media posts and press articles, have become invaluable for analyzing public discourse and opinion. The influence of online text data is so significant that 'infodemiology' research now focuses on understanding how information spreads through digital media (particularly the Internet) and among populations, as well as identifying the factors that shape this dissemination. A key area of interest in public health research is assessing how the general public responds to specific health policy decisions or recommendations—whether they are met with support, or instead provoke fear or anger. For example, the preventive measures implemented during the pandemic and vaccination campaigns sparked extensive discussions on social media and across various media platforms, which not only mirrored public opinion but also played a role in shaping it.

But how do language models work? And how can they be used responsibly in research? This course will (1) cover the basics of how language models work from an applied research perspective, (2) explain how they can be used to answer common research questions aimed at mining public opinion, and (3) finally discuss challenges of using language models (e.g., underlying biases, stereotypes) and principles of 'responsible AI'.

Objectives

By the end of the course, participants will have a basic understanding of language models, how to apply them in the context of public health research, and what to consider in the responsible use of language models.

Specifically, participants will:

1. Discuss the potential and risks of using language models in public health research.
2. Evaluate and test how different language models perform when mining public opinion using publicly available AI resources.

3. Develop competencies for the responsible use and critical evaluation of language AI

Date

18 June 2025, 14:00 – 16:00

Level

Basic

Eligibility

The course is open to PhD and MD students of the SSPH+ Inter-university Graduate Campus. It is aimed at PhD students who want to learn about language models and explore how they can be used for opinion mining in their own research. This is a basic level course, and no programming skills are required for course attendance. If you choose to complete the assignment, basic Python skills will be an advantage. You also can complete the assignment without any prior knowledge of Python as you will be provided with well commented sample code. However, the assignment is likely to take longer to complete if you have never used Python before.

Course Structure

The course will be delivered online and will consist of short lectures, group discussions and hands-on exploration of language models suitable for opinion mining. The short lectures will provide an introduction to language models in general and opinion mining in the context of public health research in particular. Participants will work in teams (breakout groups) to discuss questions about the potential, challenges, and responsible use of language models. As part of the hands-on approach of the course, participants will also explore publicly available pre-trained language models using the rich resources of the established AI platform Hugging Face (<https://huggingface.co/>). The website interface allows for convenient online testing of the language models during the course (e.g. <https://huggingface.co/j-hartmann/sentiment-roberta-large-english-3-classes>). We will also have a look at the Python topic modelling library BERTopic (<https://maartengr.github.io/BERTopic/index.html>). The results and conclusions of the groups will be presented and discussed in plenary at the end of the 2-hour online course.

Work load

Preliminary Work: 2 hours; Contact time: 2 h

Credits

(1 ECTS for active participation in 4 2h-online courses and 1 passed assessment in 1 course)

Assessment	(voluntary) If participants choose to do an assessment, they will conduct a topic modelling analysis on a public dataset of their choice (public datasets containing, e.g., tweets on vaccination or Long Covid). They will also produce a short report interpreting the results, including a critical evaluation of potential limitations of their analysis.
Location	Online course on zoom
Course Fees	SSPH+ PhD and MD Students: o.- CHF
Registration	https://www.conftool.com/ssph-phd-courses2025/
Deadline for registration	18 May 2025

This course is an SSPH+ IGC initiative to provide post-doctoral scientists at SSPH+ network institutions to teach.