

Title of Project:

(English)	Large Language Models and Data Science Applications in the Humanities and Social Sciences
-----------	---

(Chinese)	大型語言模型及數據科學在人文及社會科學的應用
-----------	------------------------

This project is a collaborative conference on “Large Language Models and Data Science Applications in the Humanities and Social Sciences” jointly organized by the Ip Ying To Lee Yu Yee School of Humanities and Languages, Felizberta Lo Padilla Tong School of Social Sciences, and Data Science Research Center (DSRC), of Saint Francis University (SFU), and the Faculty of Arts and Social Sciences of Hong Kong Chu Hai College (HKCHC).

Our motivation of holding this conference is due to the fact that the rapid advances in Artificial Intelligence (AI) research, particularly the development of Large Language Models (LLMs), has revolutionized the way scholars approach their work in humanities and social sciences. LLMs are trained with vast amounts of textual data for multiple language-related tasks such as translation, text classification, and language modelling. This gives LLMs a wide range of skills and knowledge relevant to researchers.

By leveraging cutting-edge LLM-based tools such as ChatGPT, researchers can now tap into vast amounts of data and materials that were previously inaccessible or too time-consuming to analyse manually with conventional tools. By automating tasks in text processing and data analysis, researchers have more time to focus on higher-level tasks, such as interpreting results, making connections between findings, and developing new ideas.

LLMs have become increasingly effective in summarising, analysing and retrieving large corpora of multilingual text data, allowing researchers to uncover nuanced patterns and relationships. This development has far-reaching implications for fields like linguistics, translation, literature, and history, where text-based materials are a fundamental part of research. For social scientists, LLMs can help in modelling and predicting social phenomena such as detecting hate speech in social media. It also enables them to develop new tools and methods for knowledge discovery such as emotion detection in interview recordings which can significantly enhance the research process.

Amidst the widespread adoption of LLMs in humanities and social sciences, researchers need to be mindful of the biases and errors inherent in the data used for the training of these models. For instance, LLMs are known to suffer from the so-called hallucination problem. All these require caution and safeguards in applying LLMs safely and ethically.

Researchers can reap more benefits from LLMs through collaboration and sharing of their experiences. As a rapidly developing technology relevant to diverse fields, researchers stand to learn from each other and from experts in the technology about its strengths and weaknesses as well as any risks and challenges.

To broaden our horizon on the applications of machine learning and data to the humanities and social sciences, we will invite eight prominent speakers from Hong Kong, Taiwan, Germany, United Kingdom, the United States of America to be keynote speakers of the conference. This conference will have four sessions on the conference theme: Deep Learning in the New Age

coordinated by Ip Ying To Lee Yu Yee School of Humanities and Languages, New Advances in Data Science coordinated by the Data Science Research Center, Deep Learning in the Humanities coordinated by the Faculty of Arts and Social Sciences of HKCHC, and Data Applications in Social Sciences coordinated by Felizberta Lo Padilla Tong School of Social Sciences of SFU. Despite the distinct focus of each session, there will be a strong emphasis on collaboration, synergy, and cross-disciplinary exchange among all participants. This collaborative environment will ensure that insights and innovations are shared widely, enhancing the impact of the conference. Additionally, we will invite at least 52 academic researchers, scholars, and research students to submit papers and present their findings at the conference. The paper presentations are a crucial component of the conference, not only highlighting the importance of academic contributions but also enhancing the overall value of the conference by integrating theoretical knowledge with practical applications.