Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
Consiglio Nazionale delle Ricerche

Image credits: Paul Blenkhorn on Unsplash; visual design: Rita Giuffredi

FOSSR learning session

December 10-11, 2024
CNR-ISTC (Rome) and online (hybrid event)

DATA ANALYSIS IN THE ERA OF LARGE LANGUAGE MODELS WITHIN FOSSR

FOSSR
Fostering Open Science in Social Science Research
Innovative tools and services to investigate economic and societal change

Learning session agenda

Dec. 10

14:00 **Welcome and Introduction**
Objectives and Structure of the Learning Session

14:15 **Key Concepts for Data Analysis in the era of Large Language Models**
Frontal introduction

14:30 **Development of ideas: Definitions, Challenges, and Issues Related to Key Concepts**
Group discussion

15:15 **Key Concepts in Action**
Group presentations

16:00 *Coffee break*

16:10 **Theoretical perspectives and issues**
Frontal lesson, interwoven with insights from the Groups

17:40 *Q&A and Wrap-Up*

Dec. 11

10:00 **Presentation of the exercise on data analysis with the support of an LLM**
Frontal lesson

10:30 **Group Exercise and discussion**

12:30 **Exercise execution with expert insights**
Group Exercise

13:30 **Wrap-Up**

13:40 *Closure of the learning session*

The learning session's contents are curated by ***Andrea Giovanni Nuzzolese*** (CNR-ISTC)

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
Consiglio Nazionale delle Ricerche

FOSSR learning sessions

FOSSR participatory learning sessions rely on promoting **active learning** with the guide of an instructor on how to exploit the **research potential of the FOSSR-related infrastructures**, particularly for tackling for societal issues. Their objective is to allow the emergence of **training needs** related to concrete social research possibilities based on data exploitation, give them a voice and stimulate the learning and the collaborative search for solutions by peers and experts. Contrarily to teacher-cantered methodologies, participatory learning relies on active involvement: **participants contribute to propose the learning contents, create questions, stimulate discussions, carry out critical reflections and practical activities, engage in role playing and mutually evaluate their peers' outcomes.**

This kind of approach follows the constructivist theories of learning, for which knowledge is actively constructed rather than transmitted and people learn by actively building their understanding by applying their knowledge to *meaningful* problems.

Session description

The **Learning Session "Data analysis in the era of large language models"** is based on a two-day program, scheduled for December 10-11, 2024.
The objective of the learning session is to provide participants with a **comprehensive understanding of how LLMs can be applied to data analysis**. Through a mix of theoretical insights and group exercises, participants will gain **both conceptual knowledge and hands-on experience, fostering a participatory and engaging learning environment**.

On **Day 1**, between 2:00 PM and 6:00 PM, participants will engage in a series of activities, which take an approach oriented to participatory learning. The session will start with a welcome and an introduction explaining the goals and format of the two-day learning event. The lecturer will introduce important keywords relevant to data analysis and large language models (LLMs). These keywords will lay the groundwork for understanding core LLM-based data analysis concepts with benefits and potential issues. The participants will break into groups to brainstorm ideas, define concepts, and discuss challenges related to the assigned keywords on LLM-based data analysis. This exercise will foster collaboration and a deeper understanding of the terminology. Hence, each group will present their findings, sharing insights on the meaning, application, and implications of the assigned keywords in the context of data analysis with LLMs. Then, the lecturer will provide theoretical insights into key principles of data analysis, such as prompt engineering, model interpretability, and ethical considerations, drawing connections to the ideas shared in group discussions. The day will conclude with a Q&A session.

On **Day 2**, starting at 10:00 AM and lasting until 1:40 PM, the lecturer will introduce an exercise on data analysis that asks participants to interact with an LLM to find a solution. Accordingly, the participants will work in groups to find their own solution to the assigned exercise, using an LLM-based tools to analyse a dataset relevant to FOSSR. Finally, we will conclude with a discussion about how LLMs are transforming data analysis, with a focus on FOSSR settings, and the potential they bring to various research areas. Emphasis will be placed on the importance of responsible and informed use of LLMs by social scientists.

How to apply

Applications should be submitted through the form at: ***https://l.cnr.it/fossr-learning-session-registration-form*** *by* ***December 1, 2024***. Contact for info: **fossr.dissemination@ircres.cnr.it**

Notification of acceptance by Dic. 2, 2024. It is hereby noted that, due to reasons related to the course format, participation in the learning session will be limited to a **maximum of 25 attendees**. If the number of registrations exceeds 25, a selection will be made based on the relevance of the educational or professional background, in relation to the course topics, as outlined in the submitted CV, and the motivations expressed in the corresponding registration form.

Organisational details

The learning session will be held in hybrid mode: in-presence and online, taking place in-presence at **CNR-ISTC, Via Giandomenico Romagnosi, 18a (Rome)**; the link for online participation will be provided to participants in due time before the beginning of the course, and a rich and complete interaction between on-line and in-presence participation will be curated.
Sessions will be held in Italian. Materials, thematic literature and notes on the topics covered will be available to all participants in English.

The training is organised by CNR-IRCrES and CNR-ISTC in the frame of the FOSSR project.
Local scientific/organising committee: Andrea Giovanni Nuzzolese, Serena Fabrizio, Rita Giuffredi, Andrea Orazio Spinello, Alessandra Maria Stilo.

This event was supported by FOSSR (Fostering Open Science in Social Science Research), funded by the European Union - NextGenerationEU under NRRP Grant agreement n. MUR IR0000008.