

LLM SURVEY IN HIGHER EDUCATION

Data Management Plan

Project aim:

With this survey on LLM we are exploring the impact and potential of trustworthy artificial intelligence (AI) in higher education. Our latest project focuses on understanding the current state and usage of Large Language Models (LLMs) in academic settings, and we are seeking your valuable insights through our comprehensive survey.

We seek answers to the following research questions:

Usage and Adoption:

- *To what extent are Large Language Models (LLMs) currently used in higher education institutions, and for what primary purposes (e.g., teaching, research, administrative support)?*
- *What factors influence the adoption of LLMs in higher education settings, and are there differences across departments or disciplines?*

Perceived Impact on Teaching and Learning:

- *How do educators perceive the impact of LLMs on the quality and effectiveness of teaching and learning?*
- *In what ways do LLMs change or enhance student engagement and learning outcomes in academic environments?*
- *What challenges do educators face in integrating LLMs into their teaching practices?*

Trustworthiness and Ethical Considerations:

- *How do educators and administrators perceive the trustworthiness and ethical implications of using LLMs in higher education?*
- *What concerns do higher education professionals have regarding data privacy, academic integrity, or bias when using LLMs?*
- *What policies or guidelines are in place (or needed) to ensure responsible and ethical use of LLMs in academic settings?*
- **Potential for Future Use and Development:**
- *What potential do educators and administrators see for the future integration of LLMs in higher education?*

PI:

Christa Tigerstedt

Research team:

Alessio Gallucci
Henrika Franck
Elisabeth Hildt
Pedro Kringen
Frode Ramstad Johansen
Christa Tigerstedt
Magnus Westerlund,
Emilie Wiinblad Mathez

1. GENERAL DESCRIPTION OF THE DATA

1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Additionally, give a rough estimate of the size of the data produced/collected.

Both quantitative and qualitative survey data (numbers and text: Likert scale and open-ended questions) is collected. MS Excel, CSV and SPSS is used for quantitative data processing and MS Word (N'Vivo or similar) for qualitative data processing. We have ensured reliable data collection through quality control checks during survey implementation. We use the following tool for the survey <https://e-lomake.fi/en/>.
(A survey draft is attached.)

1.2 How will the consistency and quality of data be controlled?

Standard preparatory measures for survey data are conducted, e.g. control of missing data and inconsistent responses. The research team will review data periodically during collection to ensure quality and consistency, and to identify any potential issues early.

2. ETHICAL AND LEGAL COMPLIANCE

2.1 What legal issues are related to your data management? (For example, GDPR and other legislation affecting data processing.)

Sensitive personal data are not processed nor collected in this research project. All data is collected anonymously and thus GDPR does not apply - obligations to keep records, which is expedient. We as researchers process data only on group level (such as age, gender, occupation). The studies are conducted in accordance with the guidelines drawn up by the Finnish National Board on Research Integrity (www.tenk.fi) for non-medical research involving human participants. All participation in the research is voluntary.

2.2 How will you manage the rights of the data you use, produce and share?

Data collected by the project is owned by the Graduate School and Research at Arcada, which received the project funding. Upon reasonable request and approval from the project team, an anonymous dataset may be shared with external researchers. After project completion, the team will consider making an anonymous dataset publicly available via an open-access repository under a Creative Commons Attribution-NonCommercial 4.0 International license.

GDPR regulations

Sensitive personal data are not processed in this research project. All data is collected anonymously and thus GDPR does not apply - obligations to keep records, which is expedient. We as researchers process data only on group level (such as age, gender, occupation). The studies are conducted in accordance with the guidelines drawn up by the Finnish National Board on Research Integrity for non-medical research involving human participants. All participation in the research is voluntary.

2.2 How will you manage the rights of the data you use, produce and share?

Data will not be shared outside the research team but can, in relevant cases be obtained upon request. Data files are shared with secure email between project members. Data files with respondent data are not shared in MS Teams groups.

After the project finishes an open access data repository can be considered to distribute an anonymous data set openly.

3. DOCUMENTATION AND METADATA

3.1 How will you document your data in order to make it findable, accessible, interoperable and re-usable for you and others? What kind of metadata standards, README files or other documentation will you use to help others to understand and use your data?

In the data analysis phase of the survey data software that create standardised metadata is used, such as N-Vivo, Word, Excel, which enhances interoperability and re-usage.

Published articles will be admitted a Digital Object Identifier (DOI). Relevant metadata (e.g., N participants, methods employed, RQs, any technical failures or other constraining elements observed, general review of the event) is documented in xls-files and metadata aiding in re-using of data (abbreviations in transcripts for instance) in readme/txt files. They are attributed to project data for server storage to facilitate search and re-usage of data. The research team is responsible for this task.

Documentation of the research process and pseudonymous data is continuously stored in Office 365 Sharepoint/OneDrive environment by team. After the project is ended, the data is stored in local servers of Arcada for 10 years.

4. STORAGE AND BACKUP DURING THE RESEARCH PROJECT

4.1 Where will your data be stored, and how will it be backed up?

The digital end-user data will be stored on Arcada's servers and is accessible only by the research team. The data is regularly backed up and resides within the domains of Arcada, in line with GDPR requirements. The data will be stored at least 10 years. Any analogue data and/or deciphering key will be stored in locked cupboards, separately from any research data. During the project's active phases, pseudonymous data (e.g., transcripts of interviews etc) will be stored on the OneDrive platform, allowing for access of all team members. No data that allows for identification of study participants is stored in OneDrive, but directly on

the Arcada server. When the project ends, all data will be removed for final storage on the servers of Arcada.

4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?

The PI and designated data manager will oversee access control. Any party granted data access must sign a controller-to-controller agreement, ensuring adherence to applicable laws.

5. OPENING, PUBLISHING AND ARCHIVING THE DATA AFTER THE RESEARCH PROJECT

5.1 What part of the data can be made openly available or published? Where and when will the data, or its metadata, be made available?

The research team will consider depositing an anonymized dataset in an open-access repository such as the Finnish Social Science Data Archive, subject to ethical approval. Metadata will be made available alongside the data to facilitate reuse.

5.2 Where will data with long-term value be archived, and for how long?

The digital aggregated research data will be stored on servers of Arcada and is accessible only by the research team and possible research assistants (to date, 8 persons). The data is regularly backed up and resides within the domains of Arcada, in line with GDPR requirements. The data will be stored 10 years.

6. DATA MANAGEMENT RESPONSIBILITIES AND RESOURCES

6.1 Who (for example role, position, and institution) will be responsible for data management (i.e., the data steward)?

Principal Investigator (Christa Tigerstedt) from the Graduate School and Research at Arcada will oversee data management, including access control, documentation, and long-term storage.

6.2 What resources will be required for your data management procedures to ensure that the data can be opened and preserved according to FAIR principles (Findable, Accessible, Interoperable, Re-usable)?

Required resources include access to SPSS and N-Vivo for data analysis, Arcada's secure server infrastructure for data storage, and OneDrive for secure collaborative work. Additionally, team members will need training in metadata standards and FAIR principles to ensure data is fully compliant and reusable.