Plan Overview

A Data Management Plan created using DMPonline

Title: Generative AI Facilitators for Interdisciplinary Collaborative Learning in Higher Education

Creator: Qian Shu Wang

Principal Investigator: Echo C. Wan, Dr. Malak Sadek, Dr. Minja Axelsson

Affiliation: Other

Template: DCC Template

Project abstract:

Collaborative group assignments in higher education bring together students from different disciplines to collectively solve problems and create shared artefacts. Yet these assignments often suffer from uneven participation (C1). Recent advances in large language models make it possible to use conversational AI as a lightweight facilitation aid that adaptively prompts quieter students to contribute and engage (C1).
Although prior work suggests computer supported facilitation can help, it remains unclear how generative AI can address key challenges in collaborative learning in higher education, particularly participation equity (C1). Furthermore, we do not yet understand how students perceive and experience a generative AI facilitator in interdisciplinary groupwork, including its effects on perceived collaboration quality, overall satisfaction, and perceived tool usefulness(C2).
This project aims to investigate these questions.
Accordingly, the project will design and prototype an online chat-based facilitator that prompts quieter members to contribute (C1). A controlled user study with around 20 university students across 4 text-based sessions will compare AI supported groups with unguided
groups. Using chat log content analysis and post session surveys, we will evaluate impacts on participation equity(C1), students' experiences of collaboration quality, overall satisfaction and
the perceived usefulness of the facilitator (C2).

ID: 195712

Start date: 15-12-2025

End date: 17-06-2026

Last modified: 23-01-2026

Copyright information:

Generative AI Facilitators for Interdisciplinary Collaborative Learning in Higher Education

Data Collection

What data will you collect or create?

We will collect and create three main types of research data.

1. Text based collaboration data
   Chat transcripts from each group session, exported as plain text and or CSV, plus derived participation measures such as message counts and word counts. Coverage is all messages produced during the study sessions.
2. Survey data
   Responses from short pre study and post session questionnaires, including Likert scale items and optional open ended comments, exported as CSV.
3. Research outputs created during sessions
   Group task artefacts, such as short written outputs, exported as text and or PDF.

Administration data
Email addresses for scheduling and incentive delivery may be collected separately and stored apart from research data.

How will the data be collected or created?

Chat transcripts will be exported from the chat platform after each session. Pre and post session surveys will be collected via an online tool and exported as CSV.

We will use a standardised session protocol and fixed questionnaires. Data will be stored in open formats: TXT or CSV, plus PDF for any outputs.

A simple structure will separate admin, raw, processed, analysis, and outputs folders. Files will use consistent IDs, and participants will be referenced only by study IDs.

Analysis code will be version controlled in Git. We will pilot the procedure, use automated exports, and run basic completeness and labelling checks after each session, documenting all processing steps.

Documentation and Metadata

What documentation and metadata will accompany the data?

Documentation and metadata will include a short README for the dataset, the study protocol, and a data dictionary.

The README will record the project title, creators and supervisors, contact details, dates of data collection, a description of each file type, folder structure, and access conditions. It will also describe the two study conditions, tasks, and what was collected in each session.

A data dictionary will define all variables in the survey and derived participation metrics, including

item wording, response scales, and any coding schemes for open ended responses. For chat transcripts, a schema file will describe fields such as session ID, participant ID, timestamp, and message text.

Documentation will be created alongside data collection and stored in the same secure project folder, updated after each session and after any processing steps. Analysis code and processing notes will be version controlled.

Ethics and Legal Compliance

How will you manage any ethical issues?

Participants will receive an information sheet and provide electronic consent, including consent for data storage and analysis. Any future sharing will be limited to anonymised or aggregated data and described in the consent materials.

Participants will be assigned study IDs. Contact details will be stored separately from research data. Any quotes will be anonymised and edited to remove identifying details.

Participants will be asked not to share personal data in chat. All files will be stored on secure University systems with access restricted to the research team, and transferred only via approved secure services.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Participants own their chat messages and survey responses. The research team will use them for research under participant consent. Derived metrics, codebooks, and analysis code are owned by the University or research team under University policy.

Only anonymised or aggregated data will be shared, if at all, under a standard University approved licence and or controlled access.

Raw transcripts will not be publicly shared. Any sharing will occur after publication and only in anonymised form.

Storage and Backup

How will the data be stored and backed up during the research?

All research data will be stored on University managed secure storage with access restricted to the research team. No sensitive data will be stored only on personal laptops or external drives.

University storage provides automatic, regular backups and redundancy. A working copy may be kept in a secure project folder, but the authoritative copy remains on managed storage.

The main researcher is responsible for organising files and requesting restores if needed. Backup and recovery are handled by University IT services using their standard incident recovery process.

How will you manage access and security?

Main risks are unauthorised access, accidental disclosure of identifiable data, and insecure transfer. We will minimise personal data, separate contact details from research data, and restrict access to the research team only.

Data will be stored on University managed secure systems using individual accounts, strong passwords, and multi factor authentication where available. Folder permissions will be role based and reviewed periodically.

Collaborators will access files only via approved University services. Data exports from the chat and survey platforms will be transferred directly into secure storage using encrypted connections, and will not be shared via personal email or removable media.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

We will retain anonymised or pseudonymised datasets needed to support publications and verification, including cleaned survey data, derived participation metrics, analysis scripts, codebooks, and documentation. We may also retain anonymised excerpts of transcripts used in analysis.

Raw chat transcripts will be treated as potentially identifiable and will not be openly shared. Contact details used for scheduling or incentives will be deleted once no longer needed.

Data will be retained in line with University policy and funder requirements, then securely deleted or archived as required.

Retained data support auditing and reproducibility, secondary analysis of collaboration patterns, and potential methodological reuse of instruments and coding schemes.

What is the long-term preservation plan for the dataset?

We will preserve an anonymised dataset package (cleaned survey data, derived participation metrics, codebooks, and README) in the University institutional repository or another approved University repository. Raw transcripts will not be openly archived and, if retained, will be stored under restricted access on secure University storage. Repository deposit is expected to have no additional cost. We will budget time to anonymise, convert to open formats (CSV and TXT), and prepare documentation for reuse.

Data Sharing

How will you share the data?

We will share an anonymised dataset package (cleaned survey data, derived participation metrics, codebooks, and README) via the University repository after publication or project completion, making it discoverable through the repository record and related papers. Raw chat transcripts will not be shared publicly due to re identification risk and would only be considered under restricted access on

request if consent permits.

Are any restrictions on data sharing required?

Restrictions are required because chat transcripts may contain personal data and could risk re identification.

To minimise restrictions, we will obtain consent for data use, minimise personal data during collection, and anonymise outputs. We will share only anonymised or aggregated data as standard.

The research team will require exclusive access until publication to protect participant confidentiality and to complete analysis. Any later access to more sensitive materials, if permitted at all, would be managed through restricted access and a data sharing agreement specifying approved uses, no re identification attempts, and secure storage.

Responsibilities and Resources

Who will be responsible for data management?

Qian Shu Wang will be responsible for day to day data management, including data capture and export, file organisation and naming, producing documentation and metadata, anonymisation, and maintaining analysis code and logs.

Dr Malak Sadek and Dr Minja Axelsson will oversee the DMP, review and approve procedures, and ensure compliance with University ethics and data protection requirements. They will also approve any data sharing or archiving decisions.

What resources will you require to deliver your plan?

We will use existing University provision for secure storage, backup, and standard research software.

Additional resources required are minimal and include access to an online survey tool, a text based chat platform for running sessions, and a modest participant incentive budget. An external LLM API is used for the facilitator. We will require API access and a small budget for usage costs.

We will also require a participant incentive budget to cover gift cards, for example £10 per participant (around 20 participants, total about £200).