

Group Project Description and Proposal

Overview:

For the group project, you will be working in a group for the rest of the semester. You will be investigating a dataset of software developer - ChatGPT interactions as part of the MSR 2024 Mining Challenge <https://2024.msrconf.org/track/msr-2024-mining-challenge>. At the end of the semester, you are expected to deliver a high-quality research paper that can be submitted to the mining challenge of the conference. You will have intermediate status reports and deliverables throughout the semester. The project has 4 Phases:

Phase 1 – Project Proposal – due **Wednesday, Oct 11**

Phase 2 – Project Update – due **Sunday, Nov 5**

Phase 3 – Project Final Report – due **Thursday, Nov 30**

Phase 4 – Project Final Presentation – in class during last two weeks of classes (after Thanksgiving) - **Nov 28, Nov 30, Dec 5, Dec 7**

Phase 1 – Project Proposal

As a first step, you need to come up with a project proposal. For the project proposal, you need to define the research questions you will be addressing, and then describe the methodology (tools, analysis, statistical methods, etc.) you will be using to answer the research questions you formulated based on the dataset provided. You can start from the list of research questions listed on the Mining Challenge website (<https://2024.msrconf.org/track/msr-2024-mining-challenge>) for inspiration, but you can choose maximum one question from the list and then you need to come up as a group with other research questions on your own. You should have **at least one research question per person** in your proposal.

After submitting the project proposal, this will be reviewed by the instructor and will need to be approved before you start working on the project. The instructor may request modifications to the scope of the project to ensure the effort is appropriate and the topic is appropriate.

The document should be in **.pdf** format. You can choose your own structure for the document, but the proposal should have at least the following sections:

1. *Research questions and motivation* – list the research questions you will be addressing and explain why do you want to answer them – how can answering the research questions be useful to you or other developers?
2. *Methodology* – describe the methodology you will use to answer the research questions, including any analysis techniques, statistics, data mining, etc. that you plan to use to derive the answers to your research questions from the data.
3. *Work plan* – break down your work in steps and describe each step and their order, and name the people that will be working on each step.

Student and group responsibilities:

We expect that group members collaborate with one another, but that groups work independently from one another, not exchanging results with other groups. We will check for this. Within groups, we expect that you are honest about your contribution to the group's work and that you document all your contributions thoroughly. This implies not taking credit for others' work and not covering for team members that have not contributed to the team. Please mind that **all team members need to contribute to the proposal, as well as all the subsequent work on the project.**

Deliverables and deadlines for the project:

- **Wednesday, Oct 11:** Project Proposal .pdf document due @ 11:59pm on Canvas (**100 points**).
- **Saturday, Oct 14:** You will receive **feedback** on your project proposal from the instructor and TA by **11:59pm on Canvas**.
- **Sunday, Nov 5:** Status report video presentation and report for the project due @11:59pm on Canvas and all code and artifacts up to this point are due on GitHub (**200 points**). You should be at least halfway through the project by this point. The video presentation should describe the state of the project at that point in time, describe the code/methodology/results that have been implemented/obtained so far and discuss what remains to be completed for the project until the final report is due. The report file must closely follow the formatting mentioned in the challenge (<https://2024.msrrconf.org/track/msr-2024-mining-challenge#submission>) and must be a **pdf file** and should describe in detail what each member of the team contributed to the project (what features, code, artifacts, tasks, etc. each member contributed to – **be very specific**). The report should also list any external resources that were used in the project (code fragments, libraries, etc.). Details will follow.
- **Thursday, Nov 30:** Final project report, code, and artifacts due @11:59pm in GitHub and on Canvas (**300 points**) The **GitHub repository** must include a **README file** that explains all the folders/files in the repository (what can be found where). The **final project report** must closely follow the formatting mentioned in the challenge (<https://2024.msrrconf.org/track/msr-2024-mining-challenge#submission>) and must be a **pdf file** and describe in detail all the work that has been done for the project and document in detail what each member of the team contributed to the project (what analysis, code, artifacts, tasks, etc. each member contributed to – **be very specific**). The report should also list any external resources that were used in the project (code fragments, libraries, etc.). Details will follow.
- **Nov 28, Nov 30, Dec 5, Dec 7:** In-class project presentations during lecture time (**100 points**). Each group will get to present their final project in front of the instructor, TA and the other students in the class. The presentation should demonstrate the tool in action and explain the source code implementation or should summarize the methodology used and the results of the study performed. The status of the project and individual contributions will be assessed based on the contributions visible in GitHub and in the project report. Details and schedule will follow.