Research Project Manual

CSCI 49395: Big Data Analytics and Applied Privacy, Fall 2026

1. Project Goal

The Final Research Project accounts for **25% of the course grade** and provides hands-on experience with research in one or more of the course's three main areas:

- **Big Data Systems**
- **Artificial Intelligence and Analytics**
- **Privacy**

Each team will study a research problem based on papers from a renowned conference, work with a real dataset, implement and experimentally evaluate a method, and communicate the results through project presentations and a final report. The project does **not** need to propose a novel method. A careful reproduction and evaluation of published research is an acceptable project.

2. Teams and Topic Selection

Projects will be completed in **teams of three students**. A Project Forum will be opened on Brightspace during the first week. Students should use the forum to:

- Briefly introduce themselves and their technical/research interests
- Find teammates
- Discuss possible topics
- Post their final team and preliminary topic

The instructor will provide a list of suggested project directions. Teams may select and refine one of these topics or **propose their own project**. Student-proposed topics require instructor approval. Projects should be related to at least one of the three main areas of the course, while projects connecting multiple areas are particularly encouraged.

3. Research Paper Selection

Every project must be based on **at least three research papers published in a renowned conference** in Big Data, AI/Analytics, or Privacy. During one of the **class sessions**, the instructor will discuss how to effectively read research papers and identify:

- The problem and motivation
- The main contributions
- The proposed methodology
- The dataset and experimental setup
- Evaluation metrics and baselines
- Main results and limitations

Teams should apply these skills when selecting their paper. Do not select a paper based only on its title or abstract. Before proposing a paper, the team should understand its main contributions and have an initial idea of what can

realistically be reproduced, analyzed, or extended. The instructor and TAs will provide guidance during paper and topic selection.

4. Dataset, Implementation, and Experiments

Every project must include **hands-on work with at least one dataset**. Teams should understand and explore their dataset, including its source, size, structure, important features, and relevant preprocessing. Privacy or ethical issues associated with the data should also be considered when applicable. Every project must also include an **implementation and experimental results**. A literature review alone is not sufficient.

Projects can generally take one of three forms:

Reproduction: Reproduce selected experiments or results from the research paper and compare your results with the published results.

Reproduction + Analysis: Reproduce part of the work and conduct additional experiments, such as changing parameters, evaluating additional metrics, comparing methods, or studying scalability, robustness, or privacy–utility tradeoffs.

Extension: Reproduce an appropriate baseline and extend the work; for example, by using another dataset, modifying the method, adding a privacy mechanism, comparing another approach, or investigating a new experimental question.

Novel research is encouraged but **not required**.

5. Project Progress and Midterm Presentation

The project is a **semester-long activity**. Teams should begin working with their paper, dataset, and implementation early in the semester. The instructor plans to meet with each team **several times during the semester** to discuss progress, challenges, results, and next steps. TAs will also provide support with datasets, tools, implementation, experimental setup, and other technical issues. One class session will be devoted to **Midterm Project Presentations**.

By the midterm presentation, each team should be able to present:

- Research problem and selected conference paper
- Main contributions of the paper
- Dataset and initial data exploration
- What will be reproduced, analyzed, or extended

The midterm presentation should **not be only a presentation of the research paper**. Teams are expected to have obtained and explored their dataset and started implementation by this point.

6. Final Deliverables

Each team will submit:

1. **Final Research Report**
2. **Implementation / Source Code**
3. **Instructions for running the implementation and reproducing the main experiments**
4. **Final Project Presentation**

The final report should be written in the style of a short technical/research paper and normally include introduction and motivation, background and related work, dataset and preprocessing, methodology and implementation, results, analysis and discussion, conclusion, and references. The report must clearly distinguish between the work of the original paper/authors and the work performed by the team.

Using official or publicly available code is permitted when appropriate, but its source must be acknowledged. Teams must understand the code they use and clearly explain what they implemented, modified, tested, or analyzed. Simply downloading and running an existing repository is **not sufficient**.

7. Team Responsibilities and Evaluation Philosophy

All three team members are expected to contribute and understand the overall project. Tasks may be divided among members, but every student should be able to explain the research problem, paper, dataset, methodology, implementation, experiments, and main results.

A successful project should demonstrate:

Research Understanding + Data + Implementation + Experimentation + Analysis

Obtaining better results than the original paper is **not required**. Reproduction can be difficult, and results may differ from those reported by the authors. What matters is that the team conducts meaningful experiments, reports results accurately, compares and analyzes them appropriately, discusses limitations or discrepancies, and demonstrates a clear understanding of the research problem.

Specific deadlines, presentation durations, submission formats, and the detailed grading rubric will be announced on Brightspace.