VCU College of Engineering

CS 25-336

Community-Engaged Research Evaluation: A Computational Framework for Data Collection, Processing, and Visualization

Preliminary Design Report

Prepared for

VCU Wright Center for Clinical Trials

By

Tristan Weigand, Jasper Early, Abdul Koroma, Levi Thompson

Under the supervision of

Amy Olex

Bridget McInnes

12/3/2024

Executive Summary

The Wright Regional Center for Clinical and Translational Science, in collaboration with VCU, VSU, ODU, and EVMS, aims to enhance community engagement and diversify patient populations through Community Engaged Research (CEnR). A key challenge identified is the need to quantitatively measure the impact of these initiatives, as current methods rely on low-response voluntary surveys and varied individual evaluations.

This project seeks to develop a comprehensive system for systematic data collection, processing, and evaluation of CEnR activities. By leveraging an existing dataset classified via natural language processing, the project aims to construct a robust framework that includes a database, an API, and a user-friendly web interface. This system will allow the Wright Center to effectively track, assess, and improve community engagement efforts.

Key objectives include:

- Utilizing the existing Institutional Review Board (IRB) dataset while ensuring the design is generalizable and accessible to non-programmers.
- Implementing standards that adhere to data sensitivity protocols as outlined by VCU's security and data handling guidelines.

Deliverables will encompass database designs, an API framework for data integration, and a graphical user interface for visualization and analysis. Milestones will guide the project from database design to final deployment, ensuring stakeholder feedback and continuous improvement.

Ultimately, this initiative aims to enhance the efficiency of the Wright Center’s evaluations of CEnR programs, enabling data-driven decisions that foster deeper community engagement and effective research partnerships.

Table of Contents

Section A. Problem Statement

The Wright Regional Center for Clinical and Translational Science, henceforth referred to as the Wright Center, is a regional research center formed of a collaboration between VCU, VSU, ODU, and EVMS. The Wright Center's primary goals are "to grow community engagement, diversify patient populations and encourage greater diversity among new clinician researchers entering the workforce" [1]. As an element of its community engagement goals, the Wright Center facilitates community engaged research (CEnR) which involves community members in research processes through a variety of methods and interactions. Community engaged research is predicated on the principle that community members can provide unique perspectives and insights, are able to identify problems and research areas which are of priority to their community, and are able to improve the success of implementing research-driven changes, solutions, and interventions through their involvement in a community [2]. Through CEnR, the Wright Center aims to drive authentic engagement between researchers and the community, applying principles that leverage data to drive behavioral, social, service, or policy changes, and increase the capacity of both communities and researchers.

Researchers and administration at the Wright Center would like to be able to quantitatively measure the impact of their CEnR initiative. Traditionally, data about CEnR participation is measured through voluntary surveys that return low response rates [3]. In addition, these surveys are independently entered by researchers and are subject to variation in individual evaluations of the degree or type of community involvement in a research project. Other sources of data, such as IRB submissions and manuscripts, could be theoretically evaluated to determine the degree of CEnR participation through either manual or automated (natural-language processing) methods. However, there does not currently exist a sole repository for these composite sets of data in a way which is readily accessible for processing and evaluation. Therefore, the primary goal of this project is to construct a framework of a database, an API, and a GUI which enables research data to be stored and processed so that the Wright Center may better evaluate the reach and impact of its CEnR directive on a regular basis. Our aim is to provide a method for the Wright Center's CEnR team to compile and take advantage of existing data without any requirement for significant internal technical knowledge.

This project is not the first data-processing project in collaboration with the Wright Center. However, prior projects in this area have been focused on the development of machine learning models to directly classify papers by their CEnR participation level. For example, a previous group has published a number of research papers from VCU detailing the use and training of a number of distinct machine learning algorithms in order to classify a set of research papers into six CEnR categories. The selected categories are no partnership, non-CEnR partnership, instrumental partnership, academic-led partnership, cooperative partnership, and reciprocal partnership. To develop the rules for classification, the previous group made use of Institutional Review Board (IRB) protocols to evaluate and process this set of papers to produce a classification model and a classified dataset [3] [4] [5]. These projects also required pre-processing of data in order to produce a usable dataset for these models, and this requirement

for pre-processing is a key motivator for our project. Our project makes use of this same original dataset, which will henceforth be referred to as either the IRB dataset or simply as our existing or original dataset, as well as other potential sources of research manuscripts and other relevant data types. However, rather than to work with any such models directly, our goal is to enable researchers to more easily train and run such models in the future and to reduce future requirements for data pre-processing.

Our core design will consist of a database, an API, and a web user interface. The aims of our design are to enable collection, processing, analysis, and visualization of data relevant to the CEnR initiative. While we will construct our database with the IRB dataset and scientific publications in mind, we will also ensure generality such that data from external sources and data of different input formats can be easily added, formatted, and processed. For use of this data, the API will serve to allow algorithms and models to be run on the data stored within our constructed database and return a well-formatted output for use by programmers and researchers, while the graphical user interface will include visualization elements for non-programmers such as graphs, charts, and analytical tools. Through these tools, we aim to greatly improve the efficiency of the Wright Center's evaluations of CEnR programs and allow data-driven decisions on policy and engagement strategies for greater community engagement in research.

Section B. Engineering Design Requirements

B.1 Project Goals (i.e. Client Needs)

The overall goals of this project are focused on addressing the client's needs to evaluate and improve their community-engaged research (CEnR) initiatives. These goals stem from the need to track, assess, and enhance both current and future research efforts, particularly in the context of community involvement and partnership. The project seeks to establish methods for better capturing data, utilizing external resources, and improving the reporting and impact measurement of CEnR activities. By identifying key metrics, building new models, and expanding data collection across multiple institutions, the client aims to strengthen their community engagement and collaboration efforts. The following goals outline the primary focus areas for this project:

- **Track and identify impact** to measure trust, partnerships, and impact on an annual basis.
- **Develop a database** with a modular and extensible approach similar to previous models.
- **Enhance data management and APIs** to integrate new metrics and research activities more efficiently.
- **Ingest Prototype datasets** of researchers involved in CEnR.
- **Develop Scalable API** that can be augmented to ingest datasets like NIH Reporter, NSF, and DoJ.
- **Leverage Natural Language Processing (NLP)** to analyze CEnR data and metrics.
- **Help Visualize and Implement metrics** for CTSA Reporting (RPRR) to track community engagement research.
- **Facilitate measurement of community engagement** through publications, protocols, and direct involvement with community partners.
- **Implement and visualize longitudinal metrics** to identify significant changes in CEnR activities.
- **Expand data collection** to include additional institutions and partnerships within the Wright Center and the university.
- **Incorporate bibliometrics and social network analysis** to track author collaborations and community engagement impact.

B.2 Design Objectives

The key objectives of the design are focused on ensuring that it meets the client's needs in terms of functionality, accessibility, and future expansion. These objectives are crafted to be SMART—Specific, Measurable, Achievable, Realistic, and Time-bound—so that they can be clearly defined and integrated into the design specifications during the process.

The following objectives will guide the design:

- The design will make use of the existing IRB dataset
- The design will be generalizable such that the database can be expanded
- The design will be accessible to non-programmers
- The design will be extensible by later programmers
- The metadata of the database will follow FAIR principles
- The design will implement common data models in its framework

B.3 Design Specifications and Constraints

Our design is to run on VCU's Athena HPRC Cluster and will handle Category II information. HPRC here stands for High Performance Research Computing. VCU's HPRC core operates two supercomputing clusters, Apollo and Athena. Each cluster consists of a series of nodes each with respective CPUs, GPUs, RAM, and so on. The Apollo Cluster was made to be used for Category I data, which is more classified than the data we will be using in this project, Category II. Thus we are using the Athena Cluster.

VCU divides data that it operates on into three categories based on how sensitive that data is. Category I data is data which is considered confidential and regulated. Category II data is data which is considered sensitive yet does not fall under any regulations. Category III data is data which is public and thus does not require protection. Our project is primarily concerned with research information which is considered Category II. Category II data requires password protection and cannot be shared with the public without permission. These requirements are met by handling our data exclusively on the Athena Cluster which stores the data itself and requires a VCU account to access.

- Design should be able to operate on VCU's Athena HPRC Cluster unless that becomes unrealistic for the project and some other solution becomes more reasonable (Processing Constraint)
- Design must not require constant computation needs on the cluster but will require storage (Processing & Storage Constraint)
- Design must have some protective constraints on potentially sensitive data, such as IDs, which may be linked to source documents (Data Constraint)
- Design must operate within a set of systems or platforms requiring compatibility with specific protocols or APIs (Interoperability Constraint) – SQL protocols, API of NIH (and other similar research/grant databases, eg. PubMed & Grants.gov APIs)

B.4 Codes and Standards

VCU has a long list of standards relevant to information technology. The subset of these standards which follow are one which are immediately relevant to our design, an application which uses Category II information.

- **VCU Application Security Standard I10 – Ensure encrypted session**
  - All sessions and all data transmitted during sessions must be encrypted
  - Record of "analysis and compliance" must be maintained
- **VCU Application Security Standard I15 – Physical or logical separation of application and database**
  - User interface must be physically or logically separated from backend data
- **VCU Data Handling And Storage Standard H51 – Devices storing or accessing data must not be located on a publicly accessible network without any form of reasonable authentication and authorization**
  - Data here being Category II

Section C. Scope of Work

C.1 Deliverables

Project Deliverables include not only the final goal products, but also documentation, designs, and code produced during earlier phases of the project. Many deliverables therefore serve as proof of the design process rather than goals of that process.

All project deliverables will be saved in at least one of the following three locations: the project's EduSourced page, the project's github page, and a shared google drive folder used for shared access to working copies of word documents as well as project background information. Many deliverables will be found in all 3 locations, and final versions for all deliverables will be available in EduSourced and/or from our github page.

All currently planned deliverables can be done remotely, as the nature of our project is a software design project which does not require physical access to any specified hardware. Some deliverables may require access to the VCU network, which can be accessed remotely by usage of the VPN which VCU makes available to students and faculty. For example, the VCU High Performance Research Computing (HPRC) core is only accessible through the VCU network or VPN, and the HPRC is where our data is hosted on the Athena cluster.

- Deliverable Group 0: Academic Deliverables
    - Team Contract
    - Project Proposal (This current document)
    - Preliminary Design Report
    - Fall Poster & Presentation
    - Final Design Report
    - Capstone EXPO poster & presentation
- Deliverable Group 1: Database Design
    - A set of possible database design structures. May be ideal to present multiple alternative designs with different sets of fields that can be compared for generality and coverage over not only IRB data but also any appropriate data that a user later wishes to add.
    - A software database and its associated design document serve as the final deliverables in this group.
- Deliverable Group 2: API

    - A planned API design (or proposed set of designs) for use by later programmers to run models and algorithms on/over our database.

    - A user manual with documentation of each API method. The API code and documentation will serve as a combined deliverable.
    - An API framework which can be used to obtain and format appropriate data from external datasets and incorporate that data into our designed database. This deliverable also will include documentation instructing how to extend this portion of the api to function with additional databases.

- Deliverable Group 3: Interface (Likely a web interface)
    - A front-end graphical user interface application to access the database. Also will likely include specific database views and statistical visualizations.
    - Currently planned to be a web-hosted application accessible from an authenticated user's browser. Potentially subject to change during the design phase for this portion of the project.
    - A functional prototype of this interface will be developed and tested before the live version will be deployed.
    - The code on which the web application runs will also be submitted as an independent deliverable (submitted separately from the web access interface)

C.2 Milestones

Project milestones are the key stages of the project and checkpoints to ensure the process is running smoothly. All milestones and their status can be found in three locations: the project's EduSourced page, the project's github page, and the aforementioned shared google drive folder.
To begin, the group has met with our sponsor and faculty advisor and assigned roles and responsibilities. Abdul has taken the responsibility of project manager; Tristan is the test engineer; Jasper is our main systems & manufacturing engineer; Levi is our logistics engineer and scribe. Due to the nature of this project, everyone is responsible for systems & manufacturing in one way or another.

Milestones for this project can be broken down for every major milestone or deliverable:

- Major Milestone 1: Design Database
    - Design framework
- Major Milestone 2: Design API
    - Rough plan/outline; assign specific responsibilities to group as needed
    - Implementation of design as a group
    - Test API

- Major Milestone 3: Design Wireframe
- Major Milestone 4: User-Friendly Interface
    - Decide interface format (website, mobile app, etc.)
    - Assign specific responsibilities for designing interface
    - Implement and test as a group
    - Submit for feedback from client before allowing to be live/accessible to all

C.3 Resources

The Wright Center and partners have already obtained valuable data and models that will be used in our project or as guidance. One of note are the IRB datasets that we will be using. Additionally, resources such as the Translational Science Benefits model, Social Connection Model, and bibliometrics literature have been shared with us. We also intend to search Virginia agencies, such as NIH Reporter, NSF, DoJ, PCORI, etc.

Section D. Concept Generation

Figure 1. Overall Database Schema. Figure developed as a team effort.

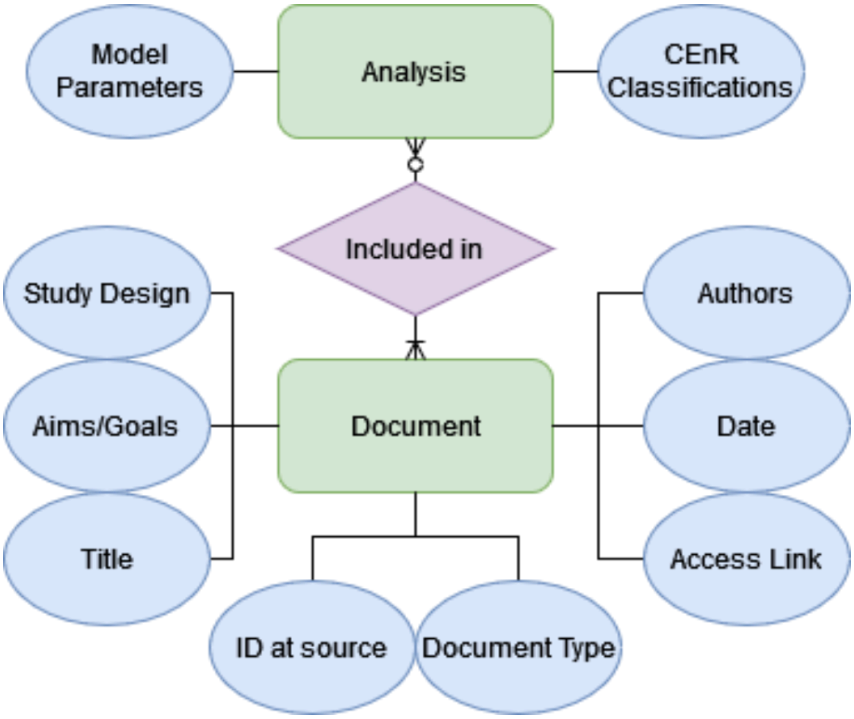


Figure 2. Natural Language Processing Pipeline. Figure developed as a team effort.

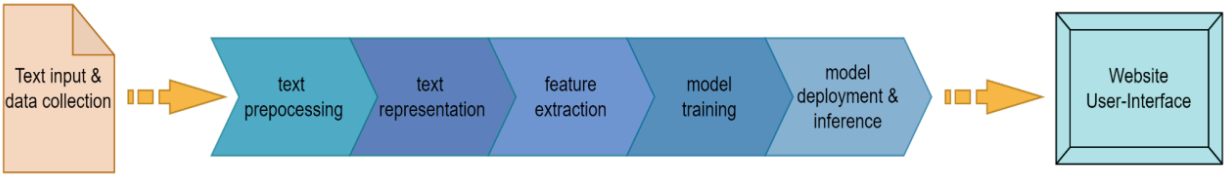
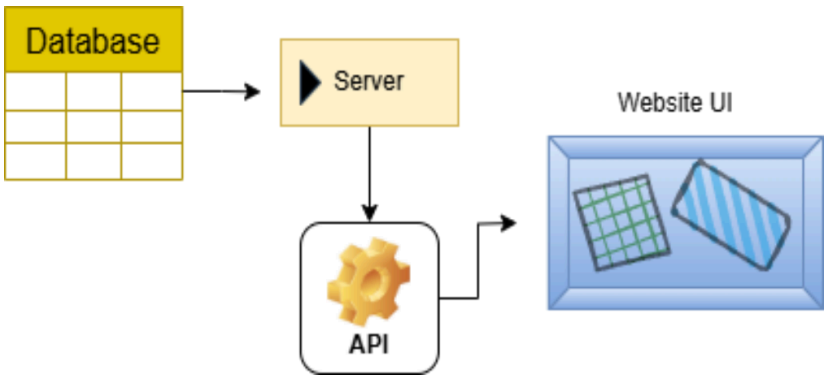


Figure 3. High Level view of server workflow. Figure developed as a team effort.



Section E. Concept Evaluation and Selection

Section not applicable to this project per mentor and faculty advisor.

Section F. Design Methodology

We will use agile design methodology which prioritizes delivering value early and continuously by breaking the design process into smaller, manageable increments. Each iteration .involves planning, designing, prototyping, and gathering user feedback, ensuring the final product aligns closely with user needs. An initial prototype pipeline will be developed per the design schematic shown in Section D. Per our mentor and advisor no further additional detail is required for this section.

Section G. Results and Design Details

Per our mentor and advisor, this section is not applicable at this stage of the project.

Section H. Societal Impacts of Design

Researchers and administration at the Wright Center would like to be able to quantitatively measure the impact of their Community-Engagement Research (CEnR) initiative. CEnR participation data is measured through voluntary surveys, Institutional Review Board (IRB) submissions and manuscripts, grants, and publications. Towards this end, this project will have the following impacts:

- User Interface – Graphical user interface to display visualization elements for non-programmers such as graphs, charts, and analytical tools
- Database – Store text documents to use for information mining and retrieval
    - This allows researchers to analyze a large batch of raw data, identify patterns, and extract useful information
- API – Allows integration between new applications and existing systems, enabling a seamless data exchange and easy scaling of applications
    - Researchers can efficiently scrape papers from sources such as PubMed, interface with other databases to easily pull data, and interact with machine learning models

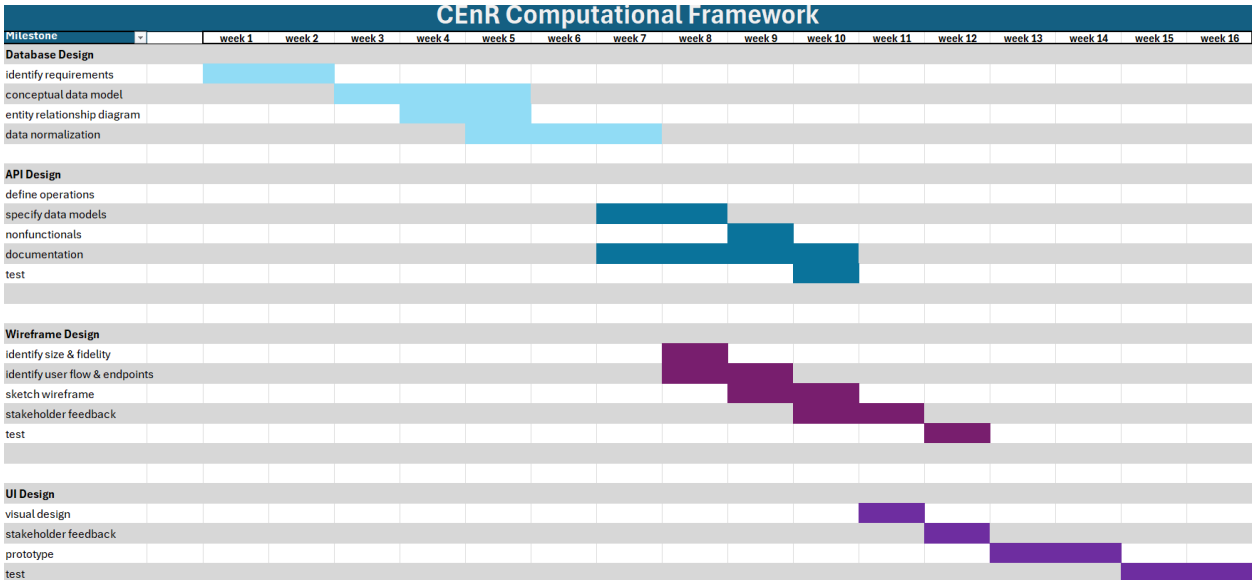
Section I. Cost Analysis

There are no expenditures.

Section J. Conclusions and Recommendations

Per our mentor and advisor, this section is premature.

Appendix 1: Project Timeline



Appendix 2: Team Contract (i.e. Team Organization)

Team Contract

Step 1: Get to Know One Another. Gather Basic Information.

Team Member Name	*Strengths each member bring to the group*	*Other Info*	*Contact Info*
Jasper Early	Working with team, being ready when needed	I'm glad to be working on a project like this	earlyj@vcu.edu
Tristan Weigand	Flexible schedule, wide software knowledge, comfortable with long work sessions	I'm eager to learn new software tools and solve real-world problems	weigandta@vcu.edu
Levi Thompson	Organization, team-worker and solo-worker, flexible	I have no preference over what language I work in (so far). Excited to work together!	thompsonle2@vcu.edu
Abdul Koroma	Experience, Communication, Team building, and conflict resolution.,	I am very eager to work with you guys.	Koromaar2@vcu.edu

Other Stakeholders	*Notes*	*Contact Info*
Bridget McInnes	*faculty advisor*	btmcinnes@vcu.edu
Amy Olex	*sponsor* *org* - VCU Wright Center for Clinical Trials	alolex@vcu.edu

Step 2: Team Culture. Clarify the Group's Purpose and Culture Goals.

Resources: More information and an example Team Culture can be found in the Biodesign Student Guide "Intentional Teamwork" page (webpage | PDF)

Culture Goals	***Actions***	***Warning Signs***
Being on time to every meeting, and if a meeting may be missed, give reasonable forewarning	- Set up meetings in shared calendar & discord events - Send discord ping reminder day before meeting	- Student misses first meeting, warning is granted - Student misses meetings afterwards – issue is brought up with faculty advisor
Informing the group of any delays in completing assignments	- Stay up to date with each other's project responsibilities - Set reasonable deadlines and note when an extension is needed	- Student shows up for weekly meeting with no considerable work done - Low communication – student may be spoken to individually
Have weekly objectives and meet those weekly objectives	- Be proactive and consistent - Work together to set objectives during each weekly meeting - Use dedicated channel to keep track of own and team objectives - If an objective broadens in scope, address it early	- Falling behind and or not meeting one objective. - laziness or lack of ambition (avoidance of equivalent work)

Step 3: Time Commitments, Meeting Structure, and Communication

Meeting Participants	***Frequency*** ***Dates and Times / Locations***	***Meeting Goals*** ***Responsible Party***
Students Only	4pm Wednesdays, on Discord	Check in with each other Go over previous week's work Prepare for meeting with advisor + sponsor
Students Only	Weekends as needed	Actively work on project
Students + Faculty advisor	6pm Thursday	Update faculty advisor and get answers to our questions (** will scribe; ** will create meeting agenda and lead meeting)
Project Sponsor	6pm Thursdays	Update project sponsor and make sure we are on the right track (** will scribe; **will create meeting agenda ** will present prototype so far)

Step 4: Determine Individual Roles and Responsibilities

Team Member	***Role(s)***	***Responsibilities***
Abdul	Project Manager	- Keeping the team on task - tracking the upcoming progress of our group - Making sure the well being of the group is good
Jasper	Systems & Manufacturing Engineer	- Understand project goals and existing systems - Work to iterate and improve systems - Assist in implementation of designs
Levi	Systems & Logistics Engineer	- Coordinate with teammates - Obtain relevant information needed for team and project - Assist in developing system architecture and interfaces - Analyze initial design specifications and lead establishment of design and product
Tristan	Test Engineer	- Determine critical features and functionality - During design & planning, determine potential issues with framework and ensure design is theoretically sound - During implementation, create automated tests to ensure proper functionality

Common Leadership Roles for Capstone

1. **Project Manager:** Manages all tasks; develops overall schedule for project; writes agendas and runs meetings; reviews and monitors individual action items; creates an environment where team members are respected, take risks and feel safe expressing their ideas.
   Required: On Edusourced, under the Team tab, make sure that this student is assigned the Project Manager role. This is required so that Capstone program staff can easily identify a single contact person, especially for items like Purchasing and Receiving project supplies.
2. **Logistics Manager:** coordinates all internal and external interactions; lead in establishing contact within and outside of organization, following up on communication of commitments, obtaining information for the team; documents meeting minutes; manages facility and resource usage.
3. **Financial Manager:** researches/benchmarks technical purchases and acquisitions; conducts pricing analysis and budget justifications on proposed purchases; carries out team purchase requests; monitors team budget.
4. **Systems Engineer:** analyzes Client initial design specification and leads establishment of product specifications; monitors, coordinates and manages integration of sub-systems in the prototype; develops and recommends system architecture and manages product interfaces.
5. **Test Engineer:** oversees experimental design, test plan, procedures and data analysis; acquires data acquisition equipment and any necessary software; establishes test protocols and schedules; oversees statistical analysis of results; leads presentation of experimental finding and resulting recommendations.
6. **Manufacturing Engineer:** coordinates all fabrication required to meet final prototype requirements; oversees that all engineering drawings meet the requirements of machine shop or vendor; reviews designs to ensure design for manufacturing; determines realistic timing for fabrication and quality; develops schedule for all manufacturing.

Step 5: Agree to the above team contract

Team Member:	*Signature: Tristan Weigand*
Team Member:	*Signature: Levi Thompson*
Team Member:	*Signature: Jasper Early*
Team Member:	*Signature: Abdul Koroma*

Note: Drs. Olex and McInnes have approved this final report for Fall 2024. All portions relevant to the project at this stage are completed.

References

[1] VCU Writing Center. (2021, September 8). *APA Citation: A guide to formatting in APA style.* Retrieved September 2, 2024. https://writing.vcu.edu/student-resources/apa-citations/

[2] Teach Engineering. *Engineering Design Process*. TeachEngineering.org. Retrieved September 2, 2024. https://www.teachengineering.org/populartopics/designprocess