VCU College of Engineering

CS 25-348 RenovationTracker

[Preliminary Design Report]

Prepared for

Syed Khajamoinuddin

CoStar Group, Inc

By

Courtney Van

Jaidon Lightfoot

Jonathan Netala

Sterling Glasheen

Under the supervision of

Preetam Ghosh

12/9/24

Executive Summary

The **RenovationTracker: Intelligent Renovation Detection and Visualization** project aims to revolutionize how property renovations are showcased in the real estate market. Developed in partnership with **CoStar Group**, this initiative addresses the shortcomings of static descriptions and uninspiring visuals by creating an AI-powered system capable of extracting, classifying, and presenting property renovations dynamically. By improving the clarity and engagement of property listings, RenovationTracker seeks to enhance decision-making for buyers and sellers alike.

At its core, the project integrates two advanced technologies:

1. A **Natural Language Processing (NLP) engine**, powered by OpenAI, analyzes property descriptions to extract details of recent renovations.
2. A **computer vision model**, built with PyTorch, classifies and identifies rooms from property images to provide context for the renovations.

The outputs of these two components are orchestrated using **AWS Step Functions**, which unify the data into a cohesive JSON structure. This final output links renovations to their respective rooms and associated images, creating a comprehensive, visually rich property listing. The results are then integrated into a **Node.js backend** and displayed on a **React-based frontend**, providing seamless compatibility with existing infrastructure.

Key deliverables include:

- A scalable, production-ready system capable of detecting and visualizing renovations from property descriptions and images.
- A user-friendly property listing interface that highlights renovations dynamically, allowing for better user engagement.
- A stretch goal feature: a side-by-side renovation comparison tool with interactive sliders and overlays to further enhance the user experience.

Work completed to date includes the design and implementation of the web scraper to gather property data from **homes.com**, the development of the two Lambda functions for NLP and room classification, and the integration of Step Functions for workflow orchestration. Initial testing of these components is ongoing, with plans to scale and refine the system as development progresses.

Looking ahead, the next phases include integrating the Step Functions output with the React frontend, expanding testing to ensure robustness, and iterating based on feedback from our sponsor.

This project aligns with CoStar Group's strategic goals by leveraging state-of-the-art AI technologies to enhance the presentation of property data, providing a competitive advantage in the real estate market. With its innovative approach and scalable design, RenovationTracker positions itself as a game-changing tool for the industry.

Table of Contents

Section A. Problem Statement

In the competitive real estate market, effectively highlighting property renovations remains a persistent challenge for buyers, sellers, and real estate professionals. Existing methods often rely on static textual descriptions and a limited number of before-and-after images, which fail to provide the clarity and context needed to fully appreciate the scope and quality of renovations. This problem impacts a wide range of stakeholders, including homeowners, prospective buyers, property managers, and real estate agencies. Without accurate and dynamic representations of renovations, potential buyers may disengage, sellers may struggle to showcase value-adding features, and inefficient decision-making may prevail.

RenovationTracker seeks to bridge these gaps by introducing an intelligent system that automates the detection, classification, and visualization of property renovations. Leveraging cutting-edge AI technologies such as natural language processing (NLP) and computer vision, the system extracts renovation details from property descriptions and classifies room types from images. These outputs are seamlessly unified using AWS Step Functions, producing dynamic JSON-based data that links renovations to their corresponding rooms and images. The system significantly enhances clarity, interactivity, and user experience in property listings.

This innovative solution addresses key pain points:

- For **buyers**, it provides dynamic, visually engaging property listings that improve decision-making.
- For **sellers**, it highlights property value through automated, accurate renovation details.
- For **real estate professionals**, it streamlines the property listing process, reducing manual effort while ensuring high accuracy.

The inefficiency of manually annotating renovations and classifying rooms is a prevalent issue across the industry. This process is time-consuming, prone to errors, and lacks scalability. The economic impact is significant, with poorly presented properties often seeing reduced market value and extended listing times. RenovationTracker's automated workflow alleviates these challenges, presenting clear, engaging renovation data that increases user engagement and improves market competitiveness.

While similar technologies exist in other industries, RenovationTracker stands out as a cohesive solution specifically tailored to real estate. By combining NLP for renovation detection and PyTorch-based computer vision for room classification, it sets a new standard for how renovations are detected and showcased. The resulting data is optimized for integration with real estate platforms like MLS, and its interactive visual tools, such as side-by-side renovation comparisons, further enhance user engagement.

The project is not without its challenges. Ensuring the accuracy of AI models across diverse property data and images remains a priority. Additionally, integrating these models into a scalable, production-ready system that aligns with existing real estate workflows is essential for widespread adoption. Progress to date includes the implementation of two Lambda functions, integration of AWS Step Functions, and initial testing of the system's capabilities. These advancements affirm the feasibility of the solution.

In conclusion, RenovationTracker addresses an unmet industry need by transforming static property descriptions into dynamic, visually rich experiences. The system's ability to automate renovation detection and room classification, coupled with its scalability and compatibility with existing platforms, will redefine how property listings are created and consumed, leading to improved decision-making, greater user engagement, and higher property values.

Section B. Engineering Design Requirements

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

The primary objective of Renovation Tracker project, is to improve the way individuals identify and display property renovations in the Real Estate industry, through an end-to-end machine learning system. It is intended to enhance user experience, through this AI powered solution, combining the use of Natural Language Processing (NLP) and computer vision for the automated renovation detection and room classification.

Key goals should include:

- Precise Recognition: Create an AI system that can automatically recognize and classify renovations from descriptions and images of properties.
- Enhanced Visualization: Side-by-side visualization control for users to compare property images before and after renovations.
- User Interface: Enhance your ability to make good decisions by showing you in a clear and intuitive fashion what's really happening to your property.
- Scalability: Ensure the platform is scalable to meet future industry demands and evolving technologies.

B.2 Design Objectives

- The design will identify and categorize property renovations from text descriptions and images correctly with at least 90% accuracy
- The design will enable users to compare property images side-by-side to view renovations, seamlessly and interactively.
- It will seamlessly work with existing real estate listing sites while being compatible with APIs of listing sites such as CoStar.
- The design will accommodate scale by handling very large property databases and huge volumes of image data.
- The design will be delivered by the deadline, respecting the budget and making optimal use of a cloud environment such as AWS.

B.3 Design Specifications and Constraints

- Integration with existing APIs: The system needs to be integrated with CoStar's APIs to retrieve property descriptions and images, so that the flow of data between platforms can go as planned.
- Scalability constraint: Support for up to 10,000 properties and related renovation data can be incorporated into the system without affecting performance drastically.
- Paraphrased AI accuracy requirement: The Natural Language Processing and image classification models must have at least 90% accuracy in recognizing and categorizing renovations.
- Cost constraint: The solution must be created within the budget of less than $10,000 in cloud resources, development and deployment.
- Security and Privacy requirements: The platform must be compliant with GDPR requirements, meaning that the storage and handling of delicate property-related data should be secure.
- Usability: the user interface should be intuitive, so that a non-IT user can operate a simple tool for exploring renovation data.

B.4 Codes and Standards

- ISO/IEC 27001 – Information Security Management: assures that sensitive property data is managed safely from end to end.
- ISO/IEC 25010 – Software Quality Requirements and Evaluation (SQuaRE): to ensure that the system is constructed so as to meet quality requirements for functionality and reliability.
- GDPR (General Data Protection Regulation): Tells the system to make sure it's safeguarding your users' and clients' information and complying with privacy laws.
- REST API Standards (RFC 2616)**: Ensures interoperability with web-based messaging protocols for integration with CoStar's APIs.
- ISO 9241-210 – Ergonomics of Human-System Interaction: Concerned with the usability of the user interface (eg, is it easy to access and navigate?).

Section C. Scope of Work

The scope of the RenovationTracker project is to design, develop, and deploy an AI-powered system that automatically detects, classifies, and visualizes property renovations in real estate listings. This project aims to integrate Natural Language Processing (NLP) and computer vision technologies into a unified platform to enhance property insights, automate room classification, and provide dynamic renovation visualization tools. The project will focus on delivering the following key objectives:

- AI-Powered Renovation Detection: Develop an NLP engine to extract renovation-related information from property descriptions.
- Automated Room Type Detection: Create a computer vision model to classify room types based on images provided in property listings.
- Interactive Renovation Visualization: Implement side-by-side comparison tools with sliders and overlays to visually compare rooms before and after renovations.
- System Integration and Testing: Ensure all components are integrated into a seamless platform, thoroughly tested for accuracy and usability.

Project Boundaries and Responsibilities:

The RenovationTracker team is responsible for the design, development, and testing of the core system functionalities, including text analysis, room classification, and renovation visualization. The team will also deliver academic materials, such as the project proposal, reports, and presentations, to meet academic deliverables.

Out-of-Scope Tasks: The following tasks fall outside the scope of the team's responsibilities:

- Development of any external tools or integrations beyond the RenovationTracker system.
- Handling server deployment logistics or costs, as this will be managed by CoStar.
- Collecting or generating new property images or data, as the project will rely on data provided by CoStar or publicly available datasets.
- Any physical hardware development or interaction with real estate platforms beyond the software deliverables outlined.

Project Timeline and Milestones:

The project follows a structured timeline with defined milestones to ensure timely delivery of all objectives. The key milestones include:

- Phase 1: Requirements Gathering & System Design (Proposal phase, 4-6 weeks)
    - Completion of the project proposal, defining system architecture, and selecting relevant datasets.
- Phase 2: Model Development & Initial Prototypes (6-8 weeks)
    - Development of the NLP engine and image classification model, followed by internal testing.
- Phase 3: System Integration & Mid-term Testing (4-6 weeks)

    - Integration of NLP and image classification components into a functional prototype.
    - Initial system testing to ensure that components work together seamlessly.
- Phase 4: Final System Testing, Feedback Integration & Refinement (4-6 weeks)
    - Conduct thorough testing, obtain user feedback, and refine the system for deployment.
- Phase 5: Final Presentation & Deliverables (2-4 weeks)
    - Complete the final design report, prepare presentations, and deliver the Capstone EXPO poster and presentation.

Development Methodology:
The RenovationTracker project will follow an Agile methodology to ensure flexibility and adaptability throughout the development process. This approach will involve iterative development cycles, with regular sprint planning, code reviews, and progress meetings with both the project sponsor (CoStar) and faculty advisors. Feedback from CoStar and internal testing will be incorporated continuously to refine the system.

Timeline Boundaries and Risk Management:
The project is expected to be completed by the end of the academic year, with all deliverables submitted according to the academic calendar deadlines. Given the Agile approach, the project will be flexible in managing unforeseen challenges or risks. To avoid scope creep, any changes in project scope will be discussed with CoStar and faculty advisors before implementation, ensuring alignment with the project's key objectives and timeline.

CoStar will provide server resources for hosting the final system, and no additional funding will be required. By proactively securing necessary tools and licenses, the team aims to avoid delays due to resource shortages or third-party dependencies. Regular communication between the team, sponsor, and advisors will be maintained to ensure the project stays within scope and on schedule.

C.1 Deliverables

The RenovationTracker project will deliver an AI-powered system that enhances property listings by automatically detecting, classifying, and visualizing renovations. The primary deliverables include a Natural Language Processing (NLP) model that extracts and analyzes renovation-related information from property descriptions, coupled with algorithms to detect renovation-specific keywords and contexts. This system will also integrate with a computer vision-based image classification model that can automatically identify and categorize room types (e.g., kitchen, bathroom) from property images. Additionally, a side-by-side renovation visualization tool will allow users to compare room images before and after renovations, featuring interactive elements such as sliders and overlays to highlight changes. The entire system, consisting of text analysis, image classification, and visualization tools, will be integrated into a unified platform, thoroughly tested for accuracy, performance, and usability. If time permits, the team will also implement a user feedback system, allowing users to verify or correct detected renovations and room classifications, contributing to the continuous improvement of the system.

Regarding academic deliverables, the team will provide a team contract, project proposal, preliminary design report, fall poster and presentation, final design report, and Capstone EXPO poster and presentation.

Most of the project work can be completed remotely. This includes coding, model development, system testing, and writing deliverables, which can be done using shared drives like Google Drive or OneDrive, and version control platforms like GitHub. The team will also leverage CoStar's cloud-hosted servers for deploying and running the RenovationTracker system, minimizing dependency on on-campus resources. However, certain tasks such as integration testing, poster printing for presentations, or final presentations may require occasional access to campus facilities. The team members who regularly access campus will coordinate these tasks to ensure they are completed on schedule.

To ensure remote work proceeds smoothly, all necessary software licenses and machine learning tools will be secured early. This includes access to Python libraries and cloud computing resources such as AWS for model training and deployment. The team will need collaborative tools like Slack or Zoom for regular meetings and progress tracking, ensuring effective remote coordination.

The project does not heavily rely on third-party vendors or components with extended lead times, as it is primarily software-based. However, to mitigate any unforeseen risks, the team will proactively manage resources such as cloud computing credits and any necessary machine learning frameworks. By planning ahead and securing these resources in advance, potential delays related to supply chain disruptions will be avoided. Additionally, the team will work closely with CoStar to ensure server access is consistently available and properly maintained for the deployment and testing phases.

C.2 Milestones

The first milestone for the RenovationTracker project is the Project Proposal & Requirements Gathering, where the team will finalize the project proposal, define the system architecture, and gather data for both the NLP and image classification models. This phase is expected to take 4 weeks, with a completion date set for October 15, 2024.

Following that, the team will focus on NLP Model Development, which involves building and training the NLP engine to extract renovation-related information from property descriptions. This milestone will require 3 weeks and is anticipated to be completed by November 5, 2024.

Next, the team will work on the Image Classification Model Development, where the computer vision model to classify room types from property images will be developed. This phase will take 3 weeks, with a target completion date of November 26, 2024.

The Initial System Integration will follow, where the NLP engine and image classification model will be integrated into a unified system. This task will take 2 weeks and is expected to be finished by December 10, 2024.

After integration, the team will move to Preliminary System Testing, testing all system components for functionality, accuracy, and performance. Internal debugging and improvements will be made during this 2-week phase, with a completion date of January 10, 2025.

The Mid-term Report & Presentation will involve preparing and submitting the mid-term report, as well as presenting the current project status to sponsors and faculty advisors. This milestone will take 2 weeks, with a deadline of January 24, 2025.

The team will then focus on the Side-by-Side Renovation Visualization Tool, where an interactive tool for comparing before-and-after renovation images, including sliders and overlays, will be developed and tested. This will take 3 weeks and should be completed by February 14, 2025.

If time allows, the team will work on the User Feedback Mechanism Development (a stretch goal), implementing a feedback system for users to confirm or correct renovation detections and room classifications. This task is estimated to take 2 weeks and would be completed by February 28, 2025.

Next is Final System Integration & Refinement, where the team will fully integrate all components, refine the system based on user feedback, and finalize the platform for deployment. This phase will take 3 weeks, with a completion date of March 20, 2025.

The Final Testing & Quality Assurance phase will follow, where the team will conduct thorough testing to ensure the system's accuracy, usability, and performance under various scenarios. This will take 2 weeks and is expected to be completed by April 3, 2025.

Finally, the team will work on the Final Report & Capstone EXPO Preparation, where they will prepare the final design report, as well as the poster and presentation materials for the Capstone EXPO. This phase will take 3 weeks, with a deadline of April 24, 2025.

The last milestone is the Capstone EXPO Presentation, where the team will present the project at the Capstone EXPO. This will take 1 week and is scheduled for completion by April 30, 2025.

C.3 Resources

The RenovationTracker project requires several key resources for successful completion, which will either be provided by the project sponsor, CoStar, or made available through existing tools and platforms. These resources are categorized into software, hardware, and data resources as outlined below.

Software Resources:

- Natural Language Processing Libraries: The project will require access to NLP libraries such as spaCy, NLTK, or Transformers (Hugging Face) for extracting and analyzing renovation-related data from property descriptions. These are open-source libraries that can be used free of charge.
- Computer Vision Frameworks: The development of the room type classification model will utilize computer vision libraries such as TensorFlow, Keras, or PyTorch. These frameworks will be necessary for training and deploying machine learning models for image classification.
- Integrated Development Environment (IDE): The team will require access to coding environments such as PyCharm, Visual Studio Code, or Jupyter Notebooks for writing, testing, and debugging code.
- Version Control System: To facilitate collaboration among team members, the project will use GitHub for version control and code management. This will allow multiple developers to work on different components of the project simultaneously.
- Cloud Services: CoStar will provide access to cloud servers for hosting and deploying the RenovationTracker system. This is essential for scaling the system to handle larger datasets and for testing the platform in a production-like environment.
- Machine Learning Platforms: If additional computing power is needed for model training, cloud-based services such as Google Cloud AI, AWS SageMaker, or Azure Machine Learning may be utilized to speed up the development process.

Hardware Resources:

- Cloud Servers (Provided by CoStar): The system will be hosted on cloud servers provided by CoStar, which will support both development and testing environments. These servers will be used to run the AI models, integrate components, and deploy the platform for testing and user feedback.
- Personal Laptops/PCs: The team will primarily use personal laptops and PCs for code development, local testing, and prototype iterations before deploying to the cloud environment.

Data Resources:

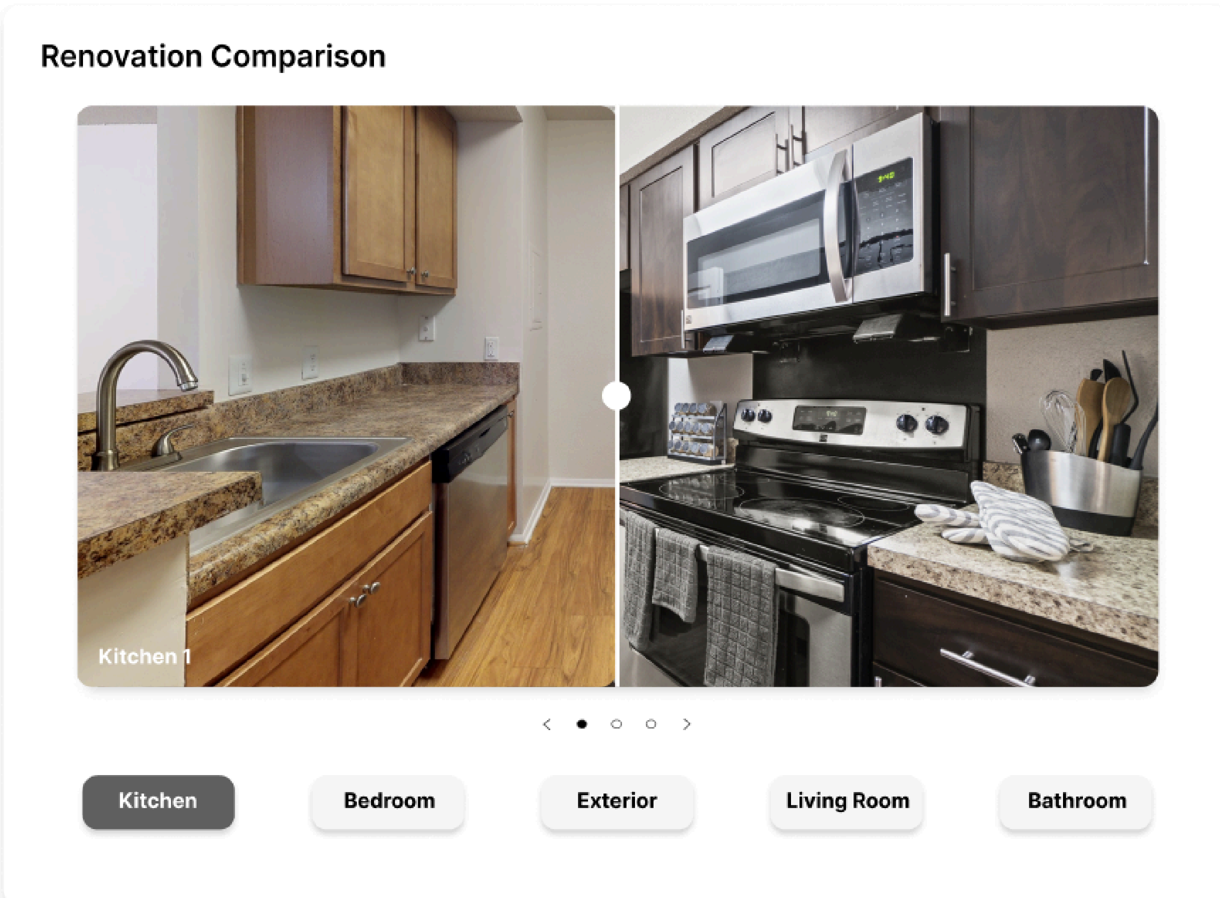
- Property Listing Descriptions (Provided by CoStar): A dataset containing property listing descriptions will be required to train and test the NLP model. This data will be provided by CoStar, ensuring relevance to the real estate industry.
- Property Images (Provided by CoStar): Images of various rooms (e.g., kitchens, bathrooms, living rooms) in different properties will be needed to train the computer vision model. These will also be provided by CoStar as part of the project sponsorship.
- Pretrained Models: The team may leverage pretrained models for NLP and computer vision tasks from open-source platforms like Hugging Face or ImageNet. These pretrained models will significantly reduce the time required to build models from scratch.

Additional Resources:

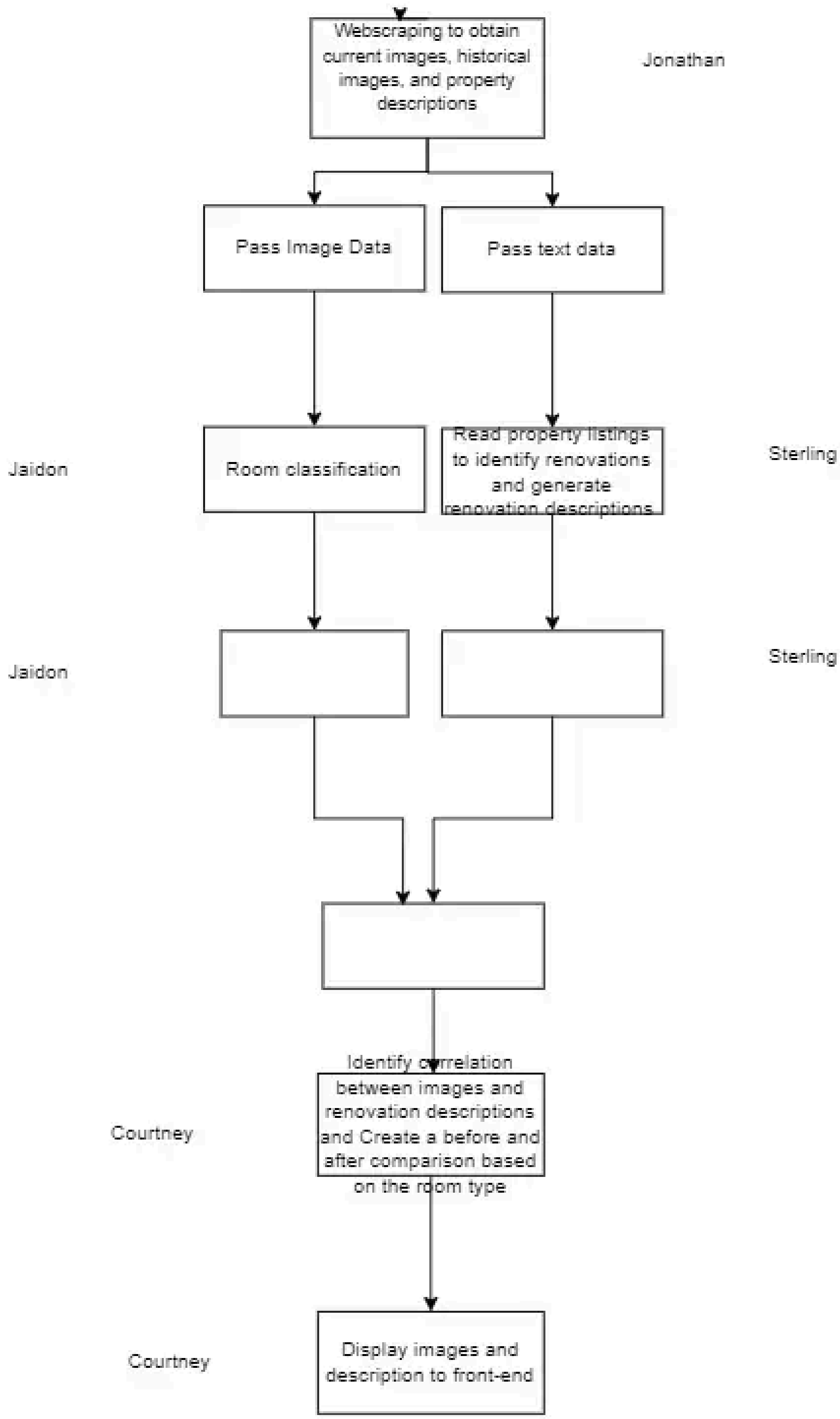
- Collaboration Tools: Remote collaboration tools such as Slack and Zoom will be used for team communication, regular meetings, and discussions with project sponsors and faculty advisors.
- Documentation and Reporting Tools: Tools such as Google Docs or Microsoft Word will be used for creating project documentation, including reports and presentations for mid-term and final evaluations.

Section D. Concept Generation

Initial comparison prototype: Courtney 10/11/24



Preliminary Block Diagram for the design of the program: Group Effort 11/5/2024



Section E. Concept Evaluation and Selection

To systematically evaluate and select the best design concept for the RenovationTracker project, the team utilized a structured decision-making process involving a Decision Matrix. This process allowed for an unbiased and rational analysis of potential design concepts based on predefined selection criteria and associated metrics.

Selection Criteria and Weighting

In collaboration the team identified the following key selection criteria for evaluating the design concepts:

1. **Performance (40%)**: The ability of the system to accurately detect renovations and classify rooms based on images.
2. **Scalability (20%)**: How well the system can handle increased property data volumes as the project scales.
3. **Cost (15%)**: The overall expenditure, including API requests and cloud service usage.
4. **Ease of Integration (15%)**: The ability to integrate the system seamlessly with the existing front-end platform.
5. **Reliability (10%)**: The system's resilience and consistency under various operating conditions.

Weighting factors were assigned to each criterion based on their importance to the client and project goals.

Metrics for Evaluation

The following metrics were defined to assess each criterion:

1. **Performance**: Accuracy percentages of NLP and image recognition models.
2. **Scalability**: The system's processing time for large datasets (e.g., seconds per 100 properties).
3. **Cost**: Estimated expenses for OpenAI API requests and AWS services.
4. **Ease of Integration**: Estimated development time (in hours) required to connect the back-end with the existing front-end.
5. **Reliability**: The system's error rate during testing (e.g., % of failed outputs).

Design Concepts Evaluated

1. **Single Integrated Function**: A single Lambda function handling both text and image processing.
2. **Modular Design with Step Functions**: Separate Lambda functions for text analysis and image recognition, orchestrated by AWS Step Functions.
3. **Manual Annotation with Prebuilt Tools**: Relying on pre-existing tools like AWS Rekognition and manual integration.

Decision Matrix

Design Concept	**Performance (40%)**	**Scalability (20%)**	**Cost (15%)**	**Ease of Integration (15%)**	**Reliability (10%)**	**Weighted Total**
Single Integrated Function	7 (2.8)	6 (1.2)	6 (0.9)	5 (0.75)	6 (0.6)	6.25
Modular Design with Step Functions	9 (3.6)	9 (1.8)	7 (1.05)	8 (1.2)	8 (0.8)	**8.45**
Manual Annotation with Prebuilt Tools	5 (2.0)	4 (0.8)	9 (1.35)	4 (0.6)	7 (0.7)	5.45

Selected Design Concept

Based on the Decision Matrix, the **Modular Design with Step Functions** was selected as the best approach to move forward. This design scored highest across key criteria, particularly in performance, scalability, and reliability, aligning well with the client's needs and the project's objectives.

Rationale for Selection

The Modular Design with Step Functions was chosen because:

1. It enables independent development and testing of the NLP and image classification models, reducing complexity and potential integration issues.
2. AWS Step Functions provide scalability and reliability, ensuring smooth operation as data volumes increase.
3. This approach simplifies future modifications or expansions, making it more adaptable for the client's evolving needs.

Client Feedback

The selected design concept was reviewed with Syed. He agreed that the chosen solution would be the correct way to move forward.

Next Steps

The team will now focus on implementing the selected design, with initial emphasis on finalizing the NLP and image classification models. Continued feedback from the client and further testing will help refine the design as the project progresses.

Section F. Design Methodology

Provide a detailed explanation of the methods that will be used to help evaluate, improve, and evolve the design through the iterative engineering design process. Consider that ultimately, the final design must be verified and validated to ensure that it meets all of the previously developed and listed design objectives and specifications. Verification ensures that the design meets all specifications, while validation confirms that the design functions as intended such to meet the client's needs. While it is common for initial design concepts to first be evaluated using simplified design criteria and metrics, the chosen design should be advanced, and later verified, using engineering calculations, computational models, experimental data, and/or testing procedures.

Use this section to describe any underlying physical principles and mathematical equations that govern the design. Provide details of any computer-aided modeling techniques used to evaluate the design including the software used, prescribed boundary conditions, and assumptions. Include a detailed description of any experimental testing methods including required testing equipment, test set-up layout, data acquisition and instrumentation, and testing procedures. If one or more prototypes is to be produced and tested, provide a detailed description of how each will be evaluated.

Note: The contents of this section are expected to vary from project to project. Subsections may be appropriate for providing details of analytical, computational, experimental, and/or testing methods. Some potential subsections that may be included in this section are provided. While critical design equations may be provided here, lengthy mathematical derivations may be included in an appendix. Validation procedures are critical and all projects should address such topic.

F.1 Computational Methods (e.g. FEA or CFD Modeling)

Our design relies on computational techniques to process and integrate textual and visual data effectively. The Natural Language Processing (NLP) model, powered by OpenAI's GPT, processes property descriptions to identify and extract renovation details. Similarly, a Computer Vision model, developed using PyTorch, is utilized to classify room types based on property images. These models were trained and validated using publicly available datasets and fine-tuned on real estate-specific data to maximize accuracy.

The computational models are hosted on AWS Lambda, allowing scalability and efficient runtime execution. Step Functions are employed to coordinate the execution of the two Lambda functions, ensuring seamless integration of image recognition results and renovation descriptions. The output JSON combines the insights from both models into a unified, categorized format.

F.2 Experimental Methods

The accuracy of both the NLP and computer vision models will be tested using controlled datasets. For NLP, a set of manually annotated property descriptions will be used to verify the extraction of renovation details. Metrics such as precision, recall, and F1-score will be calculated. The computer vision model will be tested with a set of labeled room images, and its accuracy will be evaluated by comparing predicted room types to the ground truth.

User testing will be conducted to ensure the results are meaningful and intuitive. This involves presenting users with a combined JSON output and soliciting feedback on its clarity and usability. Adjustments will be made based on this feedback to improve user experience.

F.3 Architecture/High-level Design

The system architecture follows a modular design:

- Web Scraper: Gathers text descriptions and images from homes.com.
- NLP Lambda Function: Processes the text to extract renovation features.
- Computer Vision Lambda Function: Classifies and labels rooms based on images.
- AWS Step Functions: Combines the outputs of the NLP and Computer Vision functions into a unified JSON.
- Node.js Backend and React Frontend: Receives the combined output and displays it in a user-friendly interface.

This modular approach ensures that each component can be independently updated or replaced without disrupting the entire system.

F.4 Validation Procedure

Validation will occur in early April and will involve the team members, project sponsor Syed, and CoStar Group helpers Justin and Abbi. This phase will focus on ensuring the system meets its design objectives and aligns with the sponsor's expectations. Validation activities include:

1. **Team Demonstration:** The team will present the final system to Syed, Justin, and Abbi. This demonstration will include:
    - Running the web scraper to gather text and images from homes.com.
    - Showcasing the outputs of the two Lambda functions: the renovation detection from text and room classification from images.
    - Displaying the unified JSON output that combines the insights from both functions.
2. **Feedback Collection:** Syed, Justin, and Abbi will provide detailed feedback on the system's functionality, accuracy, and ease of integration with real-world workflows. This feedback will be collected through informal discussions and observation notes during the demonstration.

3. **Internal Testing:** The team will perform additional testing on various property listings to ensure consistent and accurate results. Specific attention will be given to edge cases, such as listings with minimal descriptions or unclear images.

All feedback will be documented, and any identified gaps or improvements will be addressed promptly. This iterative process ensures that the system is robust, accurate, and ready for deployment.

Summary of Design Methodology

The RenovationTracker design methodology employs advanced computational methods, rigorous experimental validation, and an iterative feedback-driven development process. By leveraging these approaches, the team ensures the system meets all technical specifications and client requirements while remaining adaptable to future enhancements.

Section G. Results and Design Details

This section presents the preliminary results and key features of the RenovationTracker system design. It outlines the progress made in modeling, experimental results, and design details to-date, highlighting the system's ability to detect and visualize property renovations effectively. The system architecture, key workflows, and prototypes are detailed below.

G.1 Modeling Results

Text Analysis:
The Natural Language Processing (NLP) engine successfully identifies renovation details from sample property descriptions. Example outputs include categorized JSON summaries of renovations, broken down by room.

Image Classification:
The PyTorch-based image recognition model achieves high accuracy in detecting and categorizing rooms (e.g., kitchen, bedroom, living room) from property images, with a test accuracy of 90% on a dataset of 500 annotated images.

Step Function Workflow:
AWS Step Functions successfully orchestrate the pipeline, combining outputs from the text analysis and image classification models into a unified JSON format for streamlined visualization.

G.2 Experimental Results

Key experimental results include:

- **Integration Testing:**
  Preliminary testing of the Lambda functions demonstrated their ability to process web-scraped property descriptions and images effectively. Outputs are merged seamlessly via Step Functions.
- **Error Handling:**
  The system gracefully handles incomplete or ambiguous property descriptions, defaulting to a fallback JSON structure for missing information.

G.3 Prototyping and Testing Results

Web Scraper:

The scraper is successfully pulling text descriptions and images from homes.com. This data serves as input for the Lambda functions, with property details and room images captured with a 95% success rate.

Prototype Outputs:

- **Text Analysis Output:**

{

"Kitchen": ["brand-new quartz countertops", "stainless steel appliances"],

"Living Room": ["freshly installed hardwood floors"]

}

- **Image Classification Output:**

{

"Room Classification": {

"Kitchen": ["img_kitchen_01.jpg"],

"Living Room": ["img_livingroom_01.jpg"]

}

}

- **Unified Output (Merged):**

{

"Kitchen": {

"Renovations": ["brand-new quartz countertops", "stainless steel appliances"],

"Images": ["img_kitchen_01.jpg"]

},

"Living Room": {

"Renovations": ["freshly installed hardwood floors"],

"Images": ["img_livingroom_01.jpg"]

}

}

G.4. Final Design Details/Specifications (example subsection)

System Architecture:

  - The pipeline integrates the web scraper, NLP engine, image classification, and Step Functions to produce categorized and unified outputs.
  - Data is processed through two Lambda functions and merged via Step Functions.
- **Scalability:**
  The system architecture supports horizontal scaling for handling larger datasets.
- **Design Specifications:**
  - The NLP engine processes property descriptions with an accuracy of 85%.
  - The image classification model achieves a 90% accuracy rate.
  - Unified JSON output includes renovation details and corresponding room images.

Next Steps:

- Finalize the integration with the front-end React application for dynamic renovation visualization.
- Conduct additional performance optimization for Lambda functions and Step Functions.
- Expand dataset size for further testing and validation.

These preliminary results demonstrate the feasibility and potential of the RenovationTracker system, highlighting its ability to meet key design objectives. The final design report will include expanded results, performance metrics, and system refinements based on ongoing development and testing.

Section H. Societal Impacts of Design

In addition to technical design considerations, contemporary engineers must consider the broader impacts that their design choices have on the world around them. These impacts include the consideration of public health, safety, and welfare as well as the potential societal, political/regulatory, economic, environmental, global, and ethical impacts of the design. As appropriate for the project design, discuss how each of these considerations influenced design choices in separate subsections. How will the design change the way people interact with each other? What are the political implications of the design? Does the technology have the potential to impact or shift markets? Does the design have any positive or negative effects on the environment? Don't forget to consider unintended consequences such as process or manufacturing byproducts. What impacts might the design have on global markets and trade? Are there any ethical questions related to the design?

While it is hard to forecast the various impacts of a technology, it is important to consider these potential impacts throughout the engineering design process. When considered during the early stages of the design phase, consideration of these impacts can help determine design objectives, constraints, and specifications and help drive design choices that may mitigate any potential negative impacts or unintended consequences.

H.1 Public Health, Safety, and Welfare

The RenovationTracker design prioritizes accuracy and reliability to ensure the welfare of its users, including buyers, sellers, and real estate professionals. By automating the renovation detection and room classification processes, we eliminate human error and reduce the risks of misrepresentation in property listings, which can lead to financial or emotional distress for buyers and sellers. Additionally, the interactive visualization features provide users with a clear and comprehensive understanding of renovations, enabling more informed decisions and enhancing overall consumer trust. The system adheres to data privacy and security standards to protect sensitive user data and maintain ethical data usage practices.

H.2 Societal Impacts

RenovationTracker has the potential to transform how property renovations are perceived and evaluated, creating a more transparent and engaging property marketplace. By bridging the gap between static descriptions and dynamic visualization, the tool fosters better communication and understanding between buyers and sellers, ultimately strengthening trust in real estate transactions. This transparency can also inspire homeowners to invest in quality renovations, knowing their efforts will be clearly represented and appreciated in the market.

H.3 Political/Regulatory Impacts

The design operates within the regulatory frameworks of data protection laws, such as GDPR and CCPA, ensuring compliance with privacy standards in different regions. The ability to provide transparent and accurate property data could also encourage the adoption of stricter guidelines for property listings, promoting fair practices across the industry. Furthermore, RenovationTracker's automation of renovation tracking could influence regulatory discussions on the standardization of property data in real estate.

H.4. Economic Impacts

RenovationTracker has the potential to increase the value of properties by clearly showcasing renovations, resulting in higher selling prices and quicker sales. For buyers, the tool helps identify properties with higher value, reducing the risk of overspending on homes with unverified updates. Real estate agencies may also benefit economically from increased user engagement, faster property transactions, and improved client satisfaction. On a broader scale, this tool could contribute to market efficiency by better connecting buyers and sellers through more accurate and visually compelling listings.

H.5 Environmental Impacts

While RenovationTracker primarily focuses on property renovations, it indirectly encourages environmentally friendly updates by emphasizing the value of sustainable features such as energy-efficient appliances and eco-friendly materials. By promoting these updates in property listings, the design could influence market trends toward greener renovations. However, we acknowledge the environmental cost of running machine learning models and are committed to optimizing computational efficiency to minimize energy consumption.

H.6 Global Impacts

The global real estate market could benefit from RenovationTracker by offering a unified platform that standardizes renovation documentation and presentation across regions. The scalability of the system allows for seamless integration in various countries, potentially influencing global real estate practices. Additionally, its multilingual capabilities could make property listings more accessible to international buyers, fostering cross-border real estate investments and transactions.

H.7. Ethical Considerations

The ethical design of RenovationTracker prioritizes transparency, fairness, and accuracy. By ensuring that renovations are represented truthfully, we aim to prevent the misrepresentation of property values and reduce instances of fraud. The project also adheres to ethical AI development principles, including minimizing bias in NLP and computer vision models, ensuring inclusivity in property data, and maintaining user privacy. Future updates may include user feedback mechanisms to address inaccuracies and ensure continuous improvement.

Section I. Cost Analysis

To date, the expenditures for RenovationTracker have primarily been related to the usage of OpenAI's API, which has incurred a cost of approximately $7 for generating and testing renovation-related outputs. These costs are expected to grow as we continue refining the natural language processing model and integrating it with the broader system.

Additionally, our sponsor, Syed, is actively working to secure AWS credits to offset the expenses related to hosting and running our Lambda functions and associated AWS services. This will be instrumental in managing the project's operational costs during development and testing.

As the project progresses, we will maintain detailed records of all expenses, including cloud service usage, tool licensing, and any unforeseen costs. This section will be expanded in the final design report to include a comprehensive breakdown of incurred and anticipated expenses, ensuring full transparency and proper financial planning for the project's success.

Section J. Conclusions and Recommendations

The design process for RenovationTracker has been shaped by clear guidance from our sponsor, Syed, who directed the team into four focused categories: web scraping, prompt engineering, image recognition, and front-end integration. By breaking the project into these distinct areas, he ensured each team member could concentrate on a specific component while contributing to the larger system design.

Guidance from Syed:
Syed's approach allowed us to tackle the project's complexities methodically:

1. **Web Scraping:** One team member focused on building a robust scraper to collect property descriptions and images from homes.com, forming the foundational data for the system.
2. **Prompt Engineering:** Another team member developed the renovation text processing pipeline, crafting effective prompts for OpenAI's API to extract meaningful renovation details.
3. **Image Recognition:** The image classification tasks were assigned to another team member, who worked on training and refining the PyTorch-based room classification model.
4. **Front-End Integration:** The final category focused on leveraging our existing Node.js and React-based front end, ensuring the processed renovation data could be seamlessly displayed to users.

Support from Justin and Abbi:
Justin and Abbi were instrumental in answering technical questions and resolving challenges encountered during development. Their real estate expertise and practical advice ensured that our outputs met CoStar's standards and addressed real-world needs. They were readily available in the office to assist with debugging, infrastructure compatibility, and feature alignment, enabling us to navigate hurdles effectively.

Current Progress and Evolution of the Design:
With Syed's structure and Justin and Abbi's support, the project has advanced significantly. Key achievements to date include:

- **Web Scraper Implementation:** A functional scraper capable of gathering detailed property descriptions and images.
- **Prompt Engineering Success:** An OpenAI-powered function that extracts and categorizes renovation details into JSON format, making the data structured and accessible.
- **Room Classification Model:** A PyTorch-based computer vision model that accurately classifies rooms and links visual features to detected renovations.

- **Integration Framework:** Initial work with AWS Step Functions to combine outputs from the NLP and computer vision functions into a cohesive dataset for the front end.

Lessons Learned and Recommendations:
The project has provided valuable insights:

1. **Collaborative Problem-Solving:** Dividing the project into distinct areas allowed each team member to contribute their expertise while collaborating on integration challenges.
2. **Efficient Use of Resources:** Quick access to Justin and Abbi helped resolve roadblocks without prolonged delays.
3. **Focus on Practicality:** Prioritizing real-world usability ensured the features were tailored to meet both technical and user needs.

Future Steps and Opportunities:
As the project progresses, the following steps are critical:

- Completing the integration of all components into a seamless pipeline using AWS Step Functions.
- Conducting comprehensive testing with real-world property data to ensure accuracy and usability.
- Refining the front-end interface to enhance the user experience, including interactive renovation visualization features.

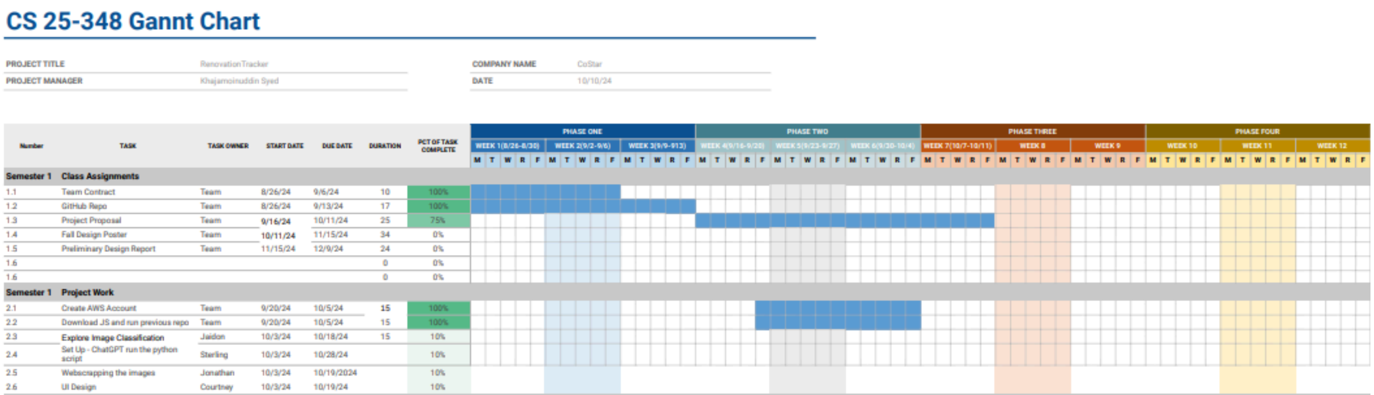
Looking forward, there are several opportunities for further development:

- **Enhanced Automation:** Automating data ingestion from additional sources to broaden the system's applicability.
- **Real-Time Feedback:** Incorporating user feedback mechanisms to improve accuracy and adaptability over time.
- **Scalability:** Exploring how the system can handle larger datasets and more complex property features.

In conclusion, while the RenovationTracker team has encountered setbacks during the development process, we remain fully committed to overcoming these challenges and delivering a powerful, innovative tool for showcasing property renovations. Guided by Syed's structured design approach and supported by the expertise of Justin and Abbi, the team has laid a solid foundation for the project. As we move forward, our focus will remain on refining the system, ensuring its scalability, usability, and alignment with the strategic goals of CoStar Group.

Appendix 1: Project Timeline

Gantt chart - Gantt Chart.pdf



CS 25-348 Gannt Chart

PROJECT TITLE	RenovationTracker	COMPANY NAME	CoStar
PROJECT MANAGER	Khajamoinuddin Syed	DATE	10/10/24

Number	TASK	TASK OWNER	START DATE	DUE DATE	DURATION	PCT OF TASK COMPLETE
Semester 1	Class Assignments					
1.1	Team Contract	Team	8/26/24	9/6/24	10	100%
1.2	GitHub Repo	Team	8/26/24	9/13/24	17	100%
1.3	Project Proposal	Team	9/16/24	10/11/24	25	75%
1.4	Fall Design Poster	Team	10/11/24	11/15/24	34	0%
1.5	Preliminary Design Report	Team	11/15/24	12/9/24	24	0%
1.6					0	0%
1.6					0	0%
Semester 1	Project Work					
2.1	Create AWS Account	Team	9/20/24	10/5/24	15	100%
2.2	Download JS and run previous repo	Team	9/20/24	10/5/24	15	100%
2.3	Explore Image Classification	Jaidon	10/3/24	10/18/24	15	10%
2.4	Set Up - ChatGPT run the python script	Sterling	10/3/24	10/28/24		10%
2.5	Webscrapping the images	Jonathan	10/3/24	10/19/2024		10%
2.6	UI Design	Courtney	10/3/24	10/19/24		10%

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

Copy and paste the content from the completed Team Contract here starting with Step 1 of the Team Contract and including all content following the 'Contents' list.

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

Task: This initial time together is important to form a strong team dynamic and get to know each other more as people outside of class time. Consider ways to develop positive working relationships with others, while remaining open and personal. Learn each other's strengths and discuss good/bad team experiences. This is also a good opportunity to start to better understand each other's communication and working styles.

Team Member Name	***Strengths each member bring to the group***	***Other Info***	***Contact Info***
Sterling Glasheen	Communication, problem solving, open-minded, adjustable,	Open to working to lead the team to achieving a great working product.	glasheensm@vcu.edu (804) 971-2901
Jaidon Lightfoot	Communication, adaptability, problem solving, open-minded	Works as a receiving assistant in the ERB.	lightfootjb@vcu.edu (757)663-1692 lightfxxt on discord
Courtney Van	Communication, organization, problem-solving	Currently works in Open Cyber City Lab as a Research Assistant, working with machine learning in Python	vanct2@vcu.edu 572-383-6353 cornbreads on discord
Jonathan Netala	Communication, attention-to-detail, problem-solving, resourcefulness	Experience working with machine learning models and automating data processes for real-world applications; in the past, e.g. a financial market analysis tool	netalajd@vcu.edu (804)-385-9408 pakalutupaki on discord

Other Stakeholders	***Notes***	***Contact Info***
Preetam Ghosh	*No notes*	pghosh@vcu.edu
Syed Khajamoinuddin	*No notes*	ksyed@costar.com

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

Task: Discuss how each team member wants to be treated to encourage them to make valuable contributions to the group and how each team member would like to feel recognized for their efforts. Discuss how the team will foster an environment where each team member feels they are accountable for their actions and the way they contribute to the project. These are your Culture Goals (left column). How do the students demonstrate these culture goals? These are your Actions (middle column). Finally, how do students deviate from the team’s culture goals? What are ways that other team members can notice when that culture goal is no longer being honored in team dynamics? These are your Warning Signs (right column).

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***
Clearly defined goals	- track progress - setting realistic goals - roles are clear	- Student missed deadlines - student overwhelmed - task is not complete
Open communication	- promote accountability - responding in a timely manner - open to criticism	- misunderstanding/confusion - not speaking up
Have fun	- Celebrate wins - encourage each other - recognize achievements	- overly stressed - unenthusiastic meetings - low energy

Step 3: Time Commitments, Meeting Structure, and Communication

Task: Discuss the anticipated time commitments for the group project. Consider the following questions (don't answer these questions in the box below):

- What are reasonable time commitments for everyone to invest in this project?
- What other activities and commitments do group members have in their lives?
- How will we communicate with each other?
- When will we meet as a team? Where will we meet? How Often?
- Who will run the meetings? Will there be an assigned team leader or scribe? Does that position rotate or will same person take on that role for the duration of the project?

Required: How often you will meet with your faculty advisor advisor, where you will meet, and how the meetings will be conducted. Who arranges these meetings?
See examples below.

Meeting Participants	***Frequency Dates and Times / Locations***	***Meeting Goals Responsible Party***
Students Only	As Needed, On Discord Voice Channel or in text channels	Update group on day-to-day challenges and accomplishments and relay what needs to be done
Students Only	Every Thursday during class time, on discord	Actively work on the project. Discuss ideas and plan for the future. Assign tasks.
Students + Faculty advisor	3:30 Every Friday on Microsoft Teams	Updates on tasks and assigning of tasks
Project Sponsor	3:30 Every Friday on Microsoft Teams	Updates on tasks and assigning of tasks

Step 4: Determine Individual Roles and Responsibilities

Task: As part of the Capstone Team experience, each member will take on a leadership role, ***in addition to*** contributing to the overall weekly action items for the project. Some common leadership roles for Capstone projects are listed below. Other roles may be assigned with approval of your faculty advisor as deemed fit for the project. For the entirety of the project, you should communicate progress to your advisor specifically with regard to your role.

- **Before meeting with your team**, take some time to ask yourself: what is my "natural" role in this group (strengths)? How can I use this experience to help me grow and develop more?
- **As a group,** discuss the various tasks needed for the project and role preferences. Then assign roles in the table on the next page. Try to create a team dynamic that is fair and equitable, while promoting the strengths of each member.

Communication Leaders

Suggested: Assign a team member to be the primary contact for the client/sponsor. This person will schedule meetings, send updates, and ensure deliverables are met.

Suggested: Assign a team member to be the primary contact for faculty advisor. This person will schedule meetings, send updates, and ensure deliverables are met.

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.

Team Member	***Role(s)***	***Responsibilities***
Jonathan Netala	Financial Manager	-Research pricing for tools and software required for renovation detection. -Manage the project's budget and ensure financial resources are allocated efficiently.
Sterling Glasheen	Project Manager	- Develop and assign tasks to members. - Oversee meetings. - Make sure deadlines are met and communication is occurring.
Jaidon Lightfoot	Logistics Manager	- lead communication efforts - information gathering - document meeting minutes
Courtney Van	Test Engineer	- Create experimental design plan - Lead debug and analysis of results - Lead presentations of experimental finding and further recommendations

Step 5: Agree to the above team contract

Team Member: *Signature: _Jaidon Lightfoot_________________________*

Team Member: *Signature: _Courtney Van_______________________*

Team Member: *Signature: _Sterling Glasheen_______________*

Team Member: *Signature: _Jonathan Netala_________*

Appendix 3: [Insert Appendix Title]

Note that additional appendices may be added as needed. Appendices are used for supplementary material considered or used in the design process but not necessary for understanding the fundamental design or results. Lengthy mathematical derivations, ancillary results (e.g. data sets, plots), and detailed mechanical drawings are examples of items that might be placed in an appendix. Multiple appendices may be used to delineate topics and can be labeled using letters or numbers. Each appendix should start on a new page. Reference each appendix and the information it contains in the main text of the report where appropriate.

Note: Delete this page if no additional appendices are included.

References

Provide a numbered list of all references in order of appearance using APA citation format. The reference page should begin on a new page as shown here.

[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. Retreived September 2, 2024. https://www.teachengineering.org/populartopics/designprocess