March 21, 2025

C964: Computer Science Capstone Template

Warning: Though it is not stated in the official resources, evaluators do not like outlines. Write narratively using paragraphs with complete sentences. Use these C964 examples to see what evaluators typically expect.

Task 2 parts A, B, C and D

Part A: Letter of Transmittal

Letter of Transmittal Requirements

The *Letter of Transmittal* should convince senior leadership to approve your project. Write a brief cover letter (suggested length 1-2 pages) describing the problem, how the application (part C) applies to the problem, the practical benefits to the organization, and a brief implementation plan. Include all artifacts typical of a professional (business) letter, e.g., subject line, date, greeting, signature, etc.

The letter should be concise and target a non-technical audience. Include the following:

- A summary of the problem.
- A proposed solution centering around your application.
- How the proposed solution benefits the organization.
- A summary of the costs, timeline, data, and any ethical concerns (if relevant).
- Your relevant expertise.

09/30/2025

Mazi Arjomand

Wag!

55 Francisco St, Suite 360, San Francisco, California 94133

Dear Mazi,

After using your wonderful service for many years, and working here for a sweet few, my great admiration and respect for the app has led me to take on a personal responsibility to help the service grow. I have recommended the app to countless friends and family through the years, with most of them absolutely loving it! There are, however, just a few instances I have observed where some of the people trying the app for the first time are hesitant and borderline paranoid about a few things. For my close friends Alisha, Brandon, Jermain, and Alex, who all have multiple dogs, their main concern was that an employee would come by and take the wrong dog outside for a walk. While I thoroughly reassured them all that the employees were very well trained and paid close attention to each client's profile and request for every service they complete, and they did all eventually did become repeat users, it became apparent to me that there could be an additional step in each service that could provide the peace of mind that potential clients who do not have a close friend with personal experience seek. When Wag! first started in 2014 technology was acclimating to robotic assistants and automation. Siri was only four years old, Alexa had not even been launched, and there were no AI or LLM buzzwords used to show off the newest

technological craze. Nowadays automation and machine learning have exploded to the point where if a brand new product or service fails to mention ‘AI’ in their pitch deck investors seemingly are completely disinterested and existing brands that fail to implement AI into their stack are projected to fall to the wayside to make room for the competitors that do. I deeply love and appreciate the service that you and every other hardworking person at the company deliver to the public and it would not only incredibly inconvenience myself and every other subscriber but it would also personally hurt to see it fall short in any way.

That is why I am proposing to implement an AI-powered Breed Identification program to be implemented as part of a walker’s workflow. After completing their walk with the dog or dogs the client has requested the service for, I propose that they not only check-in and mark the service as complete, but they also take and submit a picture back to us for verification and, if it matches, we send that image to the client. This way we can non-invasively provide that reassurance to potential clients by showing them that the service has been completed correctly while also being able to advertise that we are using AI to automatically double-check the services and correct any human errors that may occur from a busy, tired, or potentially distracted walker.

After planning it out, the creation and implementation of such a feature would be inexpensive, quick, and simple. We can utilize existing, open-sourced computer vision models and train them on freely licensed datasets to create a baseline working AI that can identify dog breeds then continue to train them on the images sent to us from walkers with the correct breed information found in the service request from our clients. For these images, we can train our walkers to not capture any people or sensitive information and down the line we can even train the AI to automatically detect and remove any personal identifiers from images in case someone submits an image that contains them unknowingly. All of the required tools and environments for development are open-sourced and free to use for business while still being incredibly powerful and flexible enough for us to create a purpose-built tool like this.

By my research and familiarity with these types of projects, I estimate that the entire process to create and implement this new feature would only take us two-quarters, one of which will be utilized to run Alpha and Beta tests where we can collect more images from our walkers and further train the model. With our AI program being relatively lightweight and minimal, we can create our own system to host it and scale it further if needed or switch to a hosting service for a marginal amount if it becomes a critical success and we expand rapidly. For the initial system, the development team’s time, and the total cost of the software and environments necessary for development, I am estimating a cost of roughly $150,000 which will be broken down in detail at the end of section B on this proposal.

Although my time with Wag! has been short, I am incredibly personally invested in the company and the brand and would love to help it grow by onboarding all of the potential clients who, like some of my close friends, have had reservations about the quality or accuracy of our services. Providing this extra peace of mind would help us with our marketing efforts, increase the number of newly signed up members, and help us continue to stay innovative with the technological advances we have available to us today.

Sincerely,

Ajmal Bari

Software Engineer I

Part B: Project Proposal Plan

The project proposal should target your client's middle management. This audience may be IT professionals but have limited computer science expertise. Use appropriate industry jargon and sufficient technical details to describe the proposed project and its application. Remember, you're establishing the technical context for your project and how it will be implemented for the client. **Write everything in the future tense.**

Project Summary

- € Describe the problem.
- € Summarize the client and their needs as related to the problem.
- € Provide descriptions of all deliverables. For example, the finished application and a user guide.
- € Provide a summary justifying how the application will benefit the client.

This project aims to provide extra peace of mind for our clients by utilizing AI to confirm that the correct dog has been walked upon service completion. The module created through this project will be integrated into a walker's normal workflow by asking them to take a picture of the dog post-walk. The picture will be sent to the AI which will determine the breed. If the determined breed matches the breed of the dog(s) from the service request, we will let the walker know they are good to go and forward that image to the client so they have a visual confirmation and do not have to worry about the service being done properly.

There are a couple of key components that need to be created for this project that ultimately compose the first phase of full integration. The scope of the project covers creating and training AI that can classify dog breeds, creating a web hosted interface to submit images to the AI for classification, and creating a way to notify the walkers of the AI's classification results.

By providing our walkers with an additional safety net we can mitigate the amount of unsatisfied clients and incorrect service fulfillments. Also, sending our clients a picture of their dog(s) will provide them the extra peace of mind that a lot of potential new clients have been waiting for before signing up. In essence, it will help us market our service more effectively to onboard new clients while reducing the number of issues encountered with our service execution.

Data Summary

- € Provide the source of the raw data, how the data will be collected, or how it will be simulated.
- € Describe how data will be processed and managed throughout the application development life cycle: design, development, maintenance, etc.
- € Justify why the data meets the needs of the project. If relevant, describe how data anomalies, e.g., outliers, incomplete data, etc., will be handled.
- € Address any ethical or legal concerns regarding the data. If there are no concerns, explain why.

The initial data necessary for training the AI is the open-sourced, freely-licensed Stanford Dogs Dataset. It can be obtained and utilized for business and research applications with no cost or contractual obligations. It contains no sensitive data and can be stored and used without any concerns for privacy. In order to use the images from the dataset to train the AI, code must be written to normalize them all to the same standard height and width while also flipping some images at random. The image modifications

allow the AI better process what it is looking for and can be executed automatically upon providing the AI with images. Training from this dataset can yield effective results that will be capable of identifying the breed of a dog using real use-case images with an accuracy of 90%. For any images that contain a mixed breed or cannot confidently be identified as a single breed by the AI, a top three result with confidence scores will be returned to the user.

After the initial AI is created, the goal is to improve its performance through Alpha and Beta testing by utilizing the walker's submitted images along with the breed listed on file for the service to further train the AI. This will make it much more robust in comprehending and accurately identifying the various breeds from real world situations. For these test phases, the AI will be adapted to automatically detect and blur any text and faces found within the scope of the images in order to avoid any potential sensitive data collection. The images stored for further training will only be the ones that have been processed to contain no sensitive information.

For this current phase of the project, both the AI and the initial training dataset will be stored on a local machine built and run locally through the IT department. With the relatively small compute power and storage space necessary for this AI module, it will help save the company in the long-run with data center costs while also providing the company a platform to further develop AI capabilities and enhancements. We will monitor resource utilization on the newly created machine as development and testing continues in order to determine if and when there is a right time to transition to a cloud hosted environment.

Implementation

- € Describe an industry-standard methodology to be used.
- € An outline of the project's implementation plan. The focus can be the project's development or the implementation of the machine learning solution.

We will be using the Agile method to efficiently conduct this project and allow room to improve upon it during development. We will be including our Software Engineering, IT, customer service, and UI/UX development teams as stakeholders to collaborate with the project developers and provide valuable insight and feedback as the project moves along. This will allow us to iterate on the core features and functionality so the final product can be as polished as possible.

In order to implement the machine learning algorithm we will be applying the SEMMA method with an Agile framework. The steps involved in the SEMMA process will be broken into different stories and sprints to help with organizing the data cleaning and training processes while keeping everyone on the same page with regards to timeline and progress. In short, the SEMMA process includes obtaining our data samples (Stanford Dogs Dataset), exploring the data to identify impactful variables, modifying the data to standardize and clean it before passing it into the model, creating the model on the modified data, then assessing the capabilities of the model before moving on to the next steps in the projects. Some phases take more time and effort, it is not a linear progression through all of the steps, but with the help of Agile planning our team will be able to accomplish each step while keeping everyone updated and aware of any updates, roadblocks, and successes along the way.

Timeline

- € Provide a projected timeline. Include each milestone and deliverable, its dependencies, resources, start and end dates, and duration. (a table is not required but encouraged).
- € Dates should be in the future. Write 'NA' where an item is not applicable.

Milestone or deliverable	**Project Dependencies**	**Resources**	**Start and End Date**	**Duration**
Prepared Data - Obtain, standardize, and organize the Stanford Dogs dataset into the three categories; training, validation, and test	Acquisition of a suitable dataset (Stanford Dogs Dataset)	Python, PyTorch, local development environment, and GIMP	11/03/2025 - 11/07/2025	One week
Local AI Development Computer - assemble a computer powerful enough to develop and host the AI model	Acquisition of two Nvidia A100 GPUs, one AMD 9995WX CPU, 256GB of ECC Memory, workstation grade motherboard, two 1500 Watt PSUs	IT Department labor for assembly, initial imaging, and stability testing of the machine	11/03/2025 - 11/07/2025	One week
Initial Trained Model - Utilize Pytorch and ResNet50 to conduct initial training of the model so it can begin breed identification	A prepared dataset, CUDA enabled computer, and a local dev environment	Python, PyTorch, local development environment with a suitable CUDA enabled GPU	11/10/2025 - 11/14/2025	Two days
Well-Tuned and optimized model - Analyze initial training results to identify key variables that affect the identification outcome so the model can be tuned around them in order to improve accuracy of the model	An initially trained model that generates detailed reports for analysis of results	Generated reports from the initial model	11/17/2025 - 12/05/2025	Two weeks
Alpha Testing - Internal testing of the AI model that incorporates collection of images	The well-tuned model and employees that will be willing to	Developers to monitor and analyze results from the new	12/08/2025 - 12/19/2025	Two weeks

from volunteering employees to further test, train, and tune the model	volunteer and submit images of their dogs to the development team	images processed by the AI and implement further tuning as required as well as internal feedback from volunteers		
Integration and UI Development - Develop a simple web interface that can hook into the existing app for walkers to easily submit their images to the AI and get a result returned to them	Finely tuned, stable, and well performing model	Web developers, UI/UX designers, and backend engineers	12/22/2025 - 01/09/2026	Three weeks (additional time considered for the holiday season)
Beta Testing - A limited beta test that a limited number of long-standing users can select whether or not to opt-in to in order for us to further train the model from more accurate real-world experience	Model that is tuned after alpha testing results have been analyzed and users that will be willing to opt-in to the beta test program	Images submitted from walkers with clients who have opted-in to the beta test, QA team to ensure the system is running as intended and improving with each tune	01/12/2026 - 02/06/2026	Four weeks
Final Evaluation and Documentation - Create the documentation necessary to explain and support the model for future users and developers. Work with marketing teams to leverage the new capabilities into advertisements and conversion efforts	Fully functional workflow that implements model inference into the existing app UI	Developers who contributed to the project, IT team, and documentation hosting service (ex: Confluence)	01/26/2026 - 02/20/2026	Four weeks

Evaluation Plan

- € Describe the verification method(s) to be used at each stage of development.
- € Describe the validation method to be used upon completion of the project.

Since we will be using the Agile method, verification and validation will be conducted throughout the project's lifecycle by utilizing feedback from stakeholders, developers, and testers. During the data processing phase we will be implementing scripts to hold all images to the same formatting standards. Once we begin training the model we can analyze the result of each Epoch (a cycle of training for the model) to spot any issues with overfitting our model and ensure it is improving and learning properly from the dataset. The Alpha test will begin our most meticulous training phase, where we can analyze the results of using images outside of the dataset with the model. The images will still be passed through the initial scripts to format them into the correct file extensions and size, but since they are not from the dataset we will obtain our first peek into the real-world performance of the model. After the results of the alpha test have been analyzed and the model is further tuned from those findings, we can begin the Beta test. Walkers with clients who have selected to opt-in to the beta test for the incentive reward will submit their images of the dogs for further analysis of the model's performance while also providing valuable feedback on the UI, the workflow, and the results returned by the model. We will analyze the results of the images to further refine the model while also taking in the walker's feedback to improve the integrated workflow and user experience of the model.

Costs

Include the itemized costs of the project. Include specific item names where applicable, e.g., 'PyCharm Professional Ed. 2024.3.5.'

- € Itemize hardware and software costs.
- € Itemize estimated labor time and costs.
- € Itemize estimated environment costs of the application, e.g., deployment, hosting, maintenance, etc.

Item	Description	Estimated Cost
AI Development Workstation	A powerful computer utilizing dual Nvidia A100 GPUs, AMD Ryzen 9995WX CPU, 256GB ECC Memory, and Workstation grade motherboard with dual power supplies. Including the IT department's time required to assemble, image, and test the machine (16 hours).	$30,000
Software and Development Environment	PyTorch and Python are open-sourced languages which will be used to program the	$0

	model and VS Code is a free IDE	
Dataset for Initial Training	Stanford Dogs Dataset is open sourced	$0
Labor	Hiring two dedicated developers to assist with the creation and implementation of the AI model across the estimated 16 week time period at a rate of $75/hour	$96,000

Part C: Application

Part C is your submitted application. This part of the document can be left blank or used to include a list of any submitted files or links.

The minimal requirements of the submitted *application* are as follows:

1. **The application functions as described.** Following the 'User Guide' in part D, the evaluator must be able to review your application on a Windows 10 machine successfully.
2. **A mathematical algorithm applied to data,** e.g., supervised, unsupervised, or reinforced machine learning method.
3. **A "user interface."** Following the 'User Guide' in part D, the client must be able to use the application to solve the proposed problem (as described in parts A, B, and D). For example, the client can input variables, and the application outputs a prediction.
4. **Three visualizations.** The visualizations can be included separately when including them in the application is not ideal or possible; e.g., the visualizations describe proprietary data, but the application is customer-facing.
5. **Submitted files and links are static and accessible.** All data, source code, and links must be accessible to evaluators on a Windows 10 machine. If parts of the project can be modified after submission, matching source files must be submitted. For example, if the application is a website or hosted notebook, the `.html` or `.ipynb` files must be submitted directly to assessments.

Ideally, submitted applications should be reviewable using either Windows or Mac OS, e.g., Jupyter notebooks, webpages, Python projects, etc. If the source files exceed the 200 MB limit, consider providing screenshots or a Panopto video of the functioning application and contact your course instructor.

Part D: Post-implementation Report

Create a post-implementation as outlined below. Provide sufficient detail so that a reader knowledgeable in computer science but unfamiliar with your project can understand what you have accomplished. Using examples and visualizations (including screenshots) beyond the three required is recommended (but not required). **Write everything in the past tense.**

Solution Summary

- € Summarize the problem and solution.
- € Describe how the application solves the problem from parts A and B.

Data Summary

- € Provide the source of the raw data, how the data was collected, or how it was simulated.
- € Describe how data was processed and managed throughout the application development life cycle: design, development, maintenance, etc.

Machine Learning

For each machine learning model (at least one is required), provide the following:

- € Identify the method and what it does (the "what"). It's advisable to include an example of the model's output.
- € Describe how the method was developed (the "how").
- € Justify the selection and development of the method (the "why").

Validation

For each machine learning algorithm described in the section above, do the following:

- € Identify the model's machine learning category, e.g., supervised, unsupervised, or reinforced. For blended approaches, identify the category most relevant to the model's application.
- € An appropriate validation method for the model's performance.

For supervised learning and reinforced learning

- € Describe an appropriate metric(s) for testing the model's performance.
- € Provide results of testing using the described metric.

For unsupervised learning

- € Describe an appropriate method(s) for testing the model's performance.
- € Provide the results of testing using the above method. The method should provide an example of the model's output and how the output is relevant or a metric measuring the model's performance, e.g., the Rand index or Silhouette Coefficient.

Visualizations

Identify the location of at least three unique visualizations. They can additionally be included here.

User Guide

Include an enumerated (steps 1, 2, 3, etc.) guide to execute and use your application.

- € Include instructions for downloading and installing any necessary software or libraries.
- € Give an example of how the client should use the application.

Reference Page

Include references for cited works, e.g., (Author, year) following an accepted writing style. References are not required; this page can be removed if no references are used. To cite sources used for code, you should include the references as code comments within the source code.

Datascience-PM, (2024, November 16th). *What is SEMMA?* https://www.datascience-pm.com/semma/

Stanford Dogs Dataset