The Title Of Your Project

**October 2025**

**By**

**Your Full Name Here**

**Student number 202XXXXXX**

**Word count: XXXX**

This document is a suggested template for your guidance. You don’t have to stick to it precisely. It may not suit your particular project. Modify it if you need to – but please discuss with your supervisor before making substantial changes to the organisation and content. Make sure you write about the all the required elements mentioned in the assignment specification.

Paragraphs in red, like this one, are instructions and extra information. You must delete them before submitting your report.

This template document has a number of paragraph styles predefined. If you use them (‘Heading 1’, ‘Heading 2’ and ‘Heading 3’) and don’t alter them, then your report will automatically have properly numbered paragraphs and your table of contents will be automatically generated with the right page numbers. Use ‘Normal’ as the style for general text paragraphs in your document. If you don’t know how to use paragraph styles in Word, now is a good time to learn.

On this page (and I *really* hope this is obvious) you must replace the words “The Title Of Your Project”, “Your Full Name Here” and “202XXXXXX” with the correct information. You’d be surprised how often people forget to do this. Don’t be one of them.

You must also replace “XXXX” with the actual word count (excluding acknowledgements, abstract, table of contents, references and appendices) of your document after you have deleted all the red bits.

This table of contents (TOC) is automatically generated by Word, based on paragraphs with styles of ‘Heading 1’, ‘Heading 2’, ‘Heading 3’ and ‘Unnumbered Heading’. It will automatically update with the correct page numbers when you load the document, but you can force it to do so at any time by right-clicking on it (the whole TOC will turn grey) and selecting “Update Field” or pressing F9.

If you manually edit the table of contents, you may find that either your changes will get overwritten or it will no longer update automatically. It’s a pain to keep a table of contents up to date manually, so you are advised not to edit it!

Delete these red paragraphs!

Contents

[1. Project background and purpose 3](#_Toc144892521)

[1.1. Introduction 3](#_Toc144892522)

[1.2. Objectives 3](#_Toc144892523)

[1.3. Scope 3](#_Toc144892524)

[1.4. Deliverables 3](#_Toc144892525)

[1.5. Constraints 3](#_Toc144892526)

[1.6. Assumptions 3](#_Toc144892527)

[2. Project rationale and operation 5](#_Toc144892528)

[2.1. Project benefits 5](#_Toc144892529)

[2.2. Project operation 5](#_Toc144892530)

[2.3. Options 5](#_Toc144892531)

[2.4. Risk analysis 5](#_Toc144892532)

[2.5. Resources required 5](#_Toc144892533)

[3. Project methodology and outcomes 6](#_Toc144892534)

[3.1. Initial project plan 6](#_Toc144892535)

[3.1.1. Tasks and milestones 6](#_Toc144892536)

[3.1.2. Schedule Gantt chart 6](#_Toc144892537)

[3.2. Project control 6](#_Toc144892538)

[3.3. Project evaluation 6](#_Toc144892539)

[4. References 7](#_Toc144892540)

[5. Appendix a 8](#_Toc144892541)

# Project background and purpose

## Introduction

Major and catastrophic events like natural disasters have lots of images and videos uploaded to social media by a lot of people such as civilians, tourists and media companies. For humanitarian and charity organisations including emergency responders, these images and videos can be very useful to spread awareness globally to people. However, when there is so much information and data in terms of velocity and volume, evaluation becomes unreliable and unsupportable which can make important decisions also unreliable. This project aims to use the techniques of Artificial Intelligence to possibly overcome the current techniques, which are very highly unreliable, to make visual data understandable and more reliable to effectively spread awareness.

The title of this project is called “Visualising natural disaster image embeddings” and it will use AI techniques to process large image data from an incident. This work will use a pre-trained Neural Network to use high-dimensional features that are called embeddings. These embeddings will get transformed into two-dimensional space using reduction algorithms like Uniform Manifold Approximation (UMAP). [1]

The results will be an interactive visualisation software with full imagery dataset visualised in a 2D map. This will help the user do quick exploratory analysis, identifying thematic patterns and clusters (damaged infrastructure, major incident locations etc) that might be impossible to discover through human inspection and evaluation. This project definition document provides the background, motivation and rationale for this project and also includes objectives and aims with a detailed methodology and an intensive roadmap and plan for a successful result and evaluation.

## Objectives

The main focus of this project is finding an efficient and better solution to not just the problem but also the current solution used right now. My solution will contain a visualisation using AI techniques to utilise large sets of high-quality images and plot them in a 2D map. This is also an exploration of how Artificial Intelligence could be used to help analyse and evaluate a crisis. Success will be decided on clear goals that have deadlines.

The project starts off with researching and finding suitable images for data. This goal of this part is to research and gather appropriate quality images through exploring the internet or finding open-source data to use as datasets. This will probably require a minimum of 5000 disaster-related images by the end of week 4 of trimester 1 but the number of images required might change depending on the strength of the current hardware. After successfully creating a dataset, we can now escalate this project more technically by creating an effective data processing pipeline to use for the datasets. The next goal is to develop a facility to pre-process all the images in that dataset to produce a set of embeddings using a pre-trained neural network deep learning model using CLIP by the end of week 7. The third goal consists of creating a functional, interactive prototype of a visualisation built using the Streamlit framework. The prototype will contain the core of project’s delivery that presents the 2D UMAP view of the image embeddings and showing useful insights for the user by the end of week 13. To help the project go beyond the basic needs of the main requirements and achieve large technical depth, we need to implement a series of secondary goals. This is because if the main aims are achieved earlier than expected then we can take this a step further by exploring clustering algorithms like HDBSCAN to search and assign thematic groups to the data [2]. Not only this, but we can compare 2 competing pre-trained deep learning models with the goal to analyse their suitability for this project to ensure that this project is supported with evidence-based analysis and conclusions.

## Scope

The project is centred around the workflow of converting a set of disaster-related images into an interactive, functional visualisation. The scope of the project is acquiring a bunch of quality images that can be used to be implemented in a data pipeline to get embeddings and to apply dimension reduction to achieve a 2D interactive projection. Not just that but also develop an operational prototype that is suitable for data exploration. The final aspect of this project includes an overall analysis and conclusion of the effectiveness of the prototype as well as the comprehensive report.

In order to keep the project clean and reliable that is within the time constraints, a number of closely related areas are excluded from the scope. The project doesn’t entail the processing of streaming or real time data from the internet but actually uses existing open sourced datasets of images only. The report does not include the information from the internet like social media posts that have images but is restricted to the visual information in the images. Furthermore, the visualisation does retain the abstract 2D similarity of the images. It also does not contain the locations in the world. At the end, the output is a prototype to be tested and proved and should not be assumed that it is industry-ready global scalable software package.

## Deliverables

The deliverables of this project will also be evidenced by the presentation of the portfolio and its major results. They mark the quality of this research project not just the development phase but also the evaluation and understanding of the activity that were carried out in this project. The significant deliverables are an interactive visualisation program written in a Python based web prototype with streamlit and plotly libraries. The system will showcase a 2D projection of the image data set which allows the user to interact and explore with it.

Another deliverable is a full repository containing the source code and heavy documentation in GitHub with all the core functionality like data preprocessing, embedding, dimensionality reduction and the visualisation software.

Data Sets are also important because transferring them into high dimensional embeddings and then extracting coordinates which are then stored in a compact HDF5 file format to aid downstream analysis.

The final report is another significant deliverable that contains a detailed and professional report that has the rationale, planning and delivery, methodology, analysis and evaluation as well as a reflective critique.

Finally, showcasing the evidence of everything in my GitHub repository by presenting them to my Supervisor Xinhui Ma of the software either with or without a video walkthrough recording that is outlined by the assessment criteria as well as answering any given question and follow ups.

## Constraints

This project gives lots of freedom of technical use but there are some significant externally imposed constraints. The strongest constraint here is the final project deadline which is in Trimester 2 which defines the project’s overall timescale and progression. The project will need to strictly stay within the academic rules of the honours Stage project module and guidelines as well as the Computer Science AI programme competency that necessitates the application of innovative AI techniques to address a real problem. All of the development will also need to be in the bounds of the computing resources that will be available and will need to observe the data protection (GDPR) and copyright law of any data sets that will be used here. Technology selection is not considered to be an external constraint but a personal choice.

D

## Assumptions

To make progress with planning and executing my project, my project has a series of real assumptions. To begin with, I suppose that I can identify and gather a good quality of open source set of a minimum thousand images related to disaster that can be licensed with permission to be used for research purposes. We can defend this assumption via preliminary research of literature that appears to hold out the promise of their existence. Secondly, we suppose that some images in the data set might be considered either noise, unreliable, corrupt or even duplicates that could affect the outcome. Lastly, our project supposes that our requisite open source software libraries (e.g. PyTorch) will be available over the period of this project and our current hardware will be able to meet the sufficient requirements for compute, disk space and RAM speed.

# Project rationale and operation

## Project benefits

This project will bring considerable value to lots of constituencies stakeholders. The major benefactors will be emergency services, government administration and humanitarian aid communities, the visualisation software in this project will represent a valuable software for improving awareness in a crisis environment. By putting raw imagery into an interactive map, the system will facilitate an automatic analysis which would be easier for humans. This will help people make better and earlier decisions. The second benefit would be the researcher and academic institutions, this project will bring a useful case study on using current AI technologies in the area of crisis and emergency informatics and a useful comparison of suitability of various visualisation protocols.

## Project operation

The project workflow will be in agile and iterative approach. Using the waterfall workflow for AI methods with a new and huge dataset would be inappropriate. This project will be organised as a series of weekly working cycles so that flexibility and the ongoing improvement of the data pipeline, pre-processing and visualisation can take place because of current results. This means that problems can be resolved whilst we are working with other parts and the direction of the project can be shifted due to the results from previous analysis and testing.

Managing the project every day will be retained through a few important methods. The master schedule which is laid out in the Gantt chart will be closely monitored against progressing work, as well as weekly supervisor meetings which is used for review and feedback. An electronic notebook will be used for details of all experimental and decisions taken. Source code will be checked and maintained on GitHub with frequent testing as well as detailed commits. The success will be approached by its flexibility for handling unexpected technical problems and making content advanced towards our goal.

## Options

The roadmap of this project is supported through of continuous assessments of some core technologies to make sure that each part is appropriate to the task. The initial decision as to use a deep learning model for the embedding of images with two competent tools. The first is the multi-modal CLIP model and the second is ResNet-50 design. The choice will be mad based on quality aesthetic coherency of the final visual clusters supported by quantitative measurements. After, we ill compare three dimensionality reduction methods: UMAP, t-SNE and PCA. The assessment will be on the final visual functionality, as well as the performance using a sizeable dataset. Lastly, the visual functionality will be implemented either Streamlit, through its capability for rapid prototyping or using Dash for deeper customisation. This will be dependent on the individual features needed to achieve the project goals. This process ensures that the final decisions is not just workable but consciously an defensibly made through a selection from multiple possibilities.

## Risk analysis and mitigation

An important aspect of project planning is future planning because it prepares for possible issues. Aive risk management is crucial in making sure that the project is progressing and meeting the goals aimed at. I have identified the vital risks which are data, technical and project management which are scored on the likelihood for possible impact. Every risk must have a clear and useful mitigation plan in preparation for possible setbacks. I have made a table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Risk ID** | **Risk description** | **Likelihood** | **Impact** | **Mitigation plan** |
| R01 | Data – Failure to get appropriate dataset | Medium | High | Immediately search for new datasets. |
| R02 | Technical – AI model produces poor results | Medium | High | Compare two models (CLIP, ResNet) and select based on result. |
| R03 | Hardware – PC performance isn’t sufficient for training | Low | Medium | Use Google Colab’s GPU/TPU or use PC’s in DAIM. |
| R04 | Personal – Coursework’s impact project time | Medium | High | Follow the Gantt chart schedule strictly. 1-2 days per week on project. |

## Resources required

This project will be developed as self-sufficient using already available resources. I intend to develop this on my computer which has appropriate specs. It is an open source software stack with the majority being written in Python along with Data science libraries like PyTorch, Scikit-learn that doesn’t require licensing. However, there might be one non-standard resource needed which are cloud computing services for large data manipulation which are available to me and will be taken care of on a limited budget if needed.

## Ethical and legal considerations

Ethical and legal considerations are very important and for our project. A full assessment is to be finished through the individual university checklist, but here we will recognise the main considerations. The main concern is data protection, to meet the standards of GDPR, the project will be using open source datasets that doesn’t have identifiable data being processed. On the Ethics side, the dataset used needs to have a license that allows academic and research to comply with data possession and copyright. There might be algorithmic bias in pre-trained AI models in this project which I have acknowledged and that itself will be discussed in the final project report when e have the results.

## Commercial considerations

What will your project cost to undertake? Could you market the end result? What would your potential market be, and who are the competitors? It may be that you have no intention of commercialising your work, but you should still consider the possibilities.

# Project methodology and outcomes

## Initial project plan

## Tasks and milestones

Present a realistic task list for the entire project, broken down to a suitable level of detail. Indicate milestones against which progress can be monitored. Make sure you include all the deliverables you mentioned earlier.

Delete the red paragraphs and replace this one with your content (use the “Normal” paragraph style).

## Schedule Gantt chart

Present a Gantt chart showing a schedule for all tasks, milestones and deliverables. Show dependencies amongst tasks. If you are intending to use SCRUM or other agile methods, be sure to go to the lectures involving project planning. Your time plan should cover the entire period of your project (and will therefore include the PDD preparation as a task and the PDD itself as a deliverable). Gantt charts work better in landscape format, so rotate yours or add a landscape format section to the document. Don’t be tempted to simply paste a wide image into a page. It needs to be readable if printed out at normal size.

Delete the red paragraphs and replace this one with your content (use the “Normal” paragraph style).

## Project control

How will you manage the project day-to-day? How will its performance be monitored? How will you judge if it has been successful?

Delete the red paragraphs and replace this one with your content (use the “Normal” paragraph style).

## Project evaluation

How will you evaluate the project’s artefacts and overall outcomes? What user evaluation will you do? Do not underestimate the importance of this, and include clear details of how you will do the evaluation. Remember that if you intend to test your outputs on people, you must declare this in your ethics review.

Delete the red paragraphs and replace this one with your content (use the “Normal” paragraph style).

# References

1. GeeksforGeeks (2024). *UMAP: Uniform Manifold Approximation and Projection*. [online] GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/machine-learning/umap-uniform-manifold-approximation-and-projection/>.

2. GeeksforGeeks (2024). *Hierarchical DensityBased Spatial Clustering of Applications with Noise (HDBSCAN)*. [online] GeeksforGeeks. Available at: https://www.geeksforgeeks.org/machine-learning/hdbscan/.

‌

‌

Delete the red paragraphs and replace this one with your content (use the “Normal” paragraph style).

# Appendix a

You may use one or more appendices (label them “Appendix a” “Appendix b” and so on), to add useful reference information which may be relevant to other sections of the report. Do not use appendices simply as a way of writing more than will fit into the main document word count. If you don't need any appendices, then delete this whole section

Delete the red paragraphs and replace this one with your content (use the “Normal” paragraph style).