

**Final Project Ideas**

***This is a collection of starting points for ideas for your Final Project. There are two for each level 6 module.***

April 2025

**Changelog**

v1.0 2025-02-28 All new; only one template for Games Development v1.1 2025-03-25 Both Games Dev templates available v1.2 2025-03-28 Only one template for NLP

# Table of Contents

1. CM3005 Data Science. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4 1.1 Project Idea 1: Data-Driven Personalised Educational Content Recommendation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

1.2 Project Idea 2: Predictive Modelling of Social Media Trend Emergence . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

1. CM3010 Databases and Advanced Data Techniques . . . . . . . . . . . . . . . 10 2.1 Project Idea 1: Working with works–a new approach to music cataloguing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10 2.2 Project Idea 2: Reconciliation service for a dataset. . . . . . . . . . . . . . 13
2. CM3015 Machine Learning and Neural Networks . . . . . . . . . . . . . . . . . . 15
   1. Project Idea 1: Neural Style Transfer . . . . . . . . . . . . . . . . . . . . . . . . . . 15 3.2 Project Idea 2: Deep Learning Breast Cancer Detection . . . . . . . . . 17
3. CM3020 Artificial Intelligence . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
   1. Project Idea 1: Orchestrating AI models to achieve a goal. . . . . . . . 19 4.2 Project Idea 2: Financial Advisor Bot . . . . . . . . . . . . . . . . . . . . . . . . . . 22
4. CM3025 Virtual Reality . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
   1. Project Idea 1: VR Game Escape Room . . . . . . . . . . . . . . . . . . . . . . . . 25 5.2 Project Idea 2: VR Educational Experience . . . . . . . . . . . . . . . . . . . . 30
5. CM3030 Games Development . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34 6.1 Project Idea 1: Custom Car Physics for Racing and Open-World Driving Games . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34 6.2 Project Idea 2: Procedural Dungeon Generation in Roguelike Games . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
6. CM3035 Advanced Web Design . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38
   1. Project Idea 1: Identity and profile management API. . . . . . . . . . . . 38 7.2 Project Idea 2: NextTrack: A music recommendation API . . . . . . . . 40
7. CM3040 Physical Computing and the Internet of Things. . . . . . . . . . . . 42 8.1 Project Idea 1: Secure IoT Device Management in a Safety-critical

Smart Environment . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42

8.2 Project Idea 2: Towards a Low-cost Resilient Early-warning Natural Disaster Monitoring, Notification and Rapid Response Platform . . 44

1. CM3045 3D Graphics and Animation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 47
   1. Project Idea 1: 3D Game with different difficulty levels . . . . . . . . . . 47
   2. Project Idea 2: Animated Short film with VFX. . . . . . . . . . . . . . . . . . . 52 10 CM3050 Mobile Development . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57 10.1 Project Idea 1: Developing a Mobile App for Local Disaster Preparedness and Response . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57

10.2 Project Idea 1: Developing a Gamified AR App for Eco-Conscious Urban Exploration . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 59

1. CM3055 Interaction Design . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61 11.1 Project Idea 1: Designing an Online Empathy Simulation to Help

Caregivers . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61

* 1. Project Idea 2: Designing an Emotion-Aware Adaptive Email and Task Manager to Reduce Workplace Stress. . . . . . . . . . . . . . . . . . . . . . . . . 63

1. CM3060 Natural Language Programming . . . . . . . . . . . . . . . . . . . . . . . . 66 12.1 Project Idea 1: Identifying research methodologies that are used in research in the computing disciplines.. . . . . . . . . . . . . . . . . . . . . . . . . . . 66
2. CM3065 Intelligent Signal Processing . . . . . . . . . . . . . . . . . . . . . . . . . . . 68 13.1 Project Idea 1: Gamified Smart Environment to Promote Mental and Physical well-being in our Everyday Life . . . . . . . . . . . . . . . . . . . . . . . 68
   1. Project Idea 2: Multi-human Detection and Activities Tracking in a Shared Space . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 71

# 1 CM3005 Data Science

## ~~1.1 Project Idea 1: Data-Driven Personalised Educational Content Recommendation~~

**What problem is this project solving, or what is the project idea?**

Can we leverage advanced data science techniques to build a highly accurate and personalised educational content recommendation system?

**What is the background and context to the question or project idea above?**

The explosion of online learning resources presents a challenge: how do learners find the most relevant materials? Traditional recommendation systems often rely on simplistic metrics. This project aims to apply sophisticated data science methods to understand individual learning patterns and preferences. By analysing large datasets of user interactions, content metadata, and learning outcomes, we can build models that predict optimal learning pathways. This involves exploring techniques from machine learning, natural language processing, and network analysis to create a system that truly personalises the educational experience, improving knowledge retention and learning efficiency.

**Here are some recommended sources for you to begin your research.**

* <https://developers.google.com/machine-learning/recommendation/>(Google’s resources on recommendation systems)
* [https://scikit-learn.org/(](https://scikit-learn.org/)Scikit-learn documentation for machine learning)
* <https://www.tensorflow.org/> (TensorFlow documentation for deep learning) • <https://www.nltk.org/> (NLTK documentation for natural language processing)
* Research papers on "Educational Data Mining" and "Learning Analytics" (search on Google Scholar or ACM Digital Library)

**What would the final product or final outcome look like?**

The final product should be a data-driven recommendation engine, likely implemented as a Python API or a web service. It should demonstrate the ability to process user data, apply machine learning models, and generate personalised recommendations. The project should include a thorough evaluation of the model’s performance using relevant data science metrics.

**What would a prototype look like?**

A prototype would demonstrate the core data science pipeline: data ingestion, preprocessing, model training, and recommendation generation. It needs to prove that the chosen algorithms can learn from data and produce meaningful recommendations. It’s important to clearly show the data transformations and model outputs. A complete user interface is not essential at this stage.

**What kinds of techniques/processes/CS fundamentals are relevant to this project?** • Machine Learning: Collaborative Filtering, Content-Based Filtering, Matrix Factorisation, Deep Learning (e.g., Recurrent Neural Networks for sequence modelling). • Natural Language Processing: Topic Modelling, Sentiment Analysis, Text Embeddings (Word2Vec, BERT).

* Data Mining: Clustering, Classification, Association Rule Mining.
* Statistical Analysis: Hypothesis testing, A/B testing.
* Data Visualization: using tools like Matplotlib, Seaborn, or Plotly to show data trends and model performance.

**What would the output of these techniques/processes/CS fundamentals look like?**

* Trained machine learning models (e.g., serialized model files).
* Recommendation scores for each content item.
* User profiles with learned preferences.
* Visualisations of model performance metrics (e.g., precision-recall curves, ROC curves).
* Dataframes containing processed and analysed data.

**How will this project be evaluated and assessed by the student (i.e. during iteration of the project)? What criteria are important?** Evaluation will focus on:

* Precision and recall of recommendations.
* Root mean squared error (RMSE) or other relevant metrics for rating prediction.
* Area under the ROC curve (AUC) for binary classification tasks.
* Statistical significance of A/B testing results.
* The use of cross validation to ensure model generalisation.

**For this brief, what might a minimum pass (e.g. 3rd) student project look like?** • Implementation of a basic recommendation algorithm (e.g., collaborative filtering using a simple matrix factorisation technique).

* Use of a publicly available dataset.
* Basic evaluation using a single metric.
* Clear documentation of the data science pipeline.

**For this brief, what might a good (e.g. 2:2 – 2:1) student project look like?**

**It should meet the minimum pass criteria and in addition:**

* Implementation of a more advanced machine learning model (e.g., a hybrid recommendation system).
* Data preprocessing and feature engineering.
* Evaluation using multiple metrics and cross-validation.
* Clear explanation of the strengths and weaknesses of the chosen approach.

**For this brief, what might an outstanding (e.g. 1st) student project look like?**

**It should meet the minimum pass criteria and in addition:**

* Development of a novel data science approach or adaptation of a state-of-the-art technique.
* Creation or curation of a high-quality dataset.
* Comprehensive evaluation, including comparisons with baseline models and statistical significance testing.
* Detailed analysis of model performance and insights.
* Potentially publishable results in a data science or educational technology venue.

## ~~1.2 Project Idea 2: Predictive Modelling of Social Media Trend Emergence~~

**What problem is this project solving, or what is the project idea?**

Can we develop a data-driven model that predicts the emergence and spread of trends on social media platforms?

**What is the background and context to the question or project idea above?**

Social media platforms are dynamic environments where trends rapidly emerge and evolve. Understanding these trends is crucial for businesses, marketers, and researchers. This project aims to apply data science techniques to analyse large-scale social media data and build predictive models. By analysing factors such as user engagement, content characteristics, and network structure, we can identify patterns that indicate the potential for a trend to gain traction. This project will explore how data science can be used to forecast and understand the dynamics of social media trends, providing valuable insights for strategic decision-making.

**Here are some recommended sources for you to begin your research.**

* <https://www.tweepy.org/> (Python library for accessing Twitter API)
* <https://networkx.org/> (Python library for network analysis)
* <https://scikit-learn.org/> (Scikit-learn documentation for machine learning)
* Research papers on "Social Media Trend Analysis," "Network Diffusion," and "Time Series Forecasting" (search on Google Scholar or ArXiv)
* <https://www.kaggle.com/datasets> (Kaggle datasets for social media data)

**What would the final product or final outcome look like?**

The final product will be a predictive model implemented in Python, along with a report detailing the data analysis, model development, and evaluation. It should demonstrate the ability to process social media data, extract relevant features, and predict the likelihood of a trend’s emergence. The project should include visualizations of trend patterns and model performance metrics.

**What would a prototype look like?**

A prototype would demonstrate the core data science pipeline: data collection (e.g., using APIs), preprocessing (e.g., text cleaning, feature extraction), model training (e.g., using time series models or machine learning classifiers), and trend prediction. It needs to prove that the model can identify patterns associated with trend emergence. It is important to show the data transformations and the model’s predictive capabilities. A real time dashboard is not important at this stage.

**What kinds of techniques/processes/CS fundamentals are relevant to this project?**

* Time Series Analysis: ARIMA, LSTM
* Network Analysis: Centrality measures, community detection
* Natural Language Processing: Sentiment analysis, topic modelling, word embeddings
* Machine Learning: Classification, regression
* Data Visualization: Time series plots, network graphs

**What would the output of these techniques/processes/CS fundamentals look like?**

* Time series forecasts of trend popularity.
* Network graphs showing the spread of trends.
* Sentiment scores and topic distributions related to trends.
* Classification probabilities for trend emergence.
* Visualizations of model performance1 (e.g., ROC curves, time series plots).

**How will this project be evaluated and assessed by the student (i.e. during iteration of the project)? What criteria are important?** Evaluation will focus on:

* Accuracy of trend prediction (e.g., precision, recall, F1-score).
* Root mean squared error (RMSE) for time series forecasting.
* Statistical significance of model results.
* Robustness of the model to different datasets and platforms.
* The use of cross validation to ensure model generalisation.

**For this brief, what might a minimum pass (e.g. 3rd) student project look like?**

* Basic data collection from a social media platform.
* Simple time series analysis or classification model.
* Basic evaluation using a single metric.
* Clear documentation of the data analysis process.

**For this brief, what might a good (e.g. 2:2 – 2:1) student project look like?**

**It should meet the minimum pass criteria and in addition:**

* Integration of multiple data sources and features.
* Implementation of a more advanced predictive model.
* Evaluation using multiple metrics and cross-validation.
* Analysis of the factors influencing trend emergence.

**For this brief, what might an outstanding (e.g. 1st) student project look like?**

**It should meet the minimum pass criteria and in addition:**

* Development of a novel predictive model or feature engineering technique.
* Analysis of trend diffusion patterns using network analysis.
* Comprehensive evaluation, including comparisons with baseline models and statistical significance testing.
* In-depth analysis of the social dynamics underlying trend emergence.
* Potentially publishable results in a data science or social science forum.

# 2 CM3010 Databases and Advanced Data Techniques

## ~~2.1 Project Idea 1: Working with works–a new approach to music cataloguing~~

**What problem is this project solving, or what is the project idea?** Create a portal for catalogues of works by classical composers

**What is the background and context to the question or project idea above?**

Book publications listing all music by a classical composer have been around for centuries

(e.g. Mozart kept one). Software for creating digital lists was developed in 2014 by the Danish national library to support a catalogue of the works of Carl Nielsen. This system uses MEI XML to encode catalogue data, a Java-based XML database, and a Java web forms manager for data entry. The original publication system could generate a PDF (for printing) and web pages.

This software is hard to maintain, and inflexible. As a result, when the Danish library stopped supporting development, and a German university took over, the publication system stopped working.

You should identify ways to work with existing catalogues, currently encoded in XML. Decide whether to keep the data as it is or transform it into a different model (e.g. JSON, SQL or Linked Data), and then make a new interactive web interface for exploring the catalogues.

**Here are some recommended sources for you to begin your research.**

About MerMEId software

**Code:** <https://github.com/Edirom/MerMEId/> (also on DockerHub) **Catalogues published using MerMEId** :

**Carl Nielsen:** <https://www.kb.dk/dcm/cnw/navigation.xq>

**Frederick Delius:** [https://delius.music.ox.ac.uk/catalogue/navigation](https://delius.music.ox.ac.uk/catalogue/navigation.html).

[html](https://delius.music.ox.ac.uk/catalogue/navigation.html) **Literature** :

**2012:** Niels Krabbe & Axel Teich Geertinger, ‘MEI (Music Encoding Initiative) as a Basis for Thematic Catalogues: Thoughts, Experiences, and Preliminary Results’, RISM Conference (2012),

[https://music-encoding.org/downloads/TeichGeertinger\_Final. pdf](https://music-encoding.org/downloads/TeichGeertinger_Final.pdf)

**2022:** Stadler et al, ‘Towards MerMEId 2.0’ Music Encoding Conference (2022), <https://works.hcommons.org/records/hkgc6-ztm37>

**What would the final product or final outcome look like?**

* A user-friendly website for exploring musical works, and their themes, performances and manuscripts
* A web API for work catalogue data
* Transformation scripts to convert MEI works catalogues into a different database model (optional).
* An evaluation of the MEI catalogue model, with recommendations (optional).

**What would a prototype look like?**

The catalogue information in MerMEId is complex. A prototype would take only the most basic data and extract it into a form that can be used for the Web API. The prototype should then consist of a basic API and web front end for this reduced data model.

**What kinds of techniques/processes/CS fundamentals are relevant to this project?** Data modelling and transformation, information retrieval, UI/UX, web programming.

**What would the output of these techniques/processes/CS fundamentals look like?**

A clear evaluation and recommendation for an alternative to the current system, first steps towards that system, including data pipeline, API and UI.

**How will this project be evaluated and assessed by the student (i.e. during iteration of the project)? What criteria are important?**

1. The project must engage with some of the complexities of the data model (not everything needs to be covered, but more than just title and composer).
2. The website should make sense and be appropriate to the data.
3. Web API and backend structure should be clearly thought through and well argued.

**For this brief, what might a minimum pass (e.g. 3rd) student project look like?** Data from a catalogue has been taken and used to build a live website.

**For this brief, what might a good (e.g. 2:2 – 2:1) student project look like?**

A significant proportion of the original data model has been implemented, and most data transformed and presented to the users in a new way.

**For this brief, what might an outstanding (e.g. 1st) student project look like?**

A perceptive evaluation of the data model and alternative approaches, which could be taken further. A significant proportion of the original data model has been implemented, and most data transformed and has been presented to the users in a new way that potentially unlocks new insights into the data. The catalogues of multiple composers can be explored together, or connected to other sources of information (such as Wikidata, Musicbrainz or IMSLP).

## ~~2.2 Project Idea 2: Reconciliation service for a dataset~~

**What problem is this project solving, or what is the project idea?**

You will implement or adapt a Reconciliation Service API for a dataset of your choice to support connecting it to other data.

**What is the background and context to the question or project idea above?**

Datasets become far more powerful when they are connected together – for example, a database with references to places can become more useful if those places can also be connected to socio-economic or historical information about those places. Identifying common entities between datasets is often a manual, time-consuming task. OpenRefine is an example of software designed to make the task easier, but is most effective at connecting data to a dataset which has a Reconciliation API available. Wikidata has this, for example.

A Reconciliation Service API allows software to query entities from a data table and retrieve possible matches from the dataset. It is possible to implement a server for this from scratch, but tools have been developed including at least one Python library.

**Here are some recommended sources for you to begin your research.**

* The API itself is maintained by a W3.org community group at [https://www.w3. org/community/reconciliation/.](https://www.w3.org/community/reconciliation/) This group develops the documentation for the evolving standard.
* An example of a library for implementing the API is here: [https://github.com/ derenrich/py-reconciliation-service-api.](https://github.com/derenrich/py-reconciliation-service-api)
* A description of the goals by a host of one service is here: [https://blog.museum-d](https://blog.museum-digital.org/2024/07/03/reconciliation-apis-arrive-to-museum-digital/)igital. [org/2024/07/03/reconciliation-apis-arrive-to-museum-digital/](https://blog.museum-digital.org/2024/07/03/reconciliation-apis-arrive-to-museum-digital/)
* The originator of the API and main client software is OpenRefine: [https://openref](https://openrefine.org/)ine.

[org/](https://openrefine.org/)

**What would the final product or final outcome look like?**

* A description of the dataset being used, including evaluation of the data consistency and reliability and assessment of any data cleaning needs;
* An implemented reconciliation API as a server. This need not be feature complete

(the standard allows implementations with minimal functionality);

* An example of data being reconciled with the original dataset, using a client that calls the API;
* Evaluation of the efficacy of the API – what works and what would need more data or more development.

**What would a prototype look like?**

A working API server, but with hard-coded or minimal data. The documentation lists a set of possible functional elements, but very little needs to be implemented for the server to conform.

**What kinds of techniques/processes/CS fundamentals are relevant to this project?** Data transformations, advanced web development, text search/information retrieval, data modelling.

**What would the output of these techniques/processes/CS fundamentals look like?**

Good system architecture, efficient database searches, good selection of fields to expose in the API, good choice of dataset(s)

**How will this project be evaluated and assessed by the student (i.e. during iteration of the project)? What criteria are important?**

1. The API must be successfully called by a client (OpenRefine, probably).
2. It must be possible to reconcile matching data.
3. Close matches should, ideally, also be retrieved, with sensible criteria for relevance.

**For this brief, what might a minimum pass (e.g. 3rd) student project look like?**

A minimal API – the reconciliation service exists, can be connected to OpenRefine, and conforms (more or less) to the standard.

**For this brief, what might a good (e.g. 2:2 – 2:1) student project look like?**

A functional API – the service exists, can be connected to from OpenRefine, and returns relevant matches when identical strings of the right datatype appear in the data.

**For this brief, what might an outstanding (e.g. 1st) student project look like?**

An interesting, well-chosen dataset is exposed through the API – the service works from within OpenRefine, and has been tested with a relevant second dataset. Approximate matching, or matching on multiple fields has been implemented. The resulting combined dataset would support further research. A test framework has been written and checks aspects of the API against the specification.

# 3 CM3015 Machine Learning and Neural Networks

## ~~3.1 Project Idea 1: Neural Style Transfer~~

**What problem is this project solving, or what is the project idea?**

Neural Style Transfer is a generative deep learning technique for altering images to make them appear to have been painted in a particular style. For example, a cityscape of skyscrapers and flyovers as if painted in the characteristic circular brush strokes Van Gogh.

**What is the background and context to the question or project idea above?**

The idea is to merge a preserve the content of an image (a cityscape) whilst adopting a reference style (for example, Van Gogh’s Starry Night). Style transfer is related to the area of texture generation in computer vision, but the power of deep learning takes the process to new level. Within the deep learning approach, content, a global property of an image, is captured by the upper levels of a convnet whereas style, a local property, is represented in multiple lower layers. Transfer proceeds by the minimisation of a loss function which measures the style and content distance between the original and the generated image.

**Here are some recommended sources for you to begin your research.**

* Gatys, L.A., Ecker, A.S. and Bethge, M., 2016. Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2414-2423).
* Francois Chollet (2018). Deep Learning with Python. Manning, Shelter Island (pp 287-295).
* Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y. and Song, M., 2019. Neural style transfer: A review. IEEE transactions on visualisation and computer graphics, 26(11), pp.33653385.

**What would the final product or final outcome look like?** A portfolio of stylised images.

A study of the relationship between deep learning architectures, hyperparameter settings and outputs.

**What would a prototype look like?**

An NST system with default settings, as suggested, for example in Chollet’s book.

**What kinds of techniques/processes/CS fundamentals are relevant to this project?**

Machine learning; deep learning; a deep learning library such as Tensorflow or PyTorch.

**What would the output of these techniques/processes/CS fundamentals look like?** A series of deep learning statistical models.

**How will this project be evaluated and assessed by the student (i.e. during iteration of the project)? What criteria are important?**

A review of content manipulation by various styles, and of different contents with the same style.

A survey of the complete hyperparameter space for a particular neural architecture.

An understanding of minimal models and maximal models and the sensitivity of outputs to hyperparameters such as the weighting between style and content distances.

**For this brief, what might a minimum pass (e.g. 3rd) student project look like?** A working NST system with some investigation of hyperparameter adjustment.

**For this brief, what might a good (e.g. 2:2 – 2:1) student project look like?**

A good survey of hyperparameter settings, different architectures (e.g. various VGG models) and production, with comment, of a variety of content and style couplings.

**For this brief, what might an outstanding (e.g. 1st) student project look like?** Novel adaptations (e.g. different optimisers) of readily available code (e.g. from tensorflow.org and Chollet’s book). Imaginative applications e.g. video neural style transfer. Qualitative and quantitative evaluation. Multiple style models. Something comparable in range to the early review paper of Jing et al.

## 3.2 Project Idea 2: Deep Learning Breast Cancer Detection

**What problem is this project solving, or what is the project idea?**

The aim is to establish if Deep Learning assisted X-ray mammography can improve the accuracy of breast cancer screening. The project will achieve this aim by modelling the Digital Database for Screening Mammography (DDSM) with convolutional neural networks (CNNs).

**What is the background and context to the question or project idea above?**

The UK national health service (NHS) launched, in early 2025, a major research project to address the above research question. If successful, a DL system could replace one of the two radiologists currently reporting on scans with the consequence of faster diagnostic turn-around and the liberation of specialists for other tasks. CNNs are deep learning neural networks that apply a succession of filters to the input layer. They are capable of impressive image recognition tasks.

**Here are some recommended sources for you to begin your research.** • Wang L. Mammography with deep learning for breast cancer detection. Front Oncol. 2024 Feb 12;14:1281922. doi: 10.3389 <https://doi.org/10.3389/fonc.2024.1281922>PMID: 38410114; PMCID: PMC10894909.

* Lee, R., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M. & Rubin, D. L. "A curated mammography data set for use in computer-aided detection and diagnosis research.", Sci Data 4, 170177 (2017). <https://doi.org/10.1038/sdata.2017.177>
* Francois Chollet (2018). Deep Learning with Python. Manning, Shelter Island
* [https://www.tensorflow.org](https://www.tensorflow.org/)

**What would the final product or final outcome look like?**

A CNN statistical model, predictions from this model and comparisons to specialist reporting accuracy.

**What would a prototype look like?**

A small CNN capable of achieving statistical power.

**What kinds of techniques/processes/CS fundamentals are relevant to this project?** Artificial neural networks; dataset splitting, model building with tensorflow, training and testing.

**What would the output of these techniques/processes/CS fundamentals look like?**

One or more test metrics for the best statistical model trained on the DDSM dataset.

**How will this project be evaluated and assessed by the student (i.e. during iteration of the project)? What criteria are important?**

The student will seek to improve the chosen test metrics by network scaling up and regularisation.

**For this brief, what might a minimum pass (e.g. 3rd) student project look like?**

A working CNN that at the minimum has statistical power. Some attemot to follow the Depp learning workflow, as described in the text book by Chollet.

**For this brief, what might a good (e.g. 2:2 – 2:1) student project look like?**

A succession of regularised CNN networks; strict adherence to the Deep Learning workflow.

**For this brief, what might an outstanding (e.g. 1st) student project look like?**

Application of transfer learning (e.g. VGG models), and any of the alternatives listed in Table 3 of Wang 2024. For a high first: near publishable results for an original model that is competitive with the best reported DL models in the literature.

# 4 CM3020 Artificial Intelligence

## 4.1 Project Idea 1: Orchestrating AI models to achieve a goal

**What problem is this project solving, or what is the project idea?**

This project is based around the idea of combining multiple pre-trained models into a workflow that achieves a particular goal. It is up to you to decide what that goal is. For example, in the artificial intelligence course, you saw how it is possible to combine multiple models to generate lyrics, music and a singing performance for a pop song. You do not need to address a creative problem like this, but you do need to combine multiple models to solve a well specified problem. Below we provide some starting ideas for the kinds of models you might work with. You should think of a problem space (e.g. music, cybersecurity, creative writing, education, healthcare) then build a system that can operate in that space based around multiple pre-trained models operating on different types of data.

**What is the background and context to the question or project idea above?**

Many pre-trained models are now available that allow ‘artificially intelligent’ computer programs to extract information from the world, and to generate information to put back into the world. As a final year computer science student, we would like to challenge you in this project to identify a problem space that you feel is important to you and to build a system that can operate in that problem space. The system should make use of several pre-trained models to make sense of the world and to put information into the world.

**Here are some recommended sources for you to begin your research.**

Example of the kinds of pre-trained models you can use are as follows. Note that these links are valid and working at the time of publication of this document. You should be able to use them as starting points to find working models.

1. Speech to text: transcribes audio containing speech into text e.g. openai whisper: <https://github.com/openai/whisper>
2. Language processing and generation: using local language models to understand and generate text e.g. ollama: <https://ollama.com/>
3. Image to text: Image analysis and description, detect objects in images, describe images e.g. mobilenet: [https://github.com/tensorflow/tfjs-models/tree/ master/mobilenet](https://github.com/tensorflow/tfjs-models/tree/master/mobilenet) and

<https://huggingface.co/Salesforce/blip-image-captioning-base>

1. Audio to text: audio scene description e.g. yamnet [https://huggingface.co/](https://huggingface.co/STMicroelectronics/yamnet)

[STMicroelectronics/yamnet](https://huggingface.co/STMicroelectronics/yamnet) and [https://github.com/tensorflow/tfhub. dev/blob/master/assets/docs/google/models/yamnet/1.md](https://github.com/tensorflow/tfhub.dev/blob/master/assets/docs/google/models/yamnet/1.md)

1. Sentiment analysis: detect emotions in text [https://huggingface.co/tabulari](https://huggingface.co/tabularisai/multilingual-sentiment-analysis)sai/ [multilingual-sentiment-analysis](https://huggingface.co/tabularisai/multilingual-sentiment-analysis)
2. Human body and face analysis, for example ml5 models for hands and poses: [https:](https://docs.ml5js.org/)

[//docs.ml5js.org/](https://docs.ml5js.org/)

1. Other sources: You can find many pre-trained models for many different tasks here:

<https://huggingface.co/models>

**What would the final product or final outcome look like?**

The final product should consist of a working piece of software that can be used to achieve your specified goal or to operate in your specified problem space. The product should include at least THREE pre-trained models, operating in different domains/ data spaces, for example, text, image and audio. You should clearly state what the purpose of the system is and show how you went about identifying, operationalising, testing etc. your models. It is likely that you will have to go through a process of testing and rejecting models to find the ideal choices for your project. We would like to see evidence of that process and your decision making process.

**What would a prototype look like?**

A prototype should show the chosen models operating successfully, generating and processing data in several different models, with a clearly described objective or purpose for the overall system.

**What kinds of techniques/processes/CS fundamentals are relevant to this project?**

This project will involve working through lots of different models – getting them working, testing them out to see if they will work with the data you want to use. We want you to use your software engineering and testing skills here.

You will need to work with different types of data, figuring out how to feed it into the different models.

You should show that you can evaluate pre-trained models for example by sending test data and validating the outputs.

Also you might need to test the performance of the models and to relate this to the system you are building and how it shall be used.

**What would the output of these techniques/processes/CS fundamentals look like?** We are keen to see evidence that you have tested several models and made decisions about which are and are not appropriate for your needs. One aspect of this will be testing the performance of the models to see if they are viable to use for your project. We will want to see evidence that you have thought about how to combine different models to achieve an overall system goal.

**How will this project be evaluated and assessed by the student (i.e. during iteration of the project)? What criteria are important?**

As noted above, we anticipate that you will be installing and testing lots of different models in order to figure out which ones are appropriate for your needs. You should evaluate your process here and be able to come up with an overarching method for doing this. You should ensure that you meet the requirement to have at least THREE pre-trained models working on different types of data in the project. You should ensure that the project delivers a working system that achieves a clear goal. Just downloading some models and running them is not enough!

**For this brief, what might a minimum pass (e.g. 3rd) student project look like?**

A minimum pass would include a working system which includes three pre-trained models working to achieve a clearly specified goal. There should be evidence that you have evaluated the models using appropriate criteria.

**For this brief, what might a good (e.g. 2:2 – 2:1) student project look like?**

A 2:2 – 2:1 project would include a working system which includes three pre-trained models working to achieve a clearly specified goal. You should have chosen a challenging goal which is not easily achievable using standard software engineering techniques and automation – it should be clear why these pre-trained models are needed. We are looking for an integrated piece of software here. We would like to see evidence of software testing, ideally with unit testing but potentially with users as well.

**For this brief, what might an outstanding (e.g. 1st) student project look like?**

An outstanding project would use an original approach to address an interesting and challenging problem. We would like to see evidence that you have thoroughly explored the space of pre-trained models relevant to your project and chosen appropriate models based on evidence presented in the report. You should choose a difficult problem and consider the best way to design the user interaction. We want to see thorough evaluation of the models and solid software testing, including user testing and iteration of the design based on the testing results.

## 4.2 Project Idea 2: Financial Advisor Bot

**What problem is this project solving, or what is the project idea?**

This project involves the creation of a financial advisor bot. The bot analyses financial data in order to make recommendations for a dynamic investment strategy, which you might also call ‘active portfolio management’.

**What is the background and context to the question or project idea above?**

You will need to identify which kind of financial systems the bot will advise about. For example, the stock market, currency exchange market, crypto-currencies, classic car trading and so on.

You will need to decide what kind of AI/ML techniques you will use to implement the bot. We recommend that you use techniques that you have seen in the course, for example, modelling the problem as ‘decision making under uncertainty’ would lend itself to reinforcement learning. Or you might attempt to evolve an investment strategy.

It should be possible for a non-technical user to interact with the bot to receive advice, for example, via a web interface. The bot should present its analysis and recommendations to the user with explanations. For example, the bot might say ‘NVIDIA stock is likely to rise by 25% so we recommend taking up a position on that now and exiting after 20% growth’. It is up to you to decide how the bot presents this advice. You could consider using local language models to format the recommendations into prose.

**Here are some recommended sources for you to begin your research.** You can find stock price data feeds: <https://github.com/ranaroussi/yfinance>Or crypto data feeds: <https://github.com/ccxt/ccxt>

You might need a backtesting system to test your advisor’s trading strategies: <https://github.com/topics/backtesting>

You might consider using local language models to help the user interact with the advisor bot. For example, you might conceive of the bot as a tool-using agent. The following systems might prove useful:

<https://ollama.com/><https://docs.openwebui.com/>

**What would the final product or final outcome look like?**

The final product would be an integrated and tested system designed for non-technical users to use. You will probably need separate systems for training models and gathering data, if you choose to do that. We would not expect the non-technical user to be able to interact with the model training system. But they should be able to interact with the advisor bot once trained. You should present evidence that you have evaluated the advice that the advisor bot generates.

**What would a prototype look like?**

A prototype would need to have the main parts functioning but it would not need to have the full user interface. So you would need a way to interact with the advisor bot, even if it is command line based. You would need to identify appropriate data sources and to gather and ingest that data.. You would need a way for the advisor bot to analyse and/ or learn from data.

**What kinds of techniques/processes/CS fundamentals are relevant to this project?** You can choose how to implement the advisor bot, but reinforcement learning, evolutionary strategies, neural networks and so on would be the typical techniques. You might want to use some language model techniques as well.

You will need to do some significant software engineering and testing here so you would need the techniques from the degree relating to those.

There is also a strong data science component around gathering and analysing data.

Web programming will be needed to create the web interface.

**What would the output of these techniques/processes/CS fundamentals look like?**

* We expect to see well tested software
* We expect to see a solid data analysis and machine learning/ AI workflow implemented
* We expect to see a working web user interface

**How will this project be evaluated and assessed by the student (i.e. during iteration of the project)? What criteria are important?**

We imagine that you will work iteratively on designing, implementing and testing the components of the system. The components include the data gathering, training, advisor bot interaction, web interface.

**For this brief, what might a minimum pass (e.g. 3rd) student project look like?**

A basic project should present a complete working system but with limited complexity of components. We would need to see some design, implementation and evaluation work.

**For this brief, what might a good (e.g. 2:2 – 2:1) student project look like?**

A 2:2-2:1 project should present a complete working system with some significant development and testing work evident in some or all of the components. You should present some detailed evaluation of the components of the system and clear planning. We would want to see some thought put into the design of the advisor bot and its algorithm in particular.

**For this brief, what might an outstanding (e.g. 1st) student project look like?**

An outstanding project should present a complete working system with some significant development and testing work evident in all of the components. You should present detailed evaluation and testing of the components of the system and clear planning. We would expect the advisor bot to implement some sort of interesting and advanced algorithm with significant effort put into it.

# 7 CM3035 Advanced Web Design

## 7.1 Project Idea 1: Identity and profile management API

**What problem is this project solving, or what is the project idea?**

We have different names, usernames and ways of presenting ourselves in different contexts. Devise a Web API to manage them securely and flexibly.

**What is the background and context to the question or project idea above?**

Many people present themselves differently in different contexts. Some have legal names that differ from the names they use in their daily lives, or use middle names or numbers to distinguish similarly-named individuals. In some cultures, their given names may be only used administratively, and in other cases, they may have special names for religious use. On marriage, people may change their legal names, but not use the new name professionally. Meanwhile, online, we manage many identities, usernames and profiles of ourselves. Which identity we use depends on context, but also on who is asking for the information and why. You will devise a REST API for reading and writing identities to a server securely and sensitively, so that users requesting information about an individual get the correct preferred name for the context about which they are enquiring. Confidential information should be concealed as necessary, and users should only be able to write data where this is appropriate.

**Here are some recommended sources for you to begin your research.**

* For context about naming, a reasonable start is the Wikipedia article ([https://en. wikipedia.org/wiki/Personal\_name)](https://en.wikipedia.org/wiki/Personal_name), but better references should be found.
* For secure account management and REST API creation, course materials should provide a good start.
* A good understanding of the HTTP protocol will be helpful – e.g. students might use Accept-language in content negotiation in this context. An overview is at [https:// developer.mozilla.org/en-US/docs/Web/HTTP/Content\_negotiation](https://developer.mozilla.org/en-US/docs/Web/HTTP/Content_negotiation)

**What would the final product or final outcome look like?**

A requirement specification (based on literature survey and, optionally user consultation) for an identity management system An implementation that at least covers names and, optionally, adds gender identities, online profile data and other relevant information.

**What would a prototype look like?**

A completed API, a backend datastore and a UI. Some form of authentication and user management will be necessary.

**What kinds of techniques/processes/CS fundamentals are relevant to this project?** Web API, Data/web security, data modelling, social computing

**What would the output of these techniques/processes/CS fundamentals look like?**

A study of the requirements for such a system and a secure implementation

**How will this project be evaluated and assessed by the student (i.e. during iteration of the project)? What criteria are important?**

* A good understanding of the user needs
* A well-thought-through API

**For this brief, what might a minimum pass (e.g. 3rd) student project look like?**

To pass, an implementation must be secure and have appropriate user management. It must at least cover names used in different contexts. Several different types of use must be implemented, and the API must have clear logic.

**For this brief, what might a good (e.g. 2:2 – 2:1) student project look like?**

A requirements document with evidence of reading and thinking about the contexts. Strong security and privacy decisions. An API that supports a good variety of use contexts. An example web front end works and is user-friendly and accessible.

**For this brief, what might an outstanding (e.g. 1st) student project look like?**

A strong requirements document that shows research and imagination. Solid security and privacy, possibly with reference to GDPR and optionally using external authentication. The data model goes beyond the simplest use cases. The API is robust and would be widely applicable. A web front end clearly demonstrates multiple use cases for the API.

## ~~7.2 Project Idea 2: NextTrack: A music recommendation API~~

**What problem is this project solving, or what is the project idea?**

Design a music recommendation API that gives a ‘next track’ based on an HTTP request, providing listening history and some preference parameters, along with data about the tracks available.

**What is the background and context to the question or project idea above?**

Music recommendation and playlist generation systems are popular and much researched. It’s quite common for these to operate within applications that users log into, and which profile that user based on their listening habits. This has implications for privacy. It can also be undermined by shared accounts and listening (for example, a household with a child, may wind up with nursery rhymes in the recommendations).

You will build a RESTful recommendation system with no user tracking instead, the user will provide a sequence of track identifiers that the next track should follow from, and some preference parameters (which you can choose). The system will use data from external sources, such as musicbrainz, genius.com, spotify, Wikidata to inform its choices.

**Here are some recommended sources for you to begin your research.**

The ISMIR conference (ismir.net) has many papers on recommender systems that should give an idea of the sorts of things you could do. There’s no necessity to use audio features – some of Brian Whitman’s work in early ISMIR conferences uses non-content information, such as metadata and online reviews.

**What would the final product or final outcome look like?**

A RESTful API that takes a set of track identifiers and other parameters and returns the id of a suitable next track. There should be some form of evaluation (for example, based on user testing or some prior work giving good sequences).

**What would a prototype look like?**

The basic API. This could return tracks at random for prototyping purpose, but it should reproduce the input/output format. If a frontend is to be implemented, a basic implementation with music player (e.g. via YouTube or logged in Spotify).

**What kinds of techniques/processes/CS fundamentals are relevant to this project?**

Web API, UI/UX, Music information Retrieval

**What would the output of these techniques/processes/CS fundamentals look like?**

A stateless playlist generating API that has a strategy for choosing new tracks that is better than random selection.

**How will this project be evaluated and assessed by the student (i.e. during iteration of the project)? What criteria are important?**

* A good overview of playlists/recommender systems and what users might want
* A well-thought-through RESTful API that offers some user control
* Sensible evaluation of the results

**For this brief, what might a minimum pass (e.g. 3rd) student project look like?**

A working API that gets some data from elsewhere, and combines it with user-provided information to choose a track.

**For this brief, what might a good (e.g. 2:2 – 2:1) student project look like?**

A literature review that identifies a good strategy for recommending the next track, a working API that offers real choice to the user. A Web application that demonstrates that the API works. Good evaluation (such as user testing)

**For this brief, what might an outstanding (e.g. 1st) student project look like?**

An insightful overview of recommendation approaches and the data that is available to a server process. An API that recommends based on well-chosen parameters. A web application that provides an interactive music playing experience. Strong user testing and reflection on the results of it (ideally, modifications based on user testing, which are themselves tested).

# 12 CM3060 Natural Language Programming

## 12.1 Project Idea 1: Identifying research methodologies that are used in research in the computing disciplines.

**What problem is this project solving, or what is the project idea?**

**What is the background and context to the question or project idea above?**

All research should be guided by a process that begins with the researcher’s philosophical world view and then details why the research has value, how the research was done, and why the particular approach was used. In addition, this process should be made known to the research community. The “how” and “why” are the chosen research methodology. The idea behind this project is to take published research (journal articles, theses, etc.) in the computing fields, determine the computing discipline (Computer Science, Information Systems, Information Technology, etc.), determine the field within the discipline and then determine what research methodologies are used in these publications.

**Here are some recommended sources for you to begin your research.**

D Jurafsky and J Martin. Speech and Language Processing. An Introduction to Natural

Language Processing, Computational Linguistics, and Speech Recognition with Language Models, Third Edition draft, University of Colorado at Boulder

1. J. Oates. Researching Information Systems and Computing. SAGE, London, 2006.
2. Pilkington and L. Pretorius. A conceptual model of the research methodology domain.In Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 2: KEOD, pages 96–107, Lisbon, Portugal, November 2015.

The Natural Language Toolkit [https://www.nltk.org](https://www.nltk.org/)

**What would the final product or final outcome look like?**

A system that classifies research publications according to computing discipline, field and the research methodologies used in that publication.

**What would a prototype look like?**

Some progress on the way to achieving the above.

**What kinds of techniques/processes/CS fundamentals are relevant to this project?**

Natural Language Processing – Information Retrieval in particular, but the whole spread of NLP is in the frame for this project.

Depending on the approach taken, some Machine Learning techniques.

Some of the NLP ‘black box’ approaches as appropriate.

**What would the output of these techniques/processes/CS fundamentals look like?** A research methodology classification pipeline for the computing disciplines.

**How will this project be evaluated and assessed by the student (i.e. during iteration of the project)? What criteria are important?**

Depending on the approach taken, the relevant and appropriate metrics should be used to obtain feedback as to whether the approach is suitable, and there should be some points at which possible other approaches are considered and justified.

**For this brief, what might a minimum pass (e.g. 3rd) student project look like?**

A system that is able to take a publication as input, and isable to determine the computing discipline.

**For this brief, what might a good (e.g. 2:2 – 2:1) student project look like?** As above, but able to also identify field within the discipline.

**For this brief, what might an outstanding (e.g. 1st) student project look like?**

A complete working classification pipeline that gives useful output.