Final Project Ideas

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

Scroll through, or use the navigation bar within your pdf viewer to jump quickly to modules and project briefs you might be interested in.

You need to choose one of the following ideas and use it as the starting point for your project proposal – which will still need to be written and submitted by yourself as part of your assessment for the Final Project module.

CM3020 - Artificial Intelligence

Project Idea Title 1: Kane and Abel: AIs that play games

What problem is this project solving?

Write this as one SHORT question.

What is the background and context to the question above in 150 words or fewer?

The problem is to implement and compare different approaches to playing games with AIs.

This project involves the implementation of two AI systems (Kane and Abel) which can play the same game. One system should have a pre-programmed, behaviour, for example using a finite state machine or other appropriate method. The other system should use some sort of statistical machine learning techniques to learn how to play the game. It is acceptable to adapt the second, machine learning based AI from published source code but the first AI should be written by the student.

AI researchers have long seen game playing as an excellent testbed for AI techniques, and as a marker of the state of the art in AI. There is a wealth of background work covering a wide range of AI techniques and how they can be used to develop game playing AIs. In this project, the student has an opportunity to dive into this fascinating body of work and to attempt to build their own systems.

List some recommended sources for students to begin their research

- Świechowski, Maciej. "Game AI competitions: Motivation for the imitation game-playing competition." 2020 15th Conference on Computer Science and Information Systems (FedCSIS). IEEE, 2020.
  → great recent reference for the various AI game player competitions out there

- Justesen, Niels, Michael S. Debus, and Sebastian Risi. "When are we done with games?." 2019 IEEE Conference on Games (CoG). IEEE, 2019.
  → Some interesting thoughts about the motivations and future of AI game players

- Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." nature 518.7540 (2015): 529-533.
  → classic paper about DQN and Atari

What would the final product look like?

(e.g. presentation, usability, functionality, results)?

Presentation: We would expect the student to present in various media, detailed information about the following:

- Review of related work, especially the background to the two game playing system designs
- Description of the game that the AIs will play
- Description of the implementation of the two AI systems
- Properly organised and commented source code for the two implementations

- Evaluation of the two systems, and comparison to human players if appropriate

What would a prototype look like?

What would it show?
What does it need to prove?
*What **IS** important to make clear?*
*What is **NOT** important at this stage?*

We recommend that at the prototype stage, the student should have the simulation environment up and running and that they should have at least one of the AI systems interacting with the environment.

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

- Review of relevant literature and description of the problem domain
- Explaining how the systems interact with the game environment
- Programming the two AI systems, noting that the machine-learning based one might be adapted from published code (e.g. DQN etc.)
- Describing the implementation of the two systems
- Evaluating the performance of the two systems and measuring the effect of different settings

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

- Review of relevant literature and description of the problem domain
    - Section in the report describing similar work in the literature and describing the problem domain
- Explaining how the systems interact with the game environment
    - Section in the report
- Programming the two AI systems, noting that the machine-learning based one might be adapted from published code (e.g. DQN etc.)
    - Well organised and commented code
- Describing the implementation of the two systems
    - Detailed technical description in the report
- Evaluating the performance of the two systems and measuring the effect of different settings
    - Presentation of appropriate tables, graphs and commentary showing how the systems perform

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

What criteria are important?

We would expect the student to evaluate the following elements:

- Are the AIs able to interact with the game?

- Are the AIs well described and implemented according to a specification?
- How well do the AIs perform?
- Is the code well organised and well commented?
- Is it absolutely clear which code has been written by the student and which has not?
- Are the descriptions in the report sufficient for a tutor to understand how the code works and how the system has been evaluated?

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

- Brief but limited review of the literature
- Working but simplistic attempt at implementing at least one AI game player
- Limited evidence of evaluation
- Limited but complete report

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

- Review of the literature which shows evidence of wide reading
- Working attempts at implementing two AI game players, with evidence of significant, iterated development effort on the part of the student
- Well presented evidence of meaningful evaluation of the two AIs
- Complete, clearly written report

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

- Extensive review of the literature which shows evidence of wide reading and critique of previous work
- Two fully working AI game players, with significant and challenging technical effort involved in their implementation by the student
- Evidence of extensive evaluation and iterated development of different aspects of the AI game players.
- Complete, clearly written report

Project Idea Title 2: Automated design using evolutionary computation

What problem is this project solving?

The problem is how to automate the design of complex, possibly moving structures. For example, moving robots or neural networks.

This project involves the creation of a system which uses evolutionary computing or artificial life techniques to carry out a directed exploration of a space of possibilities. It is up to the student to define what the space represents – the example seen in the AI course was the space of possible forms for 3D creatures. The method of exploration is also up to the student, again, the example seen in the AI course was to direct the exploration towards creatures with desirable characteristics. A successful project will contain a functional system with a well defined space and a well defined method for exploring that space.

What is the background and context to the question above in 150 words or fewer?

Statistical machine learning techniques currently dominate in many problem domains such as image analysis and speech analysis. However, there is another class of problem where the technique of learning statistical patterns in large datasets cannot necessarily be applied. Evolutionary computation and artificial life techniques can be applied to problems where there is minimal data available and no pre-existing correct and incorrect solutions.

List some recommended sources for students to begin their research

- Lehman, Joel, et al. "The surprising creativity of digital evolution: A collection of anecdotes from the evolutionary computation and artificial life research communities." Artificial life 26.2 (2020): 274-306.
  → an introduction to a range of systems that evolve forms
- Sims, Karl. "Evolving virtual creatures." Proceedings of the 21st annual conference on Computer graphics and interactive techniques. 1994.
  → classic paper evolving moving creatures
- Yee-King, Matthew John. "The use of interactive genetic algorithms in sound design: a comparative study." Computers In Entertainment 14.3 (2016).
  → evolving sound synthesis circuits
- Thompson, Adrian. "An evolved circuit, intrinsic in silicon, entwined with physics." International Conference on Evolvable Systems. Springer, Berlin, Heidelberg, 1996.
  → classic paper about evolving circuits -might be possible with a cheap, modern day FPGA, but challenging!

What would the final product look like?

(e.g. presentation, usability, functionality, results)?

Presentation: We would expect the student to present in various media, detailed information about the following:

- Review of related work
- What is the problem the system is addressing?
- How has the solution space been represented parametrically – how are the solution represented at a data level and how are they expressed such that they can be tested?
- How does the algorithm work?

- How well does the system perform? Do the solutions improve over the runtime of the algorithm? * How did the student overcome challenges and make changes to the encoding scheme, testing environment and so on to improve the system's performance?

What would a prototype look like?

What would it show?
What does it need to prove?
*What **IS** important to make clear?*
*What is **NOT** important at this stage?*

We recommend that at the prototype stage, the student has specified and ideally, implemented the encoding scheme. So it should be clear what the problem is (e.g. designing robots, designing buildings etc.) and how the system encodes the problem space (e.g. genetic encoding scheme). The student should also have a clear and viable technical plan for how the potential solutions can be evaluated (e.g. running robots in simulation).

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

- Review of relevant literature and description of the problem domain
- Encoding solutions in a manner that is appropriate for the application of evolutionary computing
- Designing a fitness function suitable for measuring the performance
- Using a genetic algorithm to iteratively evolve a population of solutions to a well specified problem
- Ensuring the evaluation of solutions is efficient enough to allow the evaluation of thousands of solutions in a reasonable time
- Evaluating the performance of the system as a whole and measuring the effect of different settings

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

- Review of relevant literature and description of the problem domain
    - Section in the report describing similar work in the literature and describing the problem domain
- Encoding solutions in a manner that is appropriate for the application of evolutionary computing
    - Genetic encoding scheme clearly described in the report and implemented in code
- Designing a fitness function suitable for measuring the performance
    - Fitness function(s) clearly described and implemented in code
- Using a genetic algorithm to iteratively evolve a population of solutions to a well specified problem
    - Genetic algorithm implemented (can use a pre-made library if that helps), genetic operators, selection and breeding working
- Ensuring the evaluation of solutions is efficient enough to allow the evaluation of thousands of solutions in a reasonable time
    - Measurements of how long it takes to evaluate solutions and demonstrating that this allows meaningful evolution.

- Evaluating the performance of the system as a whole and measuring the effect of different settings
    - Evidence presented in the report of multiple runs of the system showing variation in performance between runs. A comparison of runs with different parameter settings.

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

What criteria are important?

- Does the encoding scheme work?
- Does the fitness function work?
- Does the overall algorithm work, i.e. do the solutions improve as the algorithm proceeds?
- How do features of the encoding scheme, fitness function and evolutionary algorithm impact on the performance? Can you compare performance with different features enabled or disabled?
- Is the code well organised and well commented?
- Is it absolutely clear which code has been written by the student and which has not?
- Are the descriptions in the report sufficient for a tutor to understand how the code works and how the system has been evaluated?

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

- Brief but limited review of the literature
- Working but simplistic encoding scheme with examples of genetic data and how that is converted into solutions
- Working fitness function
- Limited evidence of evaluation
- Limited but complete report

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

- Review of the literature which shows evidence of wide reading
- Working encoding scheme with examples of genetic data and how that is converted into solutions
- Working fitness function
- Evidence of effort to optimise the performance of the system
- Evidence of meaningful evaluation of different aspects of the encoding scheme, fitness function etc.
- Evidence of significant technical work on the part of the student
- Complete, clearly written report

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

- Extensive review of the literature which shows evidence of wide reading and critique of previous work
- Working encoding scheme with examples of genetic data and how that is converted into solutions

- Working fitness function with multiple elements
- Evidence of successfully optimising the performance of the system
- Evidence of extensive evaluation and further development of different aspects of the encoding scheme, fitness function etc. involving challenging technical work by the student
- Complete, clearly written report

CM3015 Machine Learning and Neural Networks

Project Idea Title 1: Deep Learning on a public dataset

What problem is this project solving?

Choosing, based on a quantitative evaluation, a well performing machine learning model for used with a publicly available dataset.

What is the background and context to the question above in 150 words or less?

Pick a dataset from Kaggle.com – choose one that interests you or you think is important – for example: tweets, faces, lung scans, skin diseases, student grades…

Develop a deep learning classificatation/regression model for your chosen dataset by following the methodology of Deep Learning with Python. Aim to find the best model – work from simple to deep and employ the advanced techniques of Chapter 7.

List some recommended sources for students to begin their research

- F.Chollet, Deep Learning with Python, 1st ed.
- Kaggle.com

What would the final product look like?

(e.g. presentation, usability, functionality, results)?

A research project – the final product is a report

What would a prototype look like?

What would it show?
What does it need to prove?
*What **IS** important to make clear?*
*What is **NOT** important at this stage?*

The prototype is a baseline model that achieves a common sense prediction.

It is not important to achieve the accuracy of any puplished paper on this dataset (or any of the Kaggle public notebooks)

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

Jupyter notebooks

Tensorflow, matplotlib and associated Python libraries

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

Model code

Validation plots

Prediction on test set

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

What criteria are important?

Does the model significantly improve on a commonsense baseline

Have I investigated all the alternatives

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

- Any original model that runs and produces a prediction
- A basic evaluation of the model on the public dataset
- Report is well-strutured

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

In addition to minimum pass criteria:

- A sequence of orginal models of increasing depth and sophistication
- An evaluation of the different models using the public dataset, which makes it possible to draw conclusions about the effectiveness of different models and choose a preferred model
- Report: Correct application of the DL methodology; good standard of written, technical English

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

In addition to the good criteria:

- Replication of a hight quality published model(s) on the chosen dataset
- An evaluation of the models to the standard of academic research
- The report is a self-contained explanation and account of theory and experiment. There is a literature review and the work is contextualised. Critical comparison of best model and re-implemented model(s) and results in the literature.

Project Idea Title 2: Gather your own dataset

What problem is this project solving?

Gathering a data set and training a model to develop a classification system

What is the background and context to the question above in 150 words or less?

Most practical machine learning involves gathering data, labelling it and using it to train a machine learning model. The data gathering challenge as it is important to gather data that is representative of real world data and that distinguishes well between the different classes. Contemporary Deep Learning techniques often need huge datasets, into the 10s or 100s of thousands of items, which would not be feasible for this kind of project. But it is possible to build on pre-trained datasets which can be used as features extractors. The aim of this project is to gather a dataset of a moderate size (100s), including both training and testing sets. We recommend a dataset of images as there are plenty of good pre-trained models. You can either generate the dataset yourself (taking images or video) or curate existing unlabelled images. You should label the images and train a machine learning model, based on a pre-trained model, on the dataset. You should try a number of variant models, from simple to complex and using different pre-trained models if they exist, and compare the results. Iterative development should include both improvements to the model and dataset.

List some recommended sources for students to begin their research

- F.Chollet, Deep Learning with Python, 1st ed. Section 5.3 discusses how to use pretrained models
- Model Zoo is a good source of pre-trained models https://modelzoo.co/category/computer-vision

What would the final product look like?

(e.g. presentation, usability, functionality, results)?

A dataset (a collection of labelled images, or similar data), and code for training a model. The main output would be a report showing results of the data.

What would a prototype look like?

What would it show?
What does it need to prove?
*What **IS** important to make clear?*
*What is **NOT** important at this stage?*

The prototype is a small version of the dataset, tested on a simple baseline model. It needs to show that it is possible to train a basic model on the types of data needed. It needs to ensure that the data is in a suitable format for machine learning and that it works with the machine learning platform being used.

It does not have to show accurate classification at this state.

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

Basic image storage and editing. Data labelling.

Jupyter notebooks

Tensorflow, matplotlib and associated Python libraries

Pretrained models

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

A dataset (e.g. images stored in a suitable format and folder structure)

Model code

Validation plots

Prediction on test set

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

What criteria are important?

Does the model significantly improve on a commonsense baseline

Have I investigated all the alternatives

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

- A basic data set gathered
- Using the data to train on pre-trained model with some minimal training.
- Basic evaluation metrics on the result
- Report is well-strutured

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

In addition to minimum pass criteria:

- A well curated dataset that is suitable for training a machine learning model
- The data set covers a good range of possible variants of each class
- A well chosen pre-trained models used as the basis of a series of full machine learning models used to train on the dataset
- A thorough evaluation of the different models
- Report: Correct application of the DL methodology; good standard of written, technical English

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

In addition to the good criteria:

- A novel idea for a data set
- Very good coverage of different variants of each class
- A number of well designed machine learning models culminating in a complex but appropriate model
- An evaluation of the models to the standard of academic research
- The report is a self-contained explanation and account of theory and experiment. There is a literature review and the work is contextualised. Critical comparison of best model and re-implemented model(s) and results in the literature.