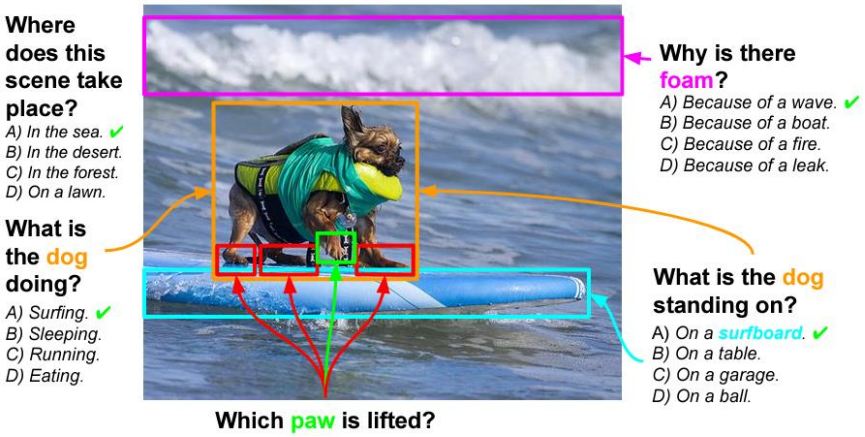
# University of Lincoln Assessment Framework
# Assessment Briefing Template 2024-2025

| | |
|---|---|
| **1. Module code & title** | CMP9137M Advanced Machine Learning |
| **2. Assessed learning outcomes** | [LO3] Use machine learning software to solve complex real-world problems in an application domain of interest. |
| **3. Assessment title** | Visual Question Answering and Game Policy Learning |
| **4. Contribution to final module mark (%)** | 50% |
| **5. Description of assessment task** | This assessment comprises two tasks on machine learning as follows.<br><br>**Task 1**<br>You are required to use Machine Learning to tackle the problem of "Multi-Choice Visual Question Answering". Given an image and candidate textual descriptions of the image, the task is to predict the correct answer from the available candidates. This implies a multi-class classification task, where the image and candidate answer correspond to each other (i.e., only one candidate answer can be considered as correct), see examples in the figure below. The data for this task will be generated by the delivery team – due to the size of data – from the original dataset, which was proposed by (Zhu et al, 2016, CVPR) to investigate methods at the intersection between machine learning, computer vision, and natural language processing.<br><br>  <br>*Figure 1 Example visual question-answers, image from (Zhu et al, 2016).* |

Your task consists of training and evaluating at least three machine learning classifiers for matching images and question-answer pairs. To do that, you will use a unique dataset (smaller than the originally proposed due to compute requirements), which will be available via Blackboard. The dataset will contain different data splits for training, validation, and test examples. You are reminded that the test data should not be involved during training. It should only be used for testing your trained classifiers, and it should be used only to report the performance of the predictions of your proposed models.

You are expected to explore a range of machine learning classifiers inspired by the materials presented and discussed within the module and beyond (i.e., from reading related literature). You will investigate and compare their performance and critique them to justify your recommendations. Your comparisons should include metrics such as Classification Accuracy, Mean Reciprocal Rank, Training & Test Times and any other metrics that you consider relevant. In this assignment you are free to train any classifier, to do any pre- and/or post-processing of the data, and to implement your own algorithm(s) whenever possible instead of only using libraries. You are allowed to use publicly available libraries to train your models. For example, you can use any deep learning frameworks and tools such as Keras, Tensorflow, Pytorch, Large Language Models (e.g., via pre-trained models available on HuggingFace, with finetuning if opted). But you will need to clearly mention your resources, acknowledge appropriately, and compare between classifiers and their results in your own report.

**Task 2**

You are required to use Machine Learning to tackle the problem of "Game Learning". Your goal in this task is to train Deep Reinforcement Learning (DRL) agents that receive image-inputs from a game simulator, and that output game actions (such as turn left/right and shoot) to play the game autonomously. The following simulator will be used to play the game:

- Link to tutorial and github code: ViZDoom (put.edu.pl)
- Reference*: M. Wydmuch, et al. "ViZDoom Competitions: Playing Doom from Pixels". IEEE Transactions on Games, 11(3), 2019*.

You are required to use your knowledge acquired in the module regarding DRL agents, and knowledge acquired from additional recommended readings. This will be useful to investigate the performance of those agents, and to compare and criticise them so you can recommend your best agent. You are expected to evaluate your agents using metrics such as Avg. Reward, Avg. Game Score, Avg. Steps Per Episode, and Training and Test Times – and any others that you consider relevant.

You are expected to train at least three different agents, which can differ in their state representation (CNN, Transformer, CNN-Transformer) and/or

| | |
|---|---|
| | different learning algorithms or training methodologies. Once you have decided the agents that you want to report, you should train them with multiple seeds (3 at least ) and average their results—to reduce the potential noise (due to randomness) in the performance of your models. If you report learning curves, they should be based on those average results instead of using a single seed (run). You are expected to justify your choices in terms of architectures, hyperparameters and algorithms.<br><br>In this assignment, you are free to train any DRL agent, in any programming language, to pre-process the data, and to implement your own solutions whenever possible. While you are free to use DRL libraries such as StableBaselines3, (reference) you should mention your resources used, acknowledge appropriately, and compare between agents in your report. Reference: A. Raffin, et al. "Stable-Baselines3: Reliable Reinforcement Learning Implementations", Journal of Machine Learning Research, 2021.<br><br>Please read the Criterion Reference Grid for details on how your work will be graded. |
| **6. Assessment submission instructions** | You must make an electronic submission of your work in **PDF format** (not MS Word) by using the assessment link on Blackboard for this component. You must attend the lectures and workshops for further details, guidance, and clarifications regarding these instructions.<br><br>DO NOT include this briefing document with your submission.<br><br>The deadline for submission of this work is included in the School Submission dates on Blackboard. |
| **7. Date for return of mark and feedback** | Please see the School assessment dates spreadsheet (available via Blackboard). |
| **8. Feedback format** | Written and numerical feedback will be provided via Blackboard, and additional feedback can be provided upon request in a meeting or via email. |
| **9. Use of Artificial Intelligence (AI) in this assessment** | In this assessment you are allowed to make use of publicly available resources including libraries or chatbots such as ChatGPT. However, the use of AI tools is not permitted in the generation of the final report for this assessment. Please note that your report should be written by yourself—even if parts of your solutions are derived from responses of a chatbot. In other words, **chatbots should only be used to increase your understanding instead of writing the assignment for you**. |
| **10. Marking criteria for assessment** | A Criterion Reference Gid (CRG) is used to evaluate your learning against a set of pre-defined criteria. |

| | |
|---|---|
| **11. Additional information (support, advice, tips etc)** | This assessment is an individually assessed component. Your work must be presented according to the Lincoln School of Computer Science guidelines for the presentation of assessed written work. Please make sure you have a clear understanding of the grading principles for this component as detailed in the accompanying Criterion Reference Grid. You are expected to take the following into account:<br>• Your submission should be a PDF file generated by one of the provided templates (in MS Word or Latex), which will be available via Blackboard under the materials of assessment item 1. This PDF file should be a concise report of **maximum 4 pages** including references. If you fail to stick to the length requirement or omitting submitting the source code of your solutions for both tasks, your submission of this assessment **will not be marked**.<br>• Do not submit the dataset provided via Blackboard.<br>• Do not submit all your models to avoid uploading and downloading really large files. Only submit the best model per task in addition to your source code.<br>• Do not submit Jupyter notebooks, do submit *.py files instead – and resource/config files (if used).<br>• Make sure that you submit **your own work** and not somebody else's writing or results. Failure to do so will incur plagiarism or collusion, which will the reported to the School for investigation of potential academic misconduct—see more on this below.<br><br>If you are unsure about any aspect of this assessment component, please seek advice with a member of the delivery team. |
| **12. Important Information on Dishonesty, Plagiarism and AI Tools** | University of Lincoln Regulations define plagiarism as *'the passing off of another person's thoughts, ideas, writings or images as one's own...'*. Examples of plagiarism include the unacknowledged use of another person's material whether in original or summary form. Plagiarism also includes the copying of another student's work'. Plagiarism is a serious offence and is treated by the University as a form of academic dishonesty. For more information on examples of Academic Offences, please see the **Academic Offence Guidance**.<br><br>Please note, if you use AI tools in the production of assessment work **where it is not permitted**, then it will be classed as an academic offence and treated by the University as a form of academic dishonesty.<br><br>Students are directed to the University Regulations for details of the procedures and penalties involved.<br><br>For further information, see www.plagiarism.org |