This development can be split into two main topics, the learning algorithm and the training environment.

The main aspects covered by this topic include the different learning algorithms, in more depth how each of the main ones work (DDPG and DQN). The learning algorithm topic includes not only the algorithms themselves but also the learning structure, such as a replay buffer used for storing multiple combinations of action, reward, state and next state, this helps to break harmful correlations and also learn from individual tuples multiple times making for better use of the experience. Another essential aspect of the learning algorithms is the learning framework, in this case, Keras. Keras, a high-level API of TensorFlow, is very helpful not only for this project and reinforcement learning but also any topic in machine learning. Understanding the capabilities and how to use the API is essential in artificial intelligence. Achieving the desired results in reinforcement learning is a complex and lengthy process. It requires extensive training and hyperparameter tuning. A large set of training sessions have been run to tune epsilon greedy, gamma, learning rate, or even the number of layers and neurons of the neural network.

The training environment is the second main topic in reinforcement learning, this includes training framework, simulation (physics engine in this case and visualizer) and data logging. After initial research on the topic, it became evident that openAI gym was the standard for reinforcement learning and became an obvious decision. After researching the already implemented environments, it became clear how it worked and helped decide on the implementation for the project's second stage. During the implementation of the second stage, it was required to set up a 2D humanoid using pymunk and pygame. Simulation was a new topic, therefore, it required more research, although the pymunk API integrates very well with pygame or even pyglet, making the implementation relatively straightforward. The last aspect of the training environment is logging, which is essential in reinforcement learning to allow comparison between different runs and hyperparameters, here the choice was wandb as it is integrated into Keras, while previous experience with similar platforms such as MLFlow was helpful to understand wandb, hands-on experience on logging for this development which includes more complex data was beneficial to understand the full capabilities of the platform better.