The initial goal of the project consisted of achieving walking for a bipedal robot, this was split into three steps of increasing complexity, the first step consisted of solving the cartpole environment, a classic reinforcement learning problem. The second step consisted of developing a 2D simulation of a simplified humanoid using the base structure developed for cartpole and the final step would be to develop a 3D simulation of the robots from the Bold Hearts team in Mujoco. At the current moment, the first two stages have been developed and the cartpole problem fully solved, the second stage is tested but a walking pattern has not been achieved yet. The third and final stage is currently being developed along with training for the second stage. While the many difficulties have been found along the way, many have been solved and became useful learning sources and documentation topics which is the reason of developing this project. Therefore, while a successful walking pattern might be hard to achieve given the constraints, the structure to achieve it and the base knowledge required are being acquired and documented.

Tasks in the development:

As prior mentioned, this project was split into three levels of complexity, each with unique tasks that would develop a ground basis for the next stage, increasingly progressing and further the understanding of the underlying technology.

The first stage consisted of the classic reinforcement learning problem, cartpole, here the main objective was to lay the ground of the development. The first task was to develop the learning algorithm, here two approaches where studied to better understand what should be used in the following steps, the first approach was using a high level layer of keras rl-agents, the second approach consisted of manually developing the learning structure using the base keras api. In this step it enabled me to fully understand how the learning process was working and what each component did, hence the reason to decide on using a manual approach for the following stages, given I had full control of the process. The second task was to setup a logging structure, using Keras I discovered the platform Wandb, where I started logging all aspects of each training session, the current code, hyperparameters, weights and renderings of specific checkpoints.

The second stage of development consisted of the development of the 2D simulation. In this stage the first major task was to decide on a physics and rendering engine, the decision was to use pygame and pymunk as both integrate very well and have large communities and documentation, this step allowed me to understand how physics engines and rendering work together and how to setup a semi-realistic simulation. The prior task revealed a problem as an incompatibility with the computer architecture lead to some problems with rendering under certain conditions. This prior problem lead to the research and comparison between local training and using a cloud service, in this case google colab. Google colab showed a very good performance without no losses to the local machine and enabled me to learn a valuable skill, running the training sessions on an offsite computer, including the rendering process and logging to an external service, including the use of secrets allowing me to expose the notebook without exposing credentials.

The third level consisted of the implementation of the currently used simmulation by the team Bold Hearts, implemented in gazebo to train walking using openai gym. The second task involves the re-adaptation of the learning algorithm to use the states from the 3d environment and the higher number of joints.

While the innitial target for the project was to get achieve walking on a simulation of the team's robot and possibly the transition to the real robot, givven the reduced time and the complexity of the project, allong with the many difficulties faced the main target of the project became the implementation of both the learning algorithm and the implementation of openai\_ros, the documentation of the results and the painpoints of development allong with documentation on how to setup this environment along with the setup of the teams environment on a new architecture (ARM64) which requires the creation of a new image and some workarrounds allows for other members to setup the environment if moved to the architecture.

The first level consited of the classic reinforcement learning problem, cartpole, where the objective is to balance a pole in a cart that can move in the horizontal axis.

This allowed to test the implementation of the learning algorithm in a simple problem it also allowed to test different implementations such as a high level implementation of keras agents and a lower level implementation of the learning algorithm, allowing for more controll and understanding of how each element worked.

Defining a structure of logging and code transferable to the next stages. This stage should also access the best training environment, including google colab/local.

The second level consisted of the development of a 2d simulation of a simplified humanoid robot, this involves the selection of an adequate physics simulator and a rendering engine. The second task is to setup the gym environment, such as step, reset, and the reward function. The third task involves adapting the learning algorithm and neural network to be able to control all of the joints from the robot simultaneously. The final task is hyperparameter search, reward function itteration and logging the data.

The third level consisted of the implementation of the currently used simmulation by the team Bold Hearts, implemented in gazebo to train walking using openai gym. The second task involves the re-adaptation of the learning algorithm to use the states from the 3d environment and the higher number of joints.

(expand how improved knowledge)

The innitial goal of the project consited of achieving walking for a bipedal robot, targeting the Bold Hearts team's robots. The project showed to be very complex for the time avilable, while getting a fully walking robot might not be achieved, the project still covers extensively the development process and all of its outcomes, including what painpoints where found and how the reward function affected the outcome, it also develops a 2D Gym environment and the integration of ROS and openai gym in order to use the gazebo simulator along with gym.