0.1 Plan

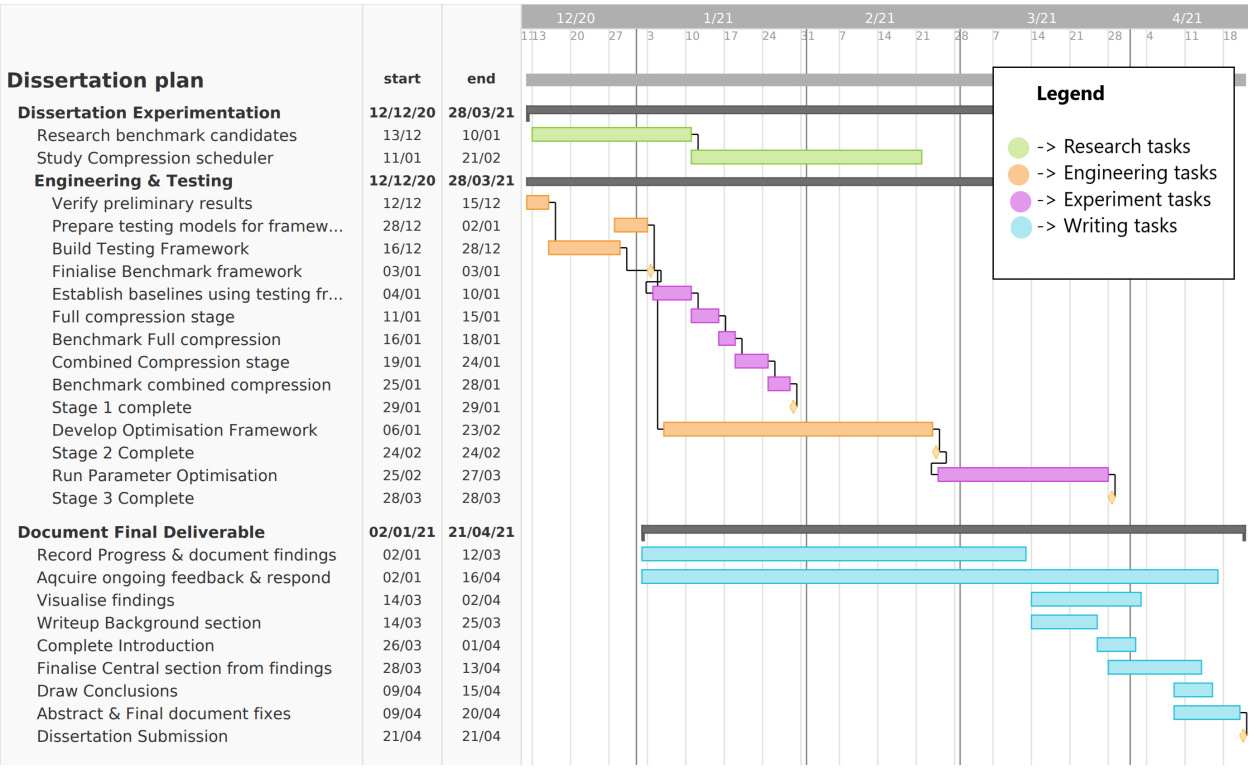

Figure 1: Project gantt chart

0.2 Risk Analysis

The following displays the anticipated risks with their accompanying mitigation strategies.

R1 Cannot compress model in Distiller (see Section **??**).

- Investigate alternative compression libraries.

R2 Issues with ONNX compatability.

- Investigate alternative intermediate representations.
- Manually export Distiller model to PyTorch and then directly to OpenVINO.

R3 Bottleneck on time to train models.

- Add additional local Training agents (see Section **??**).
- Scale the experimental model size down.

- Cloud training agents could be added.

R4 Bottleneck on time to benchmark inference.

- Investigate acquisition of a second NCS.

R5 Time management issues.

- Make use of the project plan to properly manage deadlines.
- Reduce development complexity (remove parallel training agents)

R6 Complexity of parameterising compression algorithm too high.

- Reduce scope of complexity by removing layer selection from the programatic definition and using a static layer definition.

R7 Hardware failure

- In the event of training agent hardware failure cloud resources can be accessed to perform training, the design plan is flexible to this.
- In the event of the NCS failing either a replacement will be sought out, or we could move inference off an edge environment and seek to continue latency optimisation there.

Likelihood	Near Certainty ∼90%			R1		
	Highly Likely ∼70%			R2		
	Likely ∼50%		R3	R5	R6	
	Low likelihood ∼30%					
	Not Likely ∼10%			R4	R7	
		Negligible	Minor	Moderate	Serious	Critical
		Impact of Non-Mitigated Risk				

0.3 Professional, Legal, Ethical, & Social issues

This dissertation will not use any participants in any experiments so there are no ethical or legal issues concerning the safety or well being of participants.

Care will be taken during the course of this dissertaion to maintain a professional standard of communication regarding this work with all contemporaries.

All software produced in the course of this work will be open source and licensed under the Apache License 2.0, this is in compliance with the existing Apache License 2.0 that is already in place on the Distiller and OpenVINO repositories if any modifications are made to the software of those respective libraries. Redis is licensed under the BSD license, Weights and Biases provides a free academic and open source license, this has been applied for and acquired.

The broader technological and social issues that this work may or may not play a part in are unknown, the field of machine learning is advancing at a tremendous pace (at the time of writing), and while it is vital to consider these implications they go far beyond the scope of this work. Any impact this dissertation will be just a miniscule part of a much larger technological revolution in which very few (if any) understand the full depth of the implications of their individual contributions.