Section	**Details**
Progress Overview (inkl. Reasons for Deviations of the Plan)	✔ Problem Statement ✍ Finalizing documentation of Literature Review ✍ Reproducing PASSION results – had to fix issues in current evaluation package – to reproduce an experiment on GPUHub takes 32.5h (6.5h x 5 folds) - and there are 4 experiments ✘ evaluate baseline with fairness metrics – Reproduction is not yet done
Accomplishments	✔ Categorization of the methods ✔ Got the evaluation running ✔ Progress on the report
Risks and Measures	Issues with interpreting the results • Clarifying the open questions The fine-tuning-experiments take to long • Improve checkpoint handling • Use another machine (afraid to lose information and canceled runs due to getting disconnected from GPUHub) Focus loss bc of tendency to want to fix / improve systems • Setting clear boundaries for myself: only do what must be fixed to enable running the mitigation experiments efficiently
Next Steps	• Get checkpoints for the relevant experiments • Evaluate the fairness for each subgroup on the experiment checkpoints
Discussion Points / Questions	• Have I understood the experiment result csv correctly? *(- Details in presentation)* • Do you have any tips on how to work efficiently on the GPU Hub? • How is the correct naming for the attributes in the PASSION dataset? Maybe I misunderstood the terms Features and Labels
Additional Notes	I was sick for roughly 4 days.
Attachments	• experiment_standard_split_conditions_passion.csv

Protocol	Expected to be shown next time
	• Have the fairness metrics ready
	• Know
	  – which biases should be targeted
	  – which mitigation measures will be tackled
	Performance improvement for training
	• Adjust parameters: Smaller model (Resnet18 / 34 should be sufficient, img size 256 or even smaller, ignoring folds/cross validation at the beginning)
	• GPUHub turned out to be not too much of a problem, since wanDB is used. Check if ssh connection is possible to get better editor support. Worst case: ask Jakob if there is a machine from the lab which can be used.
	Tips for evaluation
	• The values in evalTargets and evalPredictions should be in the same order like in the fold. However, since the fold is shuffled, it might be required to change the dataloader (code that creates the dataframe) to also keep track of which record is used OR the subject ID directly
	• Its also ok to rewrite the code from scratch using the dataloader classes.
	• In general, store all artefacts on wanDB and reload from there again.
	Scope Biases
	• Centers are not the focus of the thesis bc it is not a demographic bias. It might still be useful to run some experiments on it, as it could cause major biases due to image quality.
	Naming clarification
	• Features are the inputs for the model
	• Target labels are those which get predicted
	• The rest are attributes of the metadata