- Get access to MIMIC dataset (if not check Demo dataset) and be aware of the risks (measurements, units may differ) and check if the datasets come from same region/country.
- Train the models separately and then using GANS decide which is the best.
- Search for candidate datasets.
- Count the speed of sampling(?)
- For LLMs start with general knowledge
- Schedule meeting with supervisor after 19/04

REPORT

- Intro
- Problem Statement/ Research Questions
- Dataset
- Literature Review (search for studies based on LLM combined with GANS, how they combine them)
- Methodology/ Approach (List models, which GANS we use, try the recommended from Dr. Chang, check the tools on notes on google doc)
- Gant Chart
- Risk Analysis/
- We don't have to explain much on definitions for LLMs or what is synthetic data or GANS (max. 1-2 sentences), but how these models will work on our specific project (ex. conditional GANS)
- Compare the models with each other (similarities-dissimilarities, create a table to compare for presentation)

Phase 1 Presentation

-Intro
-Agenda
datasets
-Research Questions
-Dataset
-Models
-Risks?
-Gant Chart - Timetable

Phase 2 Expectations

- combine components
- try tests to produce small