# **Temporal Modelling of Clinical Data**

# Assigned to:

**Background:**

Digitally storing a patient’s multiple encounters with a healthcare system gives rise to a multivariate, longitudinal time series data. Most clinical studies - including AI in medicine, constrained by data limitations, fail to utilize temporal trends effectively for modelling. The goal of this project is to develop and evaluate a temporal modelling approach which can facilitate downstream ML tasks such as forecasting, classification and clustering.

**Problem Statement:**

Focusing on the most prevalent chronic diseases e.g. hypertension, diabetes, cancer, cholesterol etc, develop a model that uses time series EMR data to predict the future state of a patient: what are the next observations such as labs, vitals, and diagnoses likely to be.

**Tasks:**

* Collect and prepare relevant dataset from the Lakehouse. Understand the different fields and tables available and identify what data you want to use.
* Conduct a literature review to understand prior art on temporal modelling from EHR data. Some reading material will be attached.
* Explore and understand the different time series feature engineering and modelling techniques.
* Develop your models. Given an input sequence of events, the model should be able to predict the next entries in the sequence
* Experiment using your temporal modeling for different tasks to understand its efficacy e.g. instead of forecasting, what if we tried clustering similar patients, or predict the occurrence of some disease in the future?
* Evaluate your models using applicable metrics, analyze strengths and weaknesses of the trained model in perspective of real-world application.
* Report your performance and results on an unseen test set.
* Document the model's development process, training methodology, and evaluation results.

**Deliverables:**

* Trained model that can analyze a sequence of events for a patient, and predict the future state of the patient
* A visual tool to test the model on unseen patients and analyze results. The prediction as well as relevant explanations are desired
* A report detailing your literature review, model development and deployment steps, and analysis of the model’s behavior - important features and global explanations.

**Recommended Reading:**

* Various papers that can be useful for kicking off have been placed at \\dcshare\AISoft$\Ai Bootcamp projects\Temporal modeling

**Project Engineering Guidelines**

Here is a suggested research and development engineering cycle to follow:

* First do a thorough literature survey of the prior art related to the project and then shortlist the ones based on some justifiable rationale for this use case.
* Understand the models at depth including mathematics so you know what is happening inside the state-of-the-art method.
* Find suitable datasets or scientifically generate them for this use case.
* Build a software architecture diagram including class and sequence diagrams you were taught in OOP.
* Build your pipeline in real-time by using big data technologies if needed.
* Bonus would be awarded if you did model engineering with strong mathematics background instead of just doing API calls like programmers.
* Pick or create your testing and performance evaluation metrics and then measure them to show the inner working and behavior of your models on real-world scenarios. Moreover, some notion of accuracy and effectiveness of such systems need to be established.
* Add comments in the code for documentation purpose, code with no comments will be penalized heavily.