Final project

- **Project release date: 27th February**
- **Project due date: 17th March**
- **Submission mode:** via canvas upload.
  Please submit a typeset pdf and Google collab code link that I can access.

For the final project of the course, you will implement an experimental campaign to optimize for a property of interest for two materials design problems of your choice. Along with this project, you will be provided five different datasets from scholarly research published recently that aim to develop high-throughput experimental systems with closed-loop experimentation.

For any of the two selected material optimization campaigns, implement a simulated Bayesian experimental design workflow to optimize the properties of interest. Because you are given access to static data, the first goal (or evaluation metric) for this project is to build a regression model (call it $\mathcal{R}$)to accurately predict the target property. Using the model built above, create a simulated experimental system by adding an appropriate noise model to $\mathcal{R}$ such that every time you submit a query requesting the "ground truth" value from the experimental system, the value is slightly different from your model $\mathcal{R}$.

Implement an appropriate two-dimensional visualization system for the high-dimensional design space and use it to showcase how your experimental campaign evolves in the design space. Contrast it with at least a couple of choices for the acquisition functions.

Implement a random decision policy as a baseline for your experimental campaign and provide metrics to compare your Bayesian experimental design using the following definitions:

1. Enhancement factor : $\frac{\text{Top\% Bayesian}(i)}{\text{Top\% random}(i)}$
2. Acceleration factor (Top % = a) : $\frac{i_{\text{Bayesian}}}{i_{\text{random}}}$

Top % is defined as the ratio of the number of top candidates discovered to the number of top candidates possible in your static data, i corresponds to the iteration number, and a is a threshold value you can choose.

The five datasets provided for this project can be accessed from the following GitHub link:

`https://github.com/kiranvad/DOE/blob/master/project.ipynb`

The final column within each dataset is the value to be optimized for while the rest are the design variables.

Each dataset is associated with a research paper cited in the corresponding loading cell in the above notebook. Comment on if there is a particular scientific understanding you can obtain while performing this optimization task using the information provided in the paper.