# MM226: Materials Informatics
# Homework Assignment 2: AI-Generated Materials Data and Analysis

## Objective

Generate and analyze hypothetical materials data using AI tools to develop skills in data preprocessing, and exploratory analysis.

## Part 1: AI Data Generation

Based on the data collected by you in the first Assignment. Use ChatGPT, Google Gemini, Claude, and Perplexity to generate 60 data points each for your class of alloys including:

- Composition (chemical composition of alloy and the respective phases)
- Processing condition (e.g., Heat treatments at specific case, Test temperature. You have the metadata with you :))
- Microstructural parameters (grains size, phases, dislocation density)
- Mechanical properties with units (YS, UTS, Hardness, Elongation, strain rate, strain hardening exponent and coefficient)
- Request deliberate variability, missing values, and minor inconsistencies

## Part 2: Data Processing & Analysis

For each AI-generated dataset, Load the data and perform the following :

- Data Cleaning non-numeric entries and unit mismatches
- Handle missing values (e.g., mean imputation, forward-fill, row deletion). Justify the method (Hint: A little reading about the alloy system might be required).
- Compute mean, median, std deviation, min/max for all numeric fields. (as discussed in the class)
- Analyse the data and create visualization plots with interpretations. (choose relevant types: histograms, scatter plots, box plots, pair plots, etc.)
- Briefly interpret what each plot shows about the data.

# Part 3: Comparative Analysis

Compare the four AI-generated datasets based on the following criteria:

- Data plausibility and consistency - Are the values physically meaningful for your alloy system?
- Are the value ranges broad and realistic?
- what is the quality of introduced inconsistencies?
- Discovered relationships between properties

# Deliverables

Submit Report (One PDF and one Jupyter Notebook/Google Colab allowed) with:

- Your exact prompts for each AI.
- Data Cleaning methodology.
- Description of data cleansing steps.
- Comparative analysis of the four datasets. Also, plot the original data along with the AI dataset.
- Brief conclusion summarizing your insights on AI-generated data in materials science.

Everyone has to submit the assignment individually!
**Deadline: April 13, 2025**