# Build LLM: Research Based Assessment

**Deadline:** 7 May 2025
**Maximum Team Size:** 3 members

## Objective

The objective of this assignment is to deepen your understanding of Large Language Model (LLM) architectures by training them from scratch and observing the impact of various design decisions. This hands-on assignment will also help you develop insights into training dynamics by modifying hyperparameters and architectural components.

## Codefile

- **Colab Notebook:** Open in Google Colab

## Instructions

1. **Dataset Selection**
    - Choose and upload any book dataset from Project Gutenberg. You can also upload multiple books into one file.
2. **Baseline Training**
    - Train the LLM on your chosen dataset.
    - Log training and validation loss.
    - Report the baseline training results.
3. **Experiment 1: Varying Epochs and Learning Rate**
    - Train your model with different epochs (e.g., 5, 10, 20, 50, 100, 200, 500 etc) and learning rates (e.g., 1e-4, 1e-3, 1e-2).
    - Plot training and validation loss for each configuration.
    - Provide a comparative analysis and inference from your plots.

4. **Experiment 2: Varying Transformer Layers**
    - Train models with different numbers of transformer layers: 1, 3, 5, 7, and 12.
    - Plot and compare training and validation loss.
    - Explain your inference based on architectural depth.
5. **Experiment 3: Varying Attention Heads**
    - Train models with different numbers of attention heads (e.g., 1, 2, 4, 8).
    - Plot and analyze loss metrics.
    - Discuss how the number of attention heads affects model learning.
6. **Experiment 4: Ablation Studies**
    - **Remove Normalization Layers in all transformer layers:** Plot and analyze training and validation loss. Comment on inference quality.
    - **Remove Shortcut (Residual) Connections in all transformer layers:** Plot and analyze training and validation loss. Comment on inference quality.
    - **Remove Feedforward Neural Network in all transformer layers:** Plot and analyze training and validation loss. Comment on inference quality.
    - Provide clear explanations of the role of each component based on your observations.

# Submission Guidelines

- Submit a compiled PDF via Overleaf.
- Include all code as an appendix or provide a GitHub link.
- Each team must provide a 1-page summary of key findings.

# Evaluation Criteria

- Correct implementation and experimentation (30%)
- Clarity and depth of analysis (30%)
- Quality of visualizations and plots (20%)
- Presentation and formatting (10%)
- Summary insights and reflection (10%)

*Good luck, and may your models converge beautifully!*