

# Final Project

Instructor: Mork Mongkul

Deadline: [ ]

## PROJECT CORE: MODEL TRAINING AND DEPLOYMENT

---

In this project, you will collect your own dataset and train a machine learning model for one of the following tasks: **Regression, Classification or Clustering, Object Detection**, or a **Chatbot**.

Once the model is trained and evaluated, you are required to deploy it as a functional application using **Streamlit** or a **web-based interface** with a dedicated **API backend** and **frontend**.

## GRADING AND RUBRICS

---

This is a final project for this AI-Bootcamp course. Each student is expected to independently complete and submit all required components. Individual understanding of the implemented concepts is required for a passing grade.

You are required to complete one of the project tasks listed below and submit the following:

1. A **presentation slide** clearly detailing your problem statement, dataset, methodology, model architecture, and results.
2. A **Google Colab notebook** containing clean, commented, and reproducible code covering data preprocessing, model building, training, and evaluation.
3. A **GitHub repository** hosting your full project, including the source code for your deployed application with a proper `README.md`.
4. A **deployed application** accessible via Streamlit or a web-based interface with an API backend and frontend.

**Presentation:** Each student will deliver a 15-minute individual presentation and 5 minutes for questions and answers.

*The grades for the Notebook, GitHub, Deployment, and Presentation constitute your overall Project score.*

## PROJECT OPTIONS

---

Below are the descriptions for the available term projects. The goal is to explore the capabilities of learned representations, understand how much can be achieved with the resources available, and identify the specific limitations of your chosen approach. You are encouraged to discuss your methodology with the instructor as you progress.

### **Project 1: Energy Consumption Prediction**

Energy management is crucial for sustainability and cost optimization. The objective of this project is to build a regression model that predicts energy consumption based on historical data and relevant features such as weather conditions, time of day, seasonality, and building characteristics. The system should provide accurate forecasts to enable better resource planning and energy efficiency strategies.

### **Project 2: Condo Price Prediction**

Real estate valuation is a complex task influenced by multiple factors. This project focuses on developing a regression model to predict condominium prices based on features such as location, size, number of bedrooms, amenities, proximity to public transportation, and market trends. The goal is to create a reliable pricing tool that can assist buyers, sellers, and real estate professionals in making informed decisions.

### **Project 3: Malicious URL Detection**

Cybersecurity threats often begin with malicious URLs used in phishing attacks, malware distribution, and social engineering. The objective of this project is to build a classification model that can identify whether a given URL is malicious or benign. The system should analyze URL structure, domain reputation, and other features to protect users from potential security threats.

### **Project 4: Khmer Next Word Prediction**

Language modeling for low-resource languages presents unique challenges. This project aims to develop a model capable of predicting the next word in a Khmer sentence given the preceding context. The system should learn from Khmer text data to capture linguistic patterns and provide accurate suggestions, which can be useful for applications such as text completion and assisted writing.

**Project 5: Khmer Optical Character Recognition (OCR)**

Digitizing printed or handwritten Khmer text is essential for document processing and accessibility. The objective of this project is to create an OCR system that can accurately recognize and convert Khmer characters from images into machine-readable text. The system should handle various fonts, text sizes, and image qualities while maintaining high recognition accuracy.

**Project 6: Khmer Legal Compliance Chatbot**

Navigating legal regulations and compliance requirements can be challenging for businesses and individuals. The objective of this project is to develop an intelligent chatbot that can answer questions about Khmer legal matters, interpret regulations, and provide guidance on compliance issues. The system should be trained on legal documents, statutes, and regulations to offer accurate and contextually relevant responses, making legal information more accessible to users.