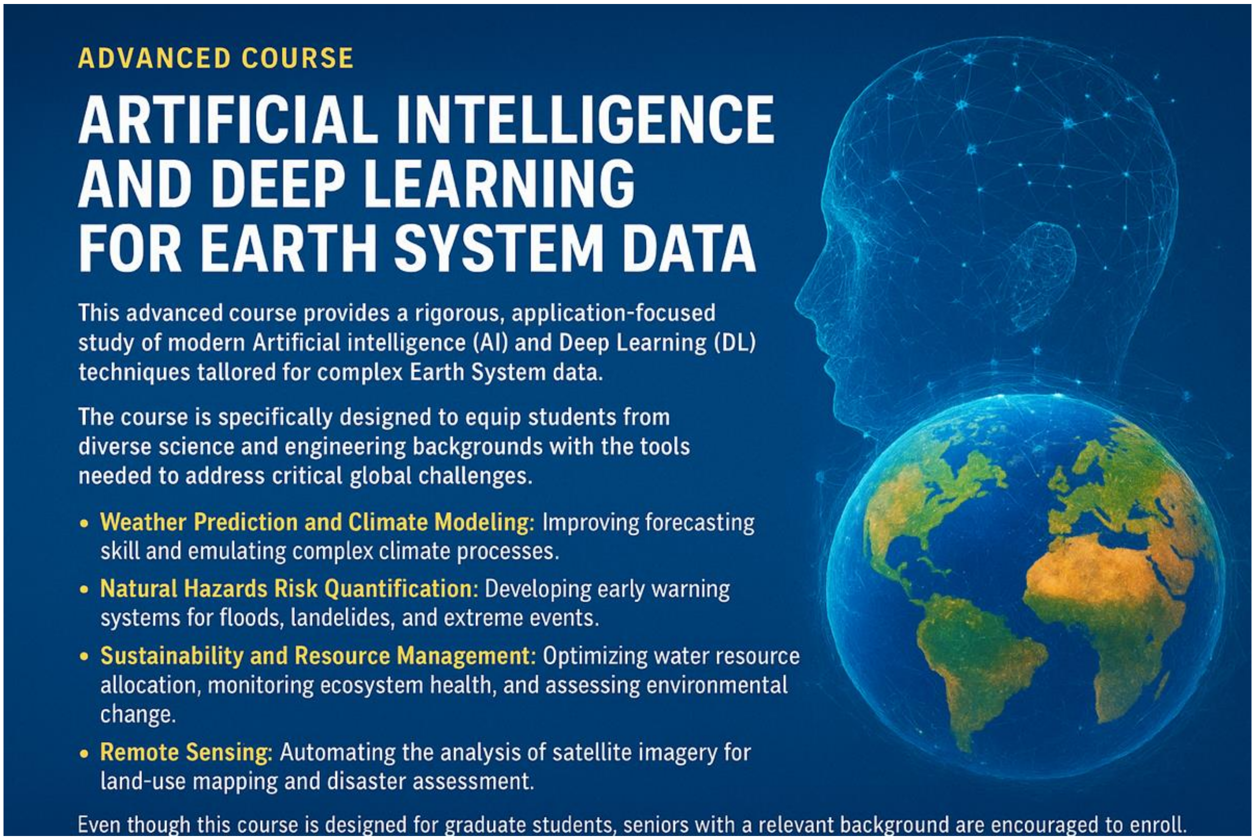
Course Syllabus: AI for Earth System Science and Engineering



Course Description

This advanced course provides a rigorous, application-focused study of modern Artificial Intelligence (AI) and Deep Learning (DL) techniques tailored for complex Earth System data. The course is specifically designed to equip students from diverse science and engineering backgrounds with the tools needed to address critical global challenges.

The curriculum covers foundational ML models like Weighted Least Squares (WLS), Logistic Regression, Support Vector Machines (SVM), and Tree-Based Methods, before transitioning to an intensive study of cutting-edge DL architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and LSTMs, Graph Neural Networks (GNNs), and Transformer Models.

The core focus is on real-world applications in Earth System Science and Engineering, including:

- Weather Prediction and Climate Modeling: Improving forecasting skill and emulating complex climate processes.
- Natural Hazards Risk Quantification: Developing early warning systems for floods, landslides, and extreme events.

- Sustainability and Resource Management: Optimizing water resource allocation, monitoring ecosystem health, and assessing environmental change.
- Remote Sensing: Automating the analysis of satellite imagery for land-use mapping and disaster assessment.

Even though this course is designed for graduate students, seniors with a relevant background are encouraged to enroll.

Learning Objectives

Upon successful completion of this course, students will be able to:

- Identify, structure, and pre-process various types of Earth System Data for AI applications.
- Implement and contrast foundational ML models, including WLS, Logistic Regression, SVM, and Tree-Based Methods with kernels.
- Design, train, and evaluate complex Deep Learning architectures (CNNs, LSTMs, GNNs) on spatial and temporal Earth System datasets using Python.
- Understand and apply state-of-the-art models like the Transformer and Generative Models.
- Complete an independent research project that applies advanced AI methods to a real-world Earth System problem.

Prerequisites

- Basic understanding of statistics, calculus, and linear algebra.
- Familiarity with high-level programming languages such as R/Python/Matlab and common scientific libraries (e.g., NumPy, Pandas, Pytorch).

Grading & Evaluation

Component	Weight
Homework	40%
Class Participation	10%
Final Project Proposal	10%
Final Project Report & Presentation	40%
Total	100%

Final Project: The final project is the cornerstone of the course. Students will work individually or in pairs to identify an earth science problem and use AI to address it. A Project Proposal is due at the end of Week 7.

Proposed Topics:

Part 1: Foundations of Earth System Data & Machine Learning (Weeks 1-4)

This section establishes the data landscape of earth science and covers essential, interpretable machine learning models used for both regression and classification.

- **Introduction and Data Foundation**
  - Overview of the field: Why AI/ML/DL are essential for Earth System Science and Engineering.
  - **Earth System Data**: sources (satellite, models, in-situ), structure, pre-processing challenges.
  - **Python** Environment Setup: Introduction to core libraries (Numpy, Pandas) and platform setup (**Kaggle** intro).
- **Supervised Learning and Regression**
  - Formal definition of the supervised learning setup.
  - **Weighted Least Squares (WLS)**: Handling heteroscedasticity and non-uniform data quality in regression.
- **Classification and Non-Linearity**
  - **Logistic Regression**: Transitioning from linear regression to basic classification tasks.
  - **Kernels**: The concept of non-linearity and implicitly mapping data into higher-dimensional space for linear separation.
- **Model-Based and Non-Parametric Methods**
  - **Support Vector Machines (SVM)**: Maximizing the margin for robust classification.
  - **Tree-Based Methods**: **Decision Trees** and the introduction to **Random Forests**.
  - Unsupervised Learning: **K-Means** for clustering and pattern discovery.

Part 2: Deep Learning (Weeks 5-10)

This core, seven-week block covers the fundamental theory and main architectures of deep learning, focusing on networks designed for image, sequence, graph, and generative tasks.

- **Deep Learning Fundamentals**
  - The structure of **Artificial Neural Networks (ANNs)** and activation functions.
  - Optimization: **Gradient Descent**, Backpropagation, and advanced optimizers.
- **Convolutional Neural Networks (CNNs) for Spatial Data**
  - **CNN** Architecture: Convolutional layers, feature maps, and parameter sharing.
  - Advanced techniques: The **U-Net** for image segmentation (e.g., cloud detection, land cover mapping).
  - Transfer Learning and data augmentation.
- **Recurrent and Sequence Models**
  - **Recurrent Neural Networks (RNNs)** and the challenge of sequential data.
  - **Long Short-Term Memory (LSTM)** networks and their gated architecture for time series forecasting (e.g., river flow).
- **Graph and Generative Models**

- **Graph Neural Networks (GNNs)**: Modeling interconnected Earth Systems (e.g., river networks).
- **Generative AI**: Introduction to **Generative Adversarial Networks (GANs)** and **Diffusion Models**.
- Applications of generative models: high-resolution data downscaling and gap filling.

Part 3: Advanced Applications & The Final Project

This section focuses on cutting-edge architectures and real-world applications relevant to engineering and science problems, including foundation models and ethical considerations.

- **Advanced Sequence Modeling**
  - The **Transformer** Architecture: Removing recurrence using the self-**Attention mechanism**.
  - Application of Transformers to long-range spatiotemporal prediction in climate and earth science.
- **Foundation Models in Practice**
  - The concept of **Geospatial** and **Weather Foundation Models** (e.g., FourCastNet, pre-trained models).
  - Leveraging large models for zero-shot/few-shot learning and model emulation.
- **Societal Impact & Ethics**
  - Integrated case studies: **Natural Hazard Risk Quantification** (floods, landslides) and early warning systems.
  - AI for **Sustainability**, water resource management, and ecosystem health monitoring.
  - **Ethics of AI in Earth Science**: addressing bias, equity, and model interpretability.

Resources:

An Introduction to Statistical Learning with Applications in Python

100 Page Python Intro

Machine Learning Basics: Machine Learning from Scratch; Stanford CS229 – Machine Learning

Deep Learning: Understanding Deep Learning; Deep Learning with PyTorch; Practical Deep Learning for Coders; Stanford CS230; Stanford CS231n – Deep Learning with Computer Vision