Artificial Intelligence For Science

Learning Objectives:

- Understand the foundational principles of artificial intelligence.
- Explore the intersection of AI with various scientific disciplines.
- Recognize the potential and limitations of AI in scientific research and applications.
- Develop hands-on experience with simple AI tools and projects.
- Cultivate a critical mindset towards the ethical considerations and consequences of AI in science.

6-Week Curriculum:

Week 1a: Introduction to Machine Learning
HW: Watch one of the videos on python for beginners

- Syllabus Intro
- What is AI? Definition, what it isn't
- AI vs. Human Intelligence – rationality, optimality, neuroscience
- History of AI. – Expert models, step function, winter, sigmoid activation, timeline
- Where does Machine Learning fit in AI? – Subset, Statistics, Agents
- Types of Learning and Objectives
    - Learning
        - Supervised
        - Unsupervised
        - Semi-Supervised
        - Self-Supervised
        - Reinforcement Learning
        - Few Shot Learning
    - Objectives
        - AI - Understanding
        - AI - Learning
        - AI - Decision Making
        - Problem Class - Regression
        - Problem Class - Classification
        - Problem Class - Clustering
        - Problem Class – Association
        - Problem Class – Dynamic Action
        - Generative vs Discriminative
        - Probabilistic vs Deterministic

Week 1b: Basic Principles of Machine Learning
HW: **Project – Part 1:**
-Problem statement
-Why/Why not machine learning

- Training process (30-40 minutes)
    - Training vs Inference
    - Training, Testing, Validation Data
    - Model Parameters
    - Cost Function
    - Backpropagation – Brief Intro
    - Auto-differentiation – Brief Intro
    - Transfer Learning and Fine-Tuning - Concept Intro

Each case gets [What, Example Usage, Pros, Cons] (40-50 minutes)

- Non-Neural Network Models
    - Linear and Multivariate Linear Regression
    - Logistic Regression
    - Support Vector Machines
    - K-Nearest Neighbor
    - K-Means
    - Decision Trees/ Random Forest
    - Gradient Boosting
    - Hidden Markov Models
    - Bayesian Networks/ Naïve Bayes Algorithm

Week 2a: Neural Networks Overview
HW: Term Matching

- Basics (20-30 minutes)
    - Feed Forward
    - Fully Connected or Not
    - Activation Functions
    - Weights and Biases

Each case gets [What, Example Usage, Pros, Cons] (60-70 minutes)

- Simple Neural Nets
- Recurrent Neural Nets
- Long-Short Term Memory
- Gated Recurrent Unit
- Auto Encoder / Variational Auto Encoder (Encoder/Latent Space/Decoder)
- Restricted Boltzmann Machine
- Convolutional Neural Network
- Generative Adversarial Neural Network
- Attention – Transformer Lite
- Multi-Modal or Modular Neural Network

Week 2b: Building and understanding basic feed-forward networks

HW: Network Labeling, Pro and Con Network Matching

- Coding a basic Machine Learning model in Python with ML libraries
- Importance of data selection, generalization, and training procedures
    - Extrapolation versus Interpolation
    - Data coverage across train/test/validation sets
    - Features of a function – Overfitting/underfitting/generalization
    - L1 and L2 regularization
    - Cross Entropy
    - Momentum
    - Dropout

Week 3a: Limitations and Challenges of AI in Scientific Research

HW: [Problem/Solution Matching

Project Part 2:

- What/how much data is needed

-What data preprocessing needs to be done

-What concerns about the data quality exist and how would you address them.]

- Data quality
- Input representation impacts
- Model structure impacts
- Uncertainty in model output
- Understandability of models
- Impact of hallucinations
- Current understanding bias
- The importance of domain knowledge.
    - Required accuracy for impact
    - What data is obtainable/trustworthy
    - Determined mathematical relationships
    - Assumptions that limit the model's reality

Week 3b: AI/ML Accessories

HW: Systematically test an LLM of your choice on its ability to teach you how to do a 1) skill you don't know well 2) complex skill you know well

- Data Preprocessing – Images, Principal Component Analysis, Removing outliers
- Active Learning
- Compute power
- Retrieval Augmented Generation (RAG)
- LLMs

Week 4a: AI in Physical Sciences

HW: [Describe how you would structure an ML study for a physical sciences problem

- I/O and why they should be functionally connected?
- How the data was or could be collected and any assumptions around the quality of the data?
- What type of ML problem are you solving?
- What model would you use and why?

Project Part 3:

-Describe what models you would try and why

-Describe what you might do to the model if:

--The model is underfit

--The model is overfit]

- AI in Physics: Simulations and predictions.
- AI in Chemistry: Materials, spectra, and synthesis
- Demos of one technology (If time available) (Pro-Con why it's impactful)

Week 4b: AI in Medicine

HW: Describe how you would structure an ML study for a given medicine problem

- I/O and why they should be functionally connected?
- How the data was or could be collected and any assumptions around the quality of the data?
- What type of ML problem are you solving?
- What model would you use and why?

- AI in drug discovery and molecular dynamics.
- Predictive modeling in genetics.
- AI in medical imaging and diagnosis.
- Demos of one technology (If time available) (Pro-Con why it's impactful)

Week 5a: AI in Space Exploration and Astronomy

HW:

[**Project Part 4:**

-Outline the limitations of the model assuming it works

-Describe how you could extend the model to new data spaces.]

- AI in satellite image analysis.
- Predicting celestial events.

Week 5b: AI Agents for Robotic Action

HW: Multiple choice questions, one short open-ended question on how robotics could be incorporated into a scientific problem area and why it would be suitable.

- What is an agent and what models guide agents
- Basics of robotics.
- The realities of online, real-time, embedded systems

- Current state of AI robots

Week 6a: Ethical Considerations of AI in Science
HW: [T/F "Is this a possible ethics problem?" scenarios
Project Part 5:
Make a video describing your proposed project.]

- Ethics 101
- Ethical concerns and examples
    - Medicine
        - Misunderstanding the problem
        - Accountability – Medical malpractice
        - Availability vs Quality Gaps (real vs AI doctor depending on what the patient can pay)
        - Human bias in data – Interpretation of patient responses
    - Data Usage
        - Informed consent (data usage and model result usage on person)
        - HIPAA
    - Robot/AI replacement of workers
    - Incorrect model decision impacts
    - Fake data
    - The interface of AI, robotics, and human interaction
    - Human-in-the-loop

Week 6b: The Future of AI in Science
HW: Course Survey

- Quantum computing and AI.
- The potential of AI in unifying scientific theories.
- Reflect on the potential and challenges of integrating AI into future scientific endeavors.
- New technology enabling faster calculations/training.
- The Data Cliff

Homework Plan

Curriculum Week	**6**
1a	Watch one of the videos on python for beginners
1b	**Project – Part 1:** -Problem statement -Why/Why not machine learning
2a	Term Matching
2b	Network Labeling, Pro and Con Network Matching

3a	Problem/Solution Matching **Project Part 2:** - What/how much data is needed -What data preprocessing needs to be done -What concerns about the data quality exist and how would you address them.
3b	Systematically test an LLM of your choice on its ability to teach you how to do a 1) skill you don't know well 2) complex skill you know well
4a	Describe how you would structure an ML study for a physical sciences problem **Project Part 3:** -Describe what models you would try and why -Describe what you might do to the model if: --The model is underfit --The model is overfit
4b	Describe how you would structure an ML study for a given medicine problem
5a	**Project Part 4:** -Outline the limitations of the model assuming it works -Describe how you could extend the model to new data spaces.
5b	Multiple choice questions, one short open-ended question on how robotics could be incorporated into a scientific problem area and why it would be suitable.
6a	T/F "Is this a possible ethics problem?" scenarios **Project Part 5:** Make a video describing your proposed project.
6b	Course Survey

Project Details – 6 weeks

- The purpose of this project is to show you how to plan out a machine learning study.
- Throughout the program you will have multiple checkpoint assignments due
- At the end of the 6 weeks you must submit a 2-5 minute video presenting how the ML study would be conducted.

Quizzes

- Cover the previous week's material
- 5 minutes each
- 5 multiple choice or fill in the blank questions each
- Grades delivered next class (will always be the (a) class of the following week) via student id list – score sheet
- Each quiz gets two chances – first time on the (b) class of the week, 2nd time on the (a) class of the following week. – Same quiz
- Best of two quiz grades kept
- Must receive an 80% overall on quizzes to get completion

Homework (HW)

- Given after each class and due by the start of the next class
- Must submit at least 90% HW's for completion

Attendance

- Students may not miss more than 4 lectures without proof of extenuating circumstance to qualify for completion.
- Students should inform the professor as soon as possible, or within two business days (whichever is justifiably greater) of extenuating circumstances they wish to be excused and to receive make-up work.
- Each unexcused lecture (up to 4) missed will subtract 2.5% from the student's total uncurved score.

Grades

- Grades in quizzes(30%), homework(30%), attendance(10%), and the project(30%) contribute to the overall completion grade.
- The total grade must be greater than 75% for completion.
- Grades may be adjusted up on a curve at the end of the course but will not be adjusted down.

Additional Help

Students needing additional help should submit questions through the Google Classroom portal.

Students should expect responses from emailed questions within 24 hours and therefore may or may not receive a response within one day of an assignment due date.