Chemistry 352/452 – Artificial Intelligence in Chemistry

Welcome to Chemistry 352/452! Artificial intelligence (AI) is rapidly becoming an essential part of modern chemical research. Methods developed in AI and machine learning are now enabling chemists to predict molecular properties, model chemical reactivity, accelerate materials discovery, and gain new insights from simulations and experimental data. From protein structure prediction and catalyst design to automated synthesis planning and spectral interpretation, AI is redefining the questions we can ask and the answers we can obtain (in chemistry).

This course will introduce students to the fundamental concepts of AI and machine learning, with an emphasis on their application to real-world problems in chemistry. Students will study core techniques such as supervised and unsupervised learning, neural networks, and generative models, while gaining hands-on experience applying these methods to tasks like molecular property prediction, reaction outcome modeling, and experimental data analysis. By combining theoretical foundations with practical implementation, the course will prepare students to critically evaluate current research, select and apply appropriate AI methods, and integrate these approaches into their own chemical studies and research.

Instructors: Yihang Wang, Ph.D.

Office: Millis G27A
Email: yxw2626@case.edu

Course Prerequisites: CHEM 105, CHEM 106, CHEM 111, ENGR 145, MATH 124, and MATH 201, or their equivalents, are recommended.

Recommended Texts:

Deep Learning for Molecules and Materials by Andrew D. White
Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
Pattern Recognition and Machine Learning by Christopher M. Bishop
Machine Learning with PyTorch and Scikit-Learn by Sebastian Raschka

Learning Objectives: By the end of this course, students will be able to: (1) explain the fundamental principles of artificial intelligence and machine learning as they apply to chemistry; (2) implement and train basic models using common computational frameworks; (3) apply supervised, unsupervised, and generative methods to problems such as molecular property prediction, reaction outcome modeling, and analysis of chemical data; (4) critically read and evaluate current research literature in AI for chemistry; and (5) develop problem-solving strategies that integrate AI methods into their own chemical studies and research.

Tentative Topics:
The course will begin with a review of the essential mathematical and computational foundations for machine learning, including linear algebra, statistics, and optimization. Building on this background, we will explore core machine learning methods relevant to chemistry: featurizing molecular data, unsupervised learning for pattern recognition and dimensionality reduction, and supervised learning for

property and reactivity prediction. We will then examine neural networks in detail, covering both fundamental principles and modern architectures such as convolutional and graph neural networks. Advanced topics will include generative models and large language models, with an emphasis on their application to chemical modeling. The second part of the course will focus on frontier applications, including AI for synthesis planning, protein structure prediction, molecular dynamics and simulation analysis, and the development of machine learning–based force fields.

Reading assignments (RAs): Prior to most lectures, a reading assignment will be posted to Canvas. You need to do the reading and complete a short assignment. The use of large language models (LLMs) will be specified in the assignments; all work must reflect your own understanding and effort.

Homework assignments (HWs):
Homework assignments will provide hands-on experience with methods. The use of LLMs is not encouraged unless specific. Discussion among classmates is allowed, but each student must complete and submit their own work. Late homework assignments will be accepted up to 72 hours after the due date, with a penalty of 20% per day.

Final Projects:
There will be individual projects and team projects with up to three students per team. Further details will be released later.
Students enrolled in CHEM 452 will be required to develop a machine learning project of their choice.
Students enrolled in CHEM 352 may participate in a project for bonus credit.

Grades:
All in-class quizzes will be counted as bonus credit and added to the final grade.
RAs are graded based on effort, with a maximum of 5 points per assignment. The total number of RAs may change as the course develops. Only the highest six RA scores will be counted toward the final grade. Due by 8:00 AM on the day of the lecture.
HWs consist of five assignments, each worth 100 points. Only the highest four homework scores will be counted toward the final grade. Due by 8:00 AM on the day of the lecture.

Grades will be calculated with the following weighting:
CHEM 352: 40% reading assignment questions, 60% homework assignments, and up to 20% bonus credit for final projects.
CHEM 452: 30% reading assignment questions, 45% homework assignments, and 25% for final projects.

Academic Honesty: Students at Case Western Reserve University are expected to uphold the highest ethical standards of academic conduct. Please review the complete academic integrity policy. Any violation of the policy will be reported to the Dean of Undergraduate Studies and the Office of Student Conduct & Community Standards.

Disabilities: In accordance with federal law, if you have a documented disability, you may be eligible to request accommodations from Disability Resources. In order to be considered for accommodations, you must first register with the Disability Resources office. Please contact their office to register at 216.368.5230 or get more information on how to begin the process. Please keep in mind that accommodations are not retroactive.

Course Schedule:

Week		Topic	Task	
Week1	Jan 12	Course overview and syllabus Introduction to Python programming		
	Jan 14	Mathematical foundations for machine learning Introduction to supervised learning	RA1 due	RA 1: Mathematical foundations I —matrix inversion
Week2	Jan 19	Supervised learning methods (regression and classification)	HW1 out	HW1: Python programming
	Jan 21	Introduction to neural networks and backpropagation	RA2 due	RA 2: Mathematical foundations II —gradients and derivatives
Week3	Jan 26	Neural network classifiers	HW2 out	HW2: First steps with neural networks
	Jan 28	Unsupervised learning Dimensionality reduction methods	RA3 due	RA 3: Building your first neural network
Week4	Feb 2	Molecular clustering and similarity analysis	HW 1 due	
	Feb 4	Unsupervised learning with neural networks representation learning		
Week5	Feb 9	Variational Autoencoder	HW2 due	
	Feb 12	Dimensionality reduction and latent spaces for molecular data		
Week6	Feb 16	Advanced network architectures I: Molecular representations: descriptors, fingerprints, and graphs	HW3 out RA4 due	RA 4: Permutation and translation invariance HW3: Deep neural network
	Feb 18	Multilayer perceptrons (MLPs) and convolutional neural networks (CNNs)		
Week7	Feb 23	Graph representations of molecules Graph convolution and message passing		
	Feb 25	Graph Neural Networks (GNNs) for chemistry applications		
Week8	Mar 2	Advanced Network Architectures II: Modeling time series data, RNNs and LSTMs	HW3 due RA5 due	RA 5: Modeling long sequences and memory
	Mar 4	Transformers and attention mechanisms		
Week9	Mar 9	Sprig break		
	Mar 11	Sprig break		
Week10	Mar 16	Vibe Coding: AI-assisted programming	Final project topic due	
	Mar 18	Generative AI I: GANs and VAEs	HW4 out	HW4: Molecular generation with generative models
Week11	Mar 23	Workflow automation tools and AI agents		

	Mar 25	Generative AI II: Normalizing flows and diffusion models	RA6 due	RA6: Operations on Gaussian distributions
Week12	Mar 30	Diffusion models with GNNs: Molecular ensemble generation	HW5 out	HW5: Processing and analyzing chemical data
	Apr 1	Guided diffusion and molecular design		
Week13	Apr 6	Predicting protein structure: AlphaFold	HW4 due	
	Apr 8	Protein large language models, model pretraining		
Week14	Apr 13	selective topic I		
	Apr 15	selective topic II	HW5 due	
Week15	Apr 20	Project project Presentations		
	Apr 22	Project project Presentations		
Week16	Apr 27	Final project presentations and Q&A		
	Apr 29	Final exam week with no final exam		

This syllabus is subject to change at the discretion of the instructor. Any modifications will be communicated promptly through Canvas and during class.