Indian Institute of Technology, Kanpur
Proposal for New Course

1. **Course No:** CHM 69**X**
2. **Course Title:** AI/ML in Chemistry
3. **Per Week Lectures:** 3 (L), **Tutorial:** 0 (T); **Laboratory:** 0 (P), **Additional Hours** [0-2]: 0 (A),
   Credits (3*L+2*T+P+A): 9
   Duration of Course: Full Semester
4. **Proposing Department/IDP:** Chemistry
   Other Departments/IDPs which may be interested in the proposed course: Physics, Biological Sciences and Bio Engineering (BSBE), Chemical Engineering, Material Sciences and Engineering
   Other faculty members interested in teaching the proposed course: Prof. Amalendu Chandra (CHM), Prof. D. Goswami (CHM)
5. **Proposing Instructor(s):** Prof. Mark Tuckerman (Currently Dist. Visiting Professor at IIT Kanpur and NYU New York), Prof. Nisanth N. Nair (CHM)
6. **Course Description:**
   A. **Objectives:** The course is intended to teach the basics of AI and ML techniques that are useful for addressing various research problems in the area of chemistry. The course will be an elective course, open to BS/MS, MSc, and PhD students in Chemistry. BTech/BS/ MTech/MS/PhD students from other departments may also be permitted who have interests in using the techniques in the area of chemistry or overlapping areas with chemical applications (such as biology, physics, chemical engineering, materials science)
   B. Contents (*preferably in the form of 5 to 10 broad titles)*:

S. No.	Broad Title	Topics	No. of lectures
1.	Introduction	• Uses and impact of AI in Chemistry: An overview. • Fundamental tasks of machine learning: Regression, classification, clustering. • Mathematical preliminaries: Crash courses in multivariable calculus, linear algebra, and graph theory. • Basics of Python programming. • Regression and classification tasks. • Some elementary machine learning models: Gaussian processes, feed-forward neural networks, tree models and random forests.	12

2.	Machine learning and the electronic structure problem:	• Machine learning interatomic potentials (MLIPs)
		  ○ Behler-Parrinello networks.
		  ○ Deep graph neural networks.
		  ○ Atomic cluster expansion (ACE) and MACE architecture.
		• Learning the electron density.
			5

3.	Machine learning for chemical compound design:	• Defining chemical space.
		• The problem of representing molecules for machines.
		• Machine learning architectures and strategies.
		  ○ Recurrent neural networks.
		  ○ Generative adversarial networks.
		  ○ Autoencoders and variational autoencoders.
		• Sequential graph generation.
		• 3D Shape generation.
		• Point clouds.
		• Molecular property prediction.
			8

	Machine learning for materials characterization and design:	• Atomic versus molecular materials
		• The problem of representing bulk crystals for machines.
		• Machine learning architectures and strategies extended to periodic structures.
		  ○ Deep graph neural networks.
		  ○ Generative adversarial networks.
		  ○ Autoencoders and variational autoencoders.
		• Predicting crystal structures and materials properties.
		• Models and strategies for materials discovery.
		  ○ Latent spaces and 3D structure searching and generation.
		  ○ Use of points clouds.
			8

	Brief overview of other models and their uses in other areas of chemistry (Instructors can adjust the topics based on the state-of-the-art areas which he/she wishes to cover at the time of teaching)	• Large language models and transformer networks. • Flow networks, generative flow networks, and diffusion models. • ML and its uses in molecular simulation. • ML and its uses in molecular spectroscopy or protein structure prediction	7

C. **Pre-requisites, if any:** None specifically; However, instructors may check that the student has sufficient mathematical background to understand the topics.

D. **Short summary for including in the Courses of Study Booklet:** This is an introductory course on artificial intelligence and machine learning tools for chemists. The course briefly goes through the basics of machine learning, mathematical background, and python-based coding in the perspective of AI/ML. Some of the major applications of AI/ML in chemistry are introduced, which include representation of potential energy of molecular systems as a substitute for electronic structure theory, machine learning for molecular design, and materials design and characterization. Some recent applications using large language models, flow and generative networks, and molecular simulations will be presented.

7. **Recommended Books:**

   - Bishop, Christopher M. Pattern Recognition and Machine Learning. New York :Springer, 2006.
   - Y. Wang et al., "On the design space between molecular mechanics and machine learning force fields", *Applied Physics Reviews* **12**, 021304 (2024). DOI 10.1063/5.0237876.
   - Jon Paul Janet and Heather J. Kulik, *Machine Learning in Chemistry*; American Chemical Society, 2020. DOI: 10.1021/acs.infocus.7e4001
   - Hugh M Cartwright (Ed.), Machine Learning in Chemistry: The Impact of Artificial Intelligence, The Royal Society of Chemistry, 2020.
   - Python Machine Learning, Sebastian Raschka & Vahid Mirjalili , Packt Publishing; 3rd ed. Edition, 2019.

8. **Any other remarks:** Additional readings will be drawn from the current machine learning literature.

Dated: _____ , Proposer: _____

Dated:_____, DUGC/DPGC Convener:_____

The course is approved / not approved

Chairman, SUGC/SPGC

Dated: