Professional Education

MACHINE LEARNING FOR MATERIALS INFORMATICS

Instructor: Markus J. Buehler, mbuehler@mit.edu

Note: All times are US Eastern Daylight Time. Schedule is subject to change.

	MONDAY	TUESDAY	WEDNESDAY	THURSDAY
9:00–10:30 AM	Foundations in materials informatics I (data science and basic concepts of machine learning). **L1**	Hands-on introduction to PyTorch (example application to fine-tuning a scientific BERT NLP model applied to protein molecular design; Hugging Face open-source tools and ecosystem). **L4**	Transformer models and applications to materials engineering and healthcare (AlphaFold and related); applications to protein design, synthesis. **L8**	Auto encoders and generative diffusion models (vision, graphs/networks, proteins, composite design) and variants. **L12**
10:30–10:45 AM	BREAK			
10:45 AM–12:30 PM	Foundations in materials informatics II (first-principles based multiscale modeling; synthetic datasets, experimental methods for data collection, self-driving labs). **L2**	Hands-on introduction to TensorFlow (example application to developing an adversarial neural network); practical guide to tensor algebra and other basic math concepts needed. **L5**	Deepening the understanding of language models applied to materials (pre-training and fine-tuning, discussion on HPC/GPU/CUDA); BERT, GPT-4/4o, Claude, and other LLMs (applications to materials problems; category theory; time-dependent material phenomena; fracture/failure mechanics). **L9**	**CLINIC 4** Participant presentations and discussion, clarifications and feedback.
12:30–1:00 PM	LUNCH BREAK			
1:00–2:30 PM	**CLINIC 1** Convolutional neural networks: Classifier, regression, and explainable/interpretable methods.	Ethics, bias and sustainability in materials informatics and AI; data science, statistics and visualization (includes review of relevant Python toolkits); collaborative research and open science practices. **L6**	**CLINIC 3** Transformer models from scratch (application to inverse problem solutions); Inference and deploying machine learning models in production and optimizing performance, high-throughput methods.	
2:30–2:45 PM	BREAK			
2:45–4:00 PM	Digging deeper: Deep neural nets, loss functions, Stochastic optimization methods (e.g., stochastic gradient descent and variants),regularization. **L3**	**LAB 1** Data, data, everywhere...data mining and dataset construction and application to build a deep neural network (covers computer vision tools; synthetic data generation, data mining, IoT, data augmentation).	Adversarial neural networks and multi-agent AI: Applications in materials design (manufacturing, inverse problem, characterization). **L10**	
4:00–5:00 PM	**CLINIC 2** Material failure analysis; physics informedand physics-aware neural networks; feedback session (participants can provide input or ask questions about the day's content).	Introduction to graph neural networks (applications to molecular systems, truss systems, alloys, proteins, and healthcare; graph transformers); feedback session. **L7**	Case study: Image segmentation and Multimodal Vision-LLMs: Applications in microscopy, medical imaging, and vision-based multimodal reasoning, distillation and design; feedback session. **L11**	
5:00–7:00 PM	**VIRTUAL RECEPTION** (includes participant introductions)	**Optional:** Time for group work and assignments (can be arranged within groups at other times)	**Optional:** Time for group work and assignments (can be arranged within groups at other times)	