Applied AI for Chemical-Physics: From Lab to Code

A Hands-on Tutorial

Workshop: Computational Chemistry and Machine Learning for Chemical Physics

June 8-12, 2026

Platform: Google Colaboratory (Python 3)

Full-Day Intensive: **June 12, 2026**

Workshop Overview

This tutorial is designed for experimentalists and theoreticians working on chemical / physics or related research areas. We will use **Google Colaboratory**, allowing us to run high-performance code and Neural Networks in the browser with zero local installation. We will focus on turning raw experimental data / artifitial data into predictive insights.

We will provide **pre-programmed Jupyter notebooks** to facilitate our hands-on tutorial. Although the session is designed to be **accessible without a programming background**, a foundational understanding is certainly an asset. We will guide you through interactng with the code to adjust variables and analyze outcomes.

Kindly note that the program may undergo slight adjustments leading up to the event.

Workshop Project Roadmap

Module 1 (Introduction): Target: Basic concepts of Machine Learning.

Module 2 (Denoising): Target: Spectral signals. *Tools:* SciPy Signal.

Module 3 (Mapping: Extraction of Main Features): Target: Spectra, Images. *Tools:* Scikit-learn PCA.

Module 4 (Morphometry): Target: Microscopy. *Tools:* Segment Anything (SAM).

Module 5 (Computer Vision): Target: Failure Detection / Main Features Detection in images. *Tools:* PyTorch CNNs.

Module 6 (Collaborative Group Projects): Target: Related Projects of General Interest (depending on the audience).

1 Module 1: Introduction to Machine Learning (09:00 – 10:15)

Tools: *Scikit-learn, NumPy, Matplotlib*

- **Supervised Learning:** Understanding how models map inputs (X, like concentration) to outputs (y, like conductivity). We distinguish between **Regression** (continuous values) and **Classification** (discrete categories).
- **Unsupervised Learning:** Discovering hidden structures in unlabeled data. Essential for exploratory research where you don't know the "answer" yet.
- **The ML Workflow:** Learning the rigor of splitting data into *Training* and *Testing* sets to ensure our model generalizes to new experiments.

- **Loss Functions:** A mathematical "ruler" (like Mean Squared Error) that tells the AI how far off its prediction is, guiding the optimization process.

2 Module 2: Data Foundations, Preprocessing & Denoising (10:30 – 12:00)

Tools: *Pandas, SciPy, Matplotlib*
Objective: Automate the transition from raw, noisy instrument output to "Publication-Ready" data.

- **Experimental Source:** Cyclic Voltammetry (CV), TGA, high-frequency rheology, spectra.
- **Data Cleaning:** Techniques to handle "human" errors in spreadsheets, such as missing values or inconsistent units, using Pandas dataframes.
- **Signal Denoising:** Applying *Savitzky-Golay* filters to smooth noisy spectra (FTIR/Raman) or electrochemical signals while preserving peak intensity and position.
- **Hands-on Baseline Correction:** Automating the removal of background drift in Cyclic Voltammetry (CV) or XRD patterns to isolate the true physical signal.

3 Module 3: Dimensionality Reduction for Spectroscopy (12:00 – 13:00)

Tools: *Scikit-learn (PCA, t-SNE)*
Objective: Visualize relationships between hundreds of chemical samples in a single 2D map.

- **Experimental Source:** SEM/TEM micrographs, Optical microscopy, AFM height maps, spectra, etc.
- **PCA Concepts:** Execute **Principal Component Analysis (PCA)** to identify the wavelengths (or another marker) responsible for 90% of variance, allowing for 2D visualization.
- **Eigenvectors & Variation:** Understanding which specific wavelengths (or marker) contribute most to the differences between your samples.
- **Hands-on Clustering:** Using *K-Means* to automatically group samples with similar (chemical or physical) signatures, identifying sub-phases or degradation states or synthesis success, etc.

4 Module 4: General Image Processing & Morphometry (14:00 – 15:15)

Tools: *OpenCV, Scikit-Image, Segment Anything (SAM)*
Objective: Quantify physical structures (pores, fibers, grains) without manual tracing.

- **Experimental Source:** SEM/TEM micrographs, Optical microscopy, or AFM height maps, etc.
- **Beyond Simple Thresholding:** Moving past "black and white" filters to AI-driven segmentation that can handle shadows and low-contrast micrographs (SEM/TEM).
- **Automated Feature Extraction:** Using Python to count particles, measure grain surface areas, and calculate porosity distributions across hundreds of images in seconds.
- **Hands-on Morphometry:** Extracting the "Aspect Ratio" and "Sphericity" of particles to correlate physical shape with macroscopic material performance.

5 Module 5: Convolutional Neural Networks (CNNs) (15:30 – 17:00)

Tools: *PyTorch or TensorFlow/Keras*
Objective: Moving beyond peak-fitting by training a Deep Learning model to "read" raw experimental waveforms / microscopy images.

- **The 1D Convolution:** Understanding how a sliding "filter" (kernel) identifies local patterns in a spectrum, such as shoulders, asymmetric broadening, or specific oscillation frequencies.
- **Translation Invariance:** Why CNNs are superior to standard regression for signals: they can recognize a chemical signature even if the peak shifts slightly due to calibration drift or temperature.
- **Hierarchical Learning:** How the first layers detect "edges" (slopes) and deeper layers detect "motifs" (complex peak patterns or redox signatures).
- **Hands-on Categorizing:** Categorize unknown samples into distinct physical states (e.g., *Cross-linked* vs. *Non-cross-linked* or *Degraded* vs. *Pristine*).

6 Module 6: Collaborative Group Projects (17:00 – 18:00)

The following are just examples of potential collaborative projects, but in practice, projects will be adapted to specific interests of participants.

6.1 High-Dimensional Data Mining

Focus: FTIR, UV-Vis-NIR, and Hyperspectral Imaging. Handling thousands of variables per sample.

6.1.1 Project 1: The Spectral Fingerprint

Goal: Extracting meaningful chemical trends from raw "Big Data" spectra.

- **Techniques:** FTIR/UV-Vis with Integrating Spheres.
- **AI Strategy:** Dimensionality Reduction (PCA) and Manifold Learning (t-SNE).
- **Hands-on:** Visualizing the "Chemical Space" of a 100-sample aging study. Identifying outliers and "hidden" degradation phases that aren't visible in individual plots.

6.1.2 Project 2: Automated Surface & Image Analysis

Goal: Turning micrographs into quantitative physical tensors.

- **Techniques:** SEM, Optical Microscopy, and Profilometry.
- **AI Strategy:** Computer Vision (Zero-Shot Segmentation with SAM).
- **Hands-on:** Automated quantification of surface roughness and pore-size distributions in IPN membranes. Transitioning from "qualitative photos" to "statistical data."

6.2 Modeling Functional Signals

Focus: EIS, TGA, and Mechanical Testing. Modeling curves where "shape" carries the physics.

6.2.1 Project 3: Smart Fitting of Complex Impedance

Goal: Replacing manual equivalent circuit "guessing" with AI-driven parameter extraction.

- **Techniques:** Electrochemical Impedance Spectroscopy (EIS).
- **AI Strategy:** Functional Data Analysis and Genetic Algorithms for Optimization.
- **Hands-on:** Using an AI agent to automatically fit Nyquist plots and extract charge-transfer resistance (R_{ct}) across a temperature series.

6.2.2 Project 4: 1D-CNNs for Signal Recognition

Goal: Training "Expert Systems" to recognize physical states from raw waveforms.

- **Techniques:** TGA/FTIR Coupling, Tensile Tests, and CV.
- **AI Strategy:** 1D Convolutional Neural Networks (1D-CNN).
- **Hands-on:** Training a model to "read" a TGA-FTIR trace and instantly identify the chemical family of the evolved gas based on the 1D waveform shape.