Toward Generalizable, Data-Efficient Self-Driving Laboratories for Organic Materials

Martin Seifrid

Organic mixed ionic-electronic conductors (OMIECs) underpin emerging technologies in sensing, bioelectronics, and energy devices, but their behavior depends on coupled chemistry, processing, structure, and environment. We are developing self-driving laboratory workflows that combine laboratory automation, uncertainty-aware machine learning, and disciplined data infrastructure to accelerate OMIEC discovery while improving transferability across chemical space and across laboratories.

Self-driving labs for OMIECs: We focus on closed-loop experimentation where automated execution is paired with explicit uncertainty quantification and structured record-keeping, so that the loop selects informative experiments and remains robust to day-to-day laboratory variability.

Generalizable ML from literature: Because SDL campaigns often start with limited internal data, we treat literature as a valuable but biased prior. We curate machine-readable datasets for conjugated polymers and evaluate models under out-of-distribution conditions that reflect real design use cases, highlighting when apparent performance under standard splits fails to translate to new chemistries, solvents, or processing histories.

Mixed kernel Gaussian processes in the low-data regime: To support sample-efficient optimization, we explore how to most effectively combine continuous process variables with non-Euclidean data such as molecular structures in Gaussian-process surrogate models. These mixed-input GPs provide calibrated uncertainty for Bayesian optimization while preserving practical usability in iterative laboratory loops.

Metadata-forward data management: We design pipelines where metadata are first-class scientific data. Capturing context such as sample history, measurement environment, and instrument configuration enables error attribution and delta-learning approaches that convert apparent noise into learnable structure.

Outlook: By unifying automation, generalizability-focused ML, mixed-input GP modeling, and metadata standards, we aim to develop OMIEC self-driving workflows that are data-efficient, reproducible, and genuinely transferable.