CURATED TRAINING SETS FOR INNOVATIVE AI PREDICTIONS

Solution success story

The challenge: Implementing a machine learning approach to innovation

A large diversified chemical company sought a first-mover advantage by developing new electronic devices with specific, novel properties. Traditionally, new materials and devices had been developed by a few senior scientists in a trial-and-error process, leveraging their unique expertise and intuition. However, domain knowledge was not easily accessible and project momentum was lost if individuals left the organization.

As the company investigated machine learning methods to augment their R&D efforts, they discovered that the required training sets were not available internally and difficult to find. While the company had collected a relatively large volume of data that could be used to begin training their predictive models, the team quickly realized that they did not have the needed training sets internally to support full model development. All available external sources contained generalized material properties and lacked the specific properties the scientists had identified as crucial for development. They needed to acquire high-quality training sets to generate innovative machine learning predictions about the relationship between a materials' specific properties and related device properties.

CAS

A division of the
American Chemical Society

The solution: Training a machine learning model with custom-curated data accelerates R&D breakthroughs

Initially, the informatics team leveraged their internal data to train and test models. Preliminary results reflected a limited view that negatively impacted the accuracy of predictions. To expand the model's narrow focus, diverse scientific data was required. Turning to CAS for help, key stakeholders reviewed the readily available data in the comprehensive CAS Content Collection™ and discussed custom-curated data sets with the CAS Custom Services team. CAS scientists went to work selecting the related materials and device properties, leveraging previously indexed content as well as sourcing additional content from published literature and patents. The highly structured content is human-curated from scientific literature to capture nuanced connections within the information.

Once trained on the highly specific and targeted datasets from CAS, the model generated precise recommendations about the substances, process conditions, layout, and assembly of the devices. The results were so promising that additional custom-curated training sets were commissioned to continue refining prediction accuracy and transferability. The emerging materials informatics team minimized risk by validating the accuracy of the model's predictive recommendations each step of the way. This systematic approach to R&D accelerated the discovery of new materials and devices, completing efforts in half the time of similar projects. Additionally, the leadership's confidence in the project was bolstered by aligning progress updates with quarterly reporting. The organization is now tracking to be first-to-market to deliver the novel, optimized device.

"Data quality is the top challenge cited by decision makers when trying to deliver AI capabilities."

Forrester

This systematic approach to R&D accelerated discovery, completing efforts in half the time of similar projects.

Company estimate

Find out how CAS Custom Services can help you transform scientific data into actionable, evidence-based insights that maximize investment and fuel success.
Learn more at cas.org

CAS

A division of the
American Chemical Society