College of Science

Bioinformatics

The University of Texas at El Paso

Getting the Most out of Published Science with Machine Learning

Dr. Gilchan Park, Computational Scientist

AI Department, Brookhaven National Laboratory, Upton, NY

The accumulated publications of the myriad scientific disciplines are unquestionably among the most valuable resources in science today. Every new experiment, discovery, theory, and analysis stands in the context of the work that came before it, and was indeed informed and made possible by that prior work. However, as scientific publication rates continue to grow, researchers become less able to keep fully abreast of their field or even all the developments in their area of specialization. Traditional keyword-based search and reference-crawling techniques are at once too rigid and too imprecise to find all and only the relevant publications a researcher seeks. Furthermore, tables and figures – containing much, if not most, of the valuable information in scientific documents – remain largely opaque to automated information extraction tools. New techniques are needed to better leverage the scientific literature, and this seminar will discuss both general approaches and specific applications where machine learning can fulfill this need. In particular, this talk will cover work in visual document analysis, named entity recognition, relation extraction, visual table parsing, automated plot recognition and value extraction, and multi-class knowledge integration. The maturity of the discussed applications and techniques varies, and ongoing work continues in multiple directions. Just as machine learning has come to be embraced as an effective set of new tools for scientific discovery and analysis, so too can it be effective in helping researchers make better use of the wealth of publications already in the hands of the scientific world.

Friday, April 24, 2026, 10:00 AM

Please pre-register for the Zoom meeting at

https://us06web.zoom.us/meeting/register/nlo4PvymQzq6A7LjViNjvg

We acknowledge the Sustainable Horizons Institute and DOE national laboratories for organizing this Computational Research Leadership Council Seminar Series.