**Rise of the machines: what is molecular data science and why should you care?**

**ABSTRACT:** This seminar will share recent advances from our research group at the intersection of molecular simulations and data science. The confluence of cheap storage, fast computers, and increases in spatial/temporal resolution of methods spanning physics-based simulations to advanced microscopy is creating challenges and opportunities for chemical engineers across many sub-disciplines. The ability to harness advanced techniques to store, visualize and process large data sets will provide a strategic research advantage and the capacity to maximize information and knowledge from our research data. Through two examples, I will highlight how the methods of data science are making an impact on my sub-disciplines of molecular simulation and reaction engineering.

After a brief discussion of context and relevant methods of data science, I will spend most of my seminar discussing a project from my group in the area of automated computational molecular design. Over the past decade, there have been many successful applications of data science towards building predictive tools for the design of small molecules or inorganic crystalline materials, primarily for energy applications such as solar energy or battery materials. We have extended this line of research into molecular design of liquids. I will explain the workflow of a molecular design algorithm we have built, as well as share examples of property design for the example of heat transfer fluids. Our workflow involves the use of a genetic algorithm for predicting new molecules, a self-learning neural network for property prediction, and molecular dynamics simulations for screening of candidate molecules. With the remaining time, I will discuss a method our group has developed for discovering complex chemical reaction networks, an application toward the degradation of common battery electrolytes, and the use of statistical models to enhance the simulations.
Jim Pfaendtner is the Bindra Career Development Professor and Associate Professor of Chemical Engineering at the University of Washington. Additional appointments include the University of Washington Associate Vice Provost for Research Computing, Senior Scientist at the Pacific Northwest National Lab and a Senior Data Science Fellow at the UW eScience Institute. He holds a B.S. in Chemical Engineering (Georgia Tech, 2001) and a Ph.D. in Chemical Engineering (Northwestern University, 2007). Jim's research focus is computational molecular science and his recent teaching interests are in the area of teaching data science skills to graduate students in chemical engineering in his role as director of an NSF graduate training program (NRT) at the intersection of data science and clean energy. Jim is also passionate about health and fitness, holds as CrossFit Level 1 trainer certificate, and lately has been honing his skills at the PlayStation 4 game “Assasins Creed: Odyssey.”