Informatics as a Service:
Machine Learning at Citrine

Sean Paradiso
Computational Scientist
Citrine Informatics
Redwood City, CA
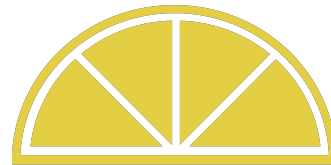
Thursday, October 6th * 3-3:45 PM * ESB 2001
Followed by reception, awards ceremony, and panel

The materials development process is long, labor intensive, and costly. In 2011, the Obama Administration launched the Materials Genome Initiative (MGI) with the aim of integrating experimental data, large-scale computation, and informatics to substantially reduce the time it takes to integrate new advanced materials into commercial applications. While sound in theory, the current state of materials data infrastructure has rendered this task nearly impossible. Experimental databases are scattered across the globe, often locked behind paywalls, and aren't structured in a format that enables mining for patterns across multiple datasets. Moreover, data aggregation and structuring is merely a necessary, not sufficient condition for making the vision of the MGI a reality. It remains an open question how such a database should be analyzed. What is the transferable basis in which multiscale physics is succinctly expressed through experimental data? In an attempt to address these challenges, Citrine has developed an extensive open data platform where industrial and academic researchers across the world are assembling to build a common database of materials properties. This talk will provide an overview of the Citrination platform, recent scientific accomplishments made in collaboration with partnering institutions, and a special focus on the machine learning infrastructure that powers the predictive analytics under the hood.

Sean studied polymer field theory under Glenn Fredrickson at UCSB. Near the end of his degree, he stumbled into a global optimization project that linked his passion for compute infrastructure with the fascinating problem of compressing high dimensional search spaces using intuition, theory, and algorithms. He discovered machine learning coincidentally while searching for an automated way to detect undesired phase transitions. He joined Citrine roughly 1 year ago as employee #3 to build out their machine learning infrastructure and is eager to share the good, bad, and ugly about joining an ambitious startup at such an early stage.