



Universität St.Gallen

Computer Science Insights @HSG - School of Computer Science

Thursday, Nov 19th, 2026 - 16:15, Room C 83-1235, House Washington

Doing Science with an AI Colleague: Collaboration, possibilities,
and mental models

Daniel M. Russell

The role of AI in scientific research is undergoing a fundamental shift, moving from passive data-processing tools to active, autonomous collaborators. As AI systems become capable of deep reasoning and autonomous experimental design, understanding how human scientists think about, interact with, trust, and understand these AI co-scientist systems is critical.

The success of agentic systems for scientific research depends entirely on having a human in the loop. We'll discuss the mental models scientists have while working with AI collaborators, noting how misaligned mental models can lead to automation bias or algorithm aversion, and then explore how we can design AI systems that actively help users build accurate, calibrated models of the AI's capabilities.

We have found that accelerating scientific discovery requires more than simply scaling compute or engineering better objective functions; it also relies on designing AI systems that align with human cognition, providing accurate mental models, while working in a safe, trusted, and productive partnership.

Dr. Daniel M. Russell has been working in Artificial Intelligence and Human-Computer Interaction for nearly 40 years at several of the top innovation companies in Silicon Valley (Google, Apple, Xerox PARC, IBM). He was in the core search engineering team at Google for over 17 years and now works on the Google Deep Mind "Co-Scientist" project to build an AI science collaborator. He currently teaches in the Human-AI group at Stanford's Computer Science department and at the University of Zürich. He has been on the faculty at Stanford, the University of Maryland, the University of Zürich, Santa Clara University, UC San Diego. His online classes have been watched by millions of students for an accumulated YouTube watch-time of > 450 years.

From insight to impact.

