Multimodal Data Capture and Modeling in AI-Driven Scientific Discovery

Patrick W. Ruch*, Antonio Foncubierta, Carlo Baldassari, David Diener, Dimitrios Christofidellis, Federico Zipoli, Irina Espejo Morales, Jannis Born, Joris Cadow-Gossweiler, Joris Schaller, Mara Graziani, Marvin Alberts, Matteo Manica, Teodoro Laino

IBM Research Europe – Zurich, 8803 Rüschlikon, Switzerland
** ruc@zurich.ibm.com*

Scientific experimentation is characterized by a variety of data modalities, but capturing detailed accounts of laboratory work is a challenging endeavor [1]. We explore the use of multimodal models for automatic action recognition in laboratory environments to provide real-time transcription of scientific actions from a first-person perspective. The importance of training data and model choices for automatic documentation of laboratory experiments is discussed. Also, we consider similar data-driven approaches to describe complex systems featuring multiple modalities, including molecular spectroscopy and electrochemical energy storage systems.

[1] G. Gabrieli, I. Espejo Morales, D. Christofidellis, M. Graziani, A. Giovannini, F. Zipoli, A. Thakkar, A. Foncubierta, M. Manica, P.W. Ruch, *Digital Discovery*, **2025**, *4*, 393-402.