

Please join the Biometric Colloquium

JOHANNES SCHWENKE

CLEAR Methods Center, Division of Clinical Epidemiology, Department of Clinical Research,
University Hospital Basel, University of Basel, Basel, Switzerland

CAN LOCAL, OPEN-SOURCE LLMs SAFELY UNLOCK ROUTINE ONCOLOGY DATA?

Wednesday 12th of November 2025 at 9:00am

Seminarraum Center for Medical Data Science
Spitalgasse 23, Room 88.03.512
Medical University of Vienna, 1090 Wien
Host: Felix Herkner

Abstract:

Large Language Models (LLMs) are increasingly used to assist or replace human data extraction in health research, with promising but varying levels of accuracy. The ability to turn unstructured clinical notes from electronic medical records into structured data at scale could enable new forms of quantitative research and data-driven changes in routine care. However, leading closed-source models are not usable in settings requiring complete patient privacy.

At the Department of Medical Oncology at the University Hospital of Basel, we are building programmatic pipelines to help quantify the entire patient journey. This requires structuring large amounts of free-text data, for example, from radiology reports. We explored whether locally hosted, open-source LLMs offer a viable, privacy-preserving solution.

We evaluated how several open-source models compare against a robust ground truth, established by duplicate human coders and senior oncologist adjudication, for tasks of varying complexity. In this talk, we will show how our hospital's GPU cluster is integrated into data extraction pipelines, compare how different models perform on these oncology-specific tasks, and share key lessons learned and paths forward for implementing these tools in a hospital setting.