

2025 EECMS Summer Internship Application Form

Main Supervisor	Tudor Groza, Rare Care Centre / Duke-NUS PRISM
Other supervisors (if applicable)	A/Prof Saumya Shekhar Jamuar, Duke-NUS PRISM, SingHealth Prof Gareth Baynam, Rare Care Centre
Project Title	Quantifying the degree of data duplication in Electronic Health Records for better AI, including more accurate medical LLM
Student location(s) for the project	KK Hospital for Women and Children, Singapore
Duration of project (ideally six weeks)	6 weeks
Project Description	<p>Devising informed care pathways for both diagnosed and undiagnosed patients - covering prognosis, surveillance, coordinated clinical care or family planning - is essential to support emotional relief, and appropriate access to resources. An accurate and comprehensive view over a patient's medical history enables the clinical team to gain a deep understanding of the trajectory of the condition and hence tailor the care pathways accordingly.</p> <p>The latest developments in large language models (LLMs) provide us with novel and more efficient ways to summarise and interpret multi-modal data. Using prompt engineering techniques on top of EHR data could lead to a more streamlined approach to build a patient's concise medical history. The nature of the clinical notes stored in the EHR may, however, raise various processing challenges, due to the predominant use of short-hand forms and the increased amount of duplication (within and across multiple specialties). Hence, using clinical notes for summarisation may introduce bias both in a Retrieval Augmented Generation (RAG) context, as well as when considering this data as input for a foundational model.</p> <p>This project aims to quantify the degree of data duplication in EHR records and to propose methods to increase the utility of the clinical encounters' content by reducing it to its most concise and up-to-date form. The study will use the clinical notes of a rare disease cohort tracked longitudinally across several years to measure the information content of each encounter at the level of the specialty and across multiple specialties within a given time period.</p>

Please email completed form to Tele Tan at t.tan@curtin.edu.au