Title: Recent Advances in Foundation Models: Continual Training, Time-Series Models and Revisiting Evals for Vision-Language Models

In this talk, I will provide an overview of some recent work in our lab on expanding both the capabilities and alignment of foundation models.

First, I will discuss the importance of the continual pretraining paradigm that allows to build upon previously trained foundation models indefinitely, as opposed to the current common practice of re-learning such models from scratch each time a new large dataset becomes available. I will discuss strategies allowing foundation models to keep absorbing new knowledge from the continually arriving datasets without "catastrophically" forgetting what these models have previously learned, and matching the performance that could have been achieved if all the data available so far were available for training at once.

Next, I will discuss recent advances in time-series foundation models (our Lag-LLama - the first open-source foundation model for time-series, as well as MOIRAI and other recent time-series foundation models). While time-series data are crucial to many practical applications of AI, from science to healthcare, the development of time series foundation models was, until very recently, lagging behind (pun intended :) the language and vision models. This situation started to change rapidly in the past year, as time-series forecasting seems to be now experiencing its own "GPT-3 moment".

Finally, the third part of this talk will focus on recent advances in evaluation approaches to large-scale vision-language models. I will present an open-source Robin suite of VLMs of different sizes, open-sourced by our team, as well as a range of evaluation approaches we explored. I will discuss some drawbacks of relying on automatic metrics and the importance of combining them with alternatives such as human- and AI-based evaluation.