



## Seminar/Talk

# Data Integration: Challenges and Opportunities for Interpolation Learning under Distribution Shifts

**Pragya Sur**

Harvard University

Host: Marco Mondelli

Min-norm interpolators naturally emerge as implicit regularized limits of modern machine learning algorithms. Recently, their out-of-distribution risk was studied when test samples are unavailable during training. However, in many applications, a limited amount of test data is typically available during training. Properties of min-norm interpolation in this setting are not well understood. In this talk, I will present a characterization of the risk of pooled min-L2-norm interpolation under covariate and concept shifts. I will show that the pooled interpolator captures both early fusion and a form of intermediate fusion. Our results have several implications. For example, under concept shift, adding data always hurts prediction when the signal-to-noise ratio is low. However, for higher signal-to-noise ratios, transfer learning helps as long as the shift-to-signal ratio lies below a threshold that I will define. Our results also show that under covariate shift, if the source sample size is small relative to the dimension, heterogeneity between domains improves the risk. Time permitting, I will introduce a novel anisotropic local law that allows to achieve some of these characterizations and is of independent interest in random matrix theory. This is based on joint work with Yanke Song and Sohom Bhattacharya.

**Monday, September 15, 2025 11:00am - 12:00pm**

Office Bldg West / Ground floor / Heinzl Seminar Room (I21.EG.101)



This invitation is valid as a ticket for the ISTA Shuttle from and to Heiligenstadt Station.

Please find a schedule of the ISTA Shuttle on our webpage:

<https://ista.ac.at/en/campus/how-to-get-here/> The ISTA Shuttle bus is marked ISTA Shuttle (#142) and has the Institute Logo printed on the side.