INDUCTION
STEP 2

Behavior Clustering

Aggregate the aspects from **M** (trajectory, feedback) pairs, and cluster them into **N** behavior clusters with common concepts. These concepts are the induced **metrics**.

Aspects from M trajectories

LLM →

Metric Description · Good Behaviors · Bad Behaviors

xN

AutoLibra **Metrics**