Robustness of Grammar Skill Detection

Synthetic Data — 33% (Mean), 95% (Validation/Test)

Manually Picked Data — 65% (Mean), 82% (Validation/Test)

Automatically Picked Data — 8% (Validation)

Legend: Mean (dashed), Validation, Test

X-axis: Precision (0% to 100%)