Average Log Likelihood over Training Steps

Legend:
- math (in-dist)
- math (out-of-dist)
- code (in-dist)
- code (out-of-dist)
- ifeval (in-dist)
- ifeval (out-of-dist)

X-axis: Training Step
Y-axis: Average Log Likelihood