# Optimizing dialog policies with ML

**Supervised learning:**

- You train to imitate the observed actions of an expert
- Often requires a large amount of expert-labeled data
- Even with a large amount of training data, parts of the dialogue state space may not be well-covered in the training data

**Reinforcement learning:**

- Given only a reward signal, the agent can optimize a dialogue policy through interaction with users.
- RL can require many samples from an environment, making learning from scratch with real users impractical
- That's why we need *simulated users* for RL

https://arxiv.org/pdf/1703.07055.pdf