DFA-conditioned RL setting

task

a_t

o_t

q_0
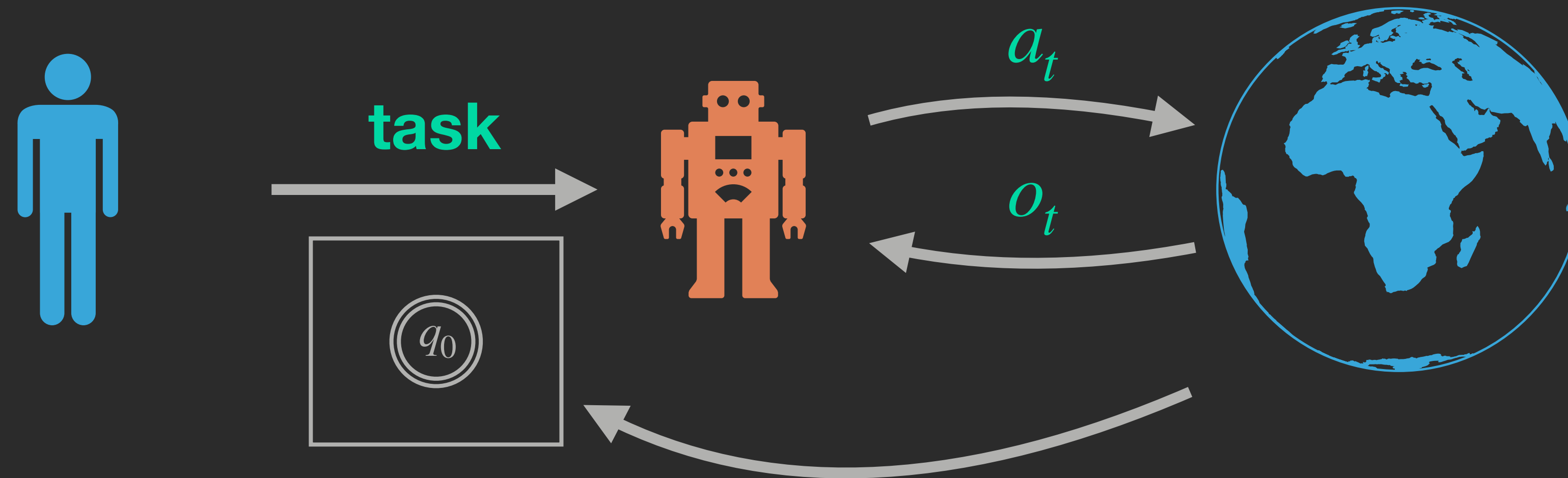
Q: How can we learn a DFA-conditioned policy?

Q: How can we learn a DFA-conditioned policy?

$$g_t$$

$$a_t$$

$$o_t$$

learn a "universal policy" that encodes the task and then feeds it into a neural policy

Schaul, Tom, et al. "Universal value function approximators." *International conference on machine learning*. PMLR, 2015.