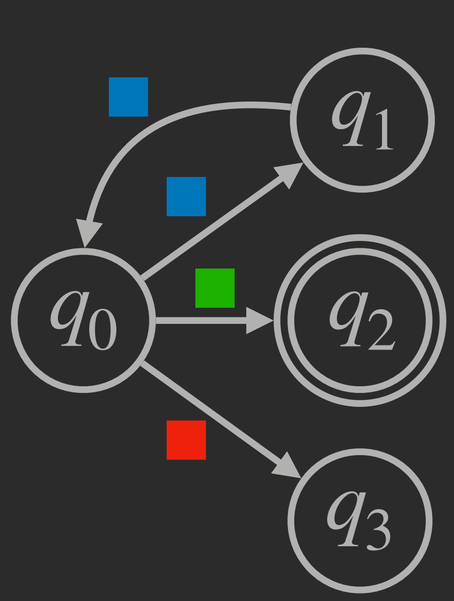
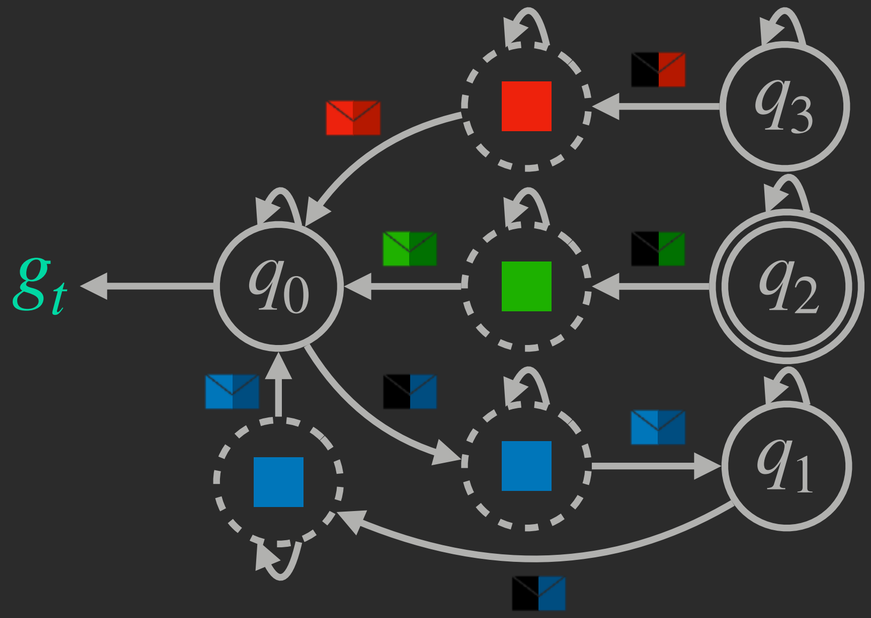
DFA featurization



1. Introduce a new node for each transition
2. Flip edges
3. Add self loops



Challenge: No Canonical Prior Distribution

How can we define a *pretraining distribution* for DFAs?

Countably infinite concept class, any distribution is necessarily biased