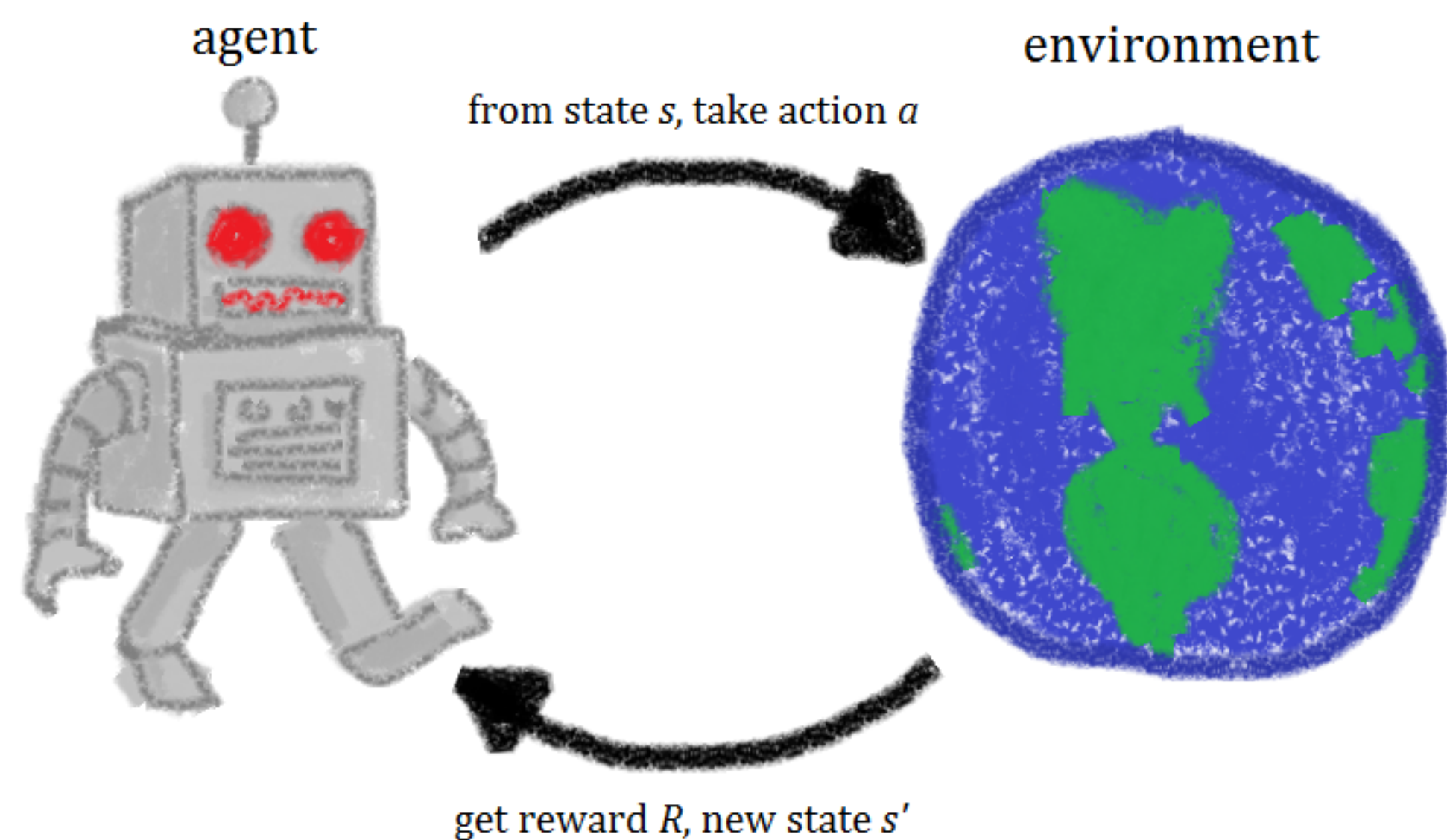


RLHF



$$V(s) = \max_a \left(R(s, a) + \gamma \sum s' P(s, a, s') V(s') \right)$$

