**Dreamer RL**

Two models:

* RSSM (Encode environment)
* Actor-Critic (output actions)

RSSM

* Deterministic hidden state
  + Encodes a hidden state to represent the environment
  + GRU/RNN takes the current hidden state and the stochastic state and current and outputs the next hidden state
  + Recursively to get the next state based on observations and actions
* Prior model
  + Outputs the sampling distribution of the stochastic state based on the hidden state and observation
  + Can use Normal Distribution or a simple MLP (with a random Gaussian Normal input)
  + Used when we have an actual observation
* Posterior model
  + Outputs the sampling distribution of the stochastic state based on the hidden state only (we don’t have an observation in this case)
  + Used when we are imagining or rolling out
* Stochastic state
  + Captures uncertainty of what we don’t know. Sampled from either prior or posterior model
  + Sampled from either prior or posterior model
* Combined state
  + Concatenation of both the deterministic hidden state and stochastic state
* Observation model
  + Decodes the environment (for example, the image that it sees)
  + Responsible for reconstruction
  + Important for training as the model knows whether it is effectively able to encode and decode the environment
* Reward model
  + Predict external reward based on combined state
  + Actor-Critic uses this reward as a basis during rollout (because we don’t actual reward)

Actor-Critic

* Actor
  + Trained to carry the best possible action based on the advantage
* Critic
  + Evaluates how much an action improves the reward over time, which is measured recursively
  + Future rewards are weighted by a discount factor