Guidelines while doing RL research (John Schulman)

General Baselines (First thing when doing anything)

- Try Cross-entropy method
- Tuned Policy Gradient
- Tuned SARSA
- Tuned Q-learning

Setup Benchmarks
Use multiple random seeds

How to approach a new algorithm?

- Use a familiar and small test problem.
- Do multiple experiments quickly
- Play with lots of hyperparameter values
- Interpret and visualise the learning process: state visitation, Value function

What to do when approached with a new task?

- Simplify the problem as much as you can
- Give good input features, shape reward function
- See what happens in the case of a random policy?
- Is the task humanly solvable with the given data?
- Plot time series of observations and rewards to check their scale

Usually try more samples than expected during the start.

Always do sensitivity analysis to each parameter.

Indicators to see how algorithm health

- VF fit quality
- Policy entropy
- Update size in output space and parameter space

Standardize Data

If the range of the observations and rewards are unknown, compute running estimate of mean and standard deviation

Clip rewards to a given range

Rescale rewards but don't shift their mean

Important Params:

- Discount (γ)
- Action frequency

Observe min/max/mean/stdev of episode returns
Look at episode lengths (Solving problems faster, losing games slower?)

Policy Gradient Strategies:

- Observe policy entropy
- KL divergence of old policy and new policy

Initialization of policy is extremely important as it determines the initial states that will be visited