# M5 — Reinforcement Learning Optimization

\*\*Purpose:\*\*

Optimize CAR-T design parameters (from M4) using a reinforcement learning (RL) loop to improve predicted efficacy and safety.

## Data

- \*\*Expression Data:\*\* TCGA-STAD.star\_counts.tsv.gz

- \*\*Phenotype Data:\*\* TCGA-STAD\_curated\_survival.txt

- \*\*Target Gene:\*\* ENSG00000066405 (CLDN18)

## Requirements

- Python 3.10+

- Packages: `numpy`, `pandas`, `matplotlib`, `scikit-learn`, `tensorflow` (or `pytorch` if preferred)

Install dependencies:

```bash

pip install -r requirements.txt

## Command

cd M5\_reinforcement\_learning

python m5\_rl\_optimize.py \

--expr "../data/TCGA-STAD.star\_counts.tsv.gz" \

--pheno "../M4\_feedback\_simulation/input/TCGA-STAD\_curated\_survival.txt" \

--gene "ENSG00000066405" \

--outdir "../m5\_out"

## Outputs

**rl\_learning\_curve.png** — RL training reward curve

**rl\_best\_policy.json** — Parameters for best-performing CAR design

**rl\_summary.txt** — Summary statistics and evaluation metrics

## Example Results

| **Iter** | **min\_detect\_prop** | **Sim\_p\_logrank** | **Tumor/Normal** |
| --- | --- | --- | --- |
| 1 | 0.10 | 0.882 | 1.85 |
| 3 | 0.20 | 0.520 | 2.35 |

## Notes

The RL agent was trained for 200 episodes with ε-greedy exploration.

The reward function combined tumor specificity, predicted patient survival benefit, and low off-target risk.

Initial parameters came from M4 module output.

🔗 **GitHub Repository:** [AI-CAR-Loop-1.0](https://github.com/ohahouhui/AI-CAR-Loop-1.0)  
🔗 **Data Source:** [UCSC Xena TCGA-STAD STAR-counts](https://xenabrowser.net/datapages/?dataset=TCGA-STAD.star_counts.tsv&host=https://gdc.xenahubs.net)

---