

Toy bandit problem with Gaussian-like Q-values (online fine-tuning)

