

**a**

Rollout policy

SL policy network

RL policy network

Value network



Policy gradient



Neural network

Classification

Classification

Self Play

Regression

Human expert positions

Self-play positions

Data

