

SAFE + EXPLAINABLE INTELLIGENT CONTROL

Human Input/Feedback

Data/Demonstrations

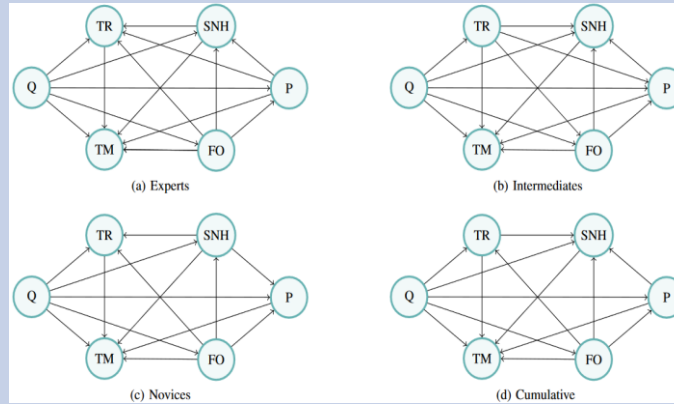


Environments

Formal Task Specifications φ

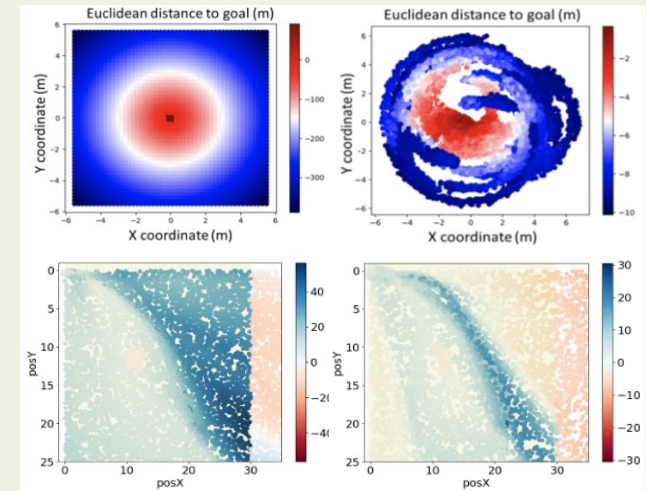
- Temporal Logic: LTL, STL
- Automata

Performance and Skill Evaluation

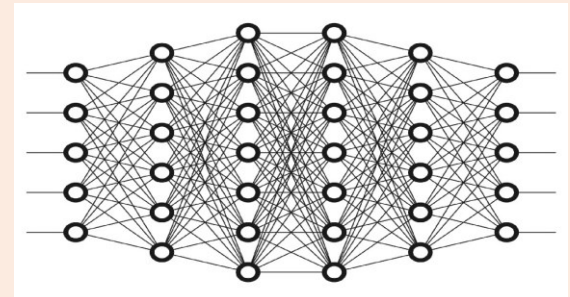


Explainable/Interpretable Behaviors (xAI)

Reward Inference



Deep Reinforcement Learning Control Policy π



Logic-Based Control

Formal
Verification
(in-the-loop)