



Adaptive Reactive Synthesis

Specification Inference

Verification

`Ensure controller satisfies the goal'

Trustworthy AI



RL



RL

3

7



Toward trustworthy decision making leveraging Formal Methods and RL

Trustworthy AI

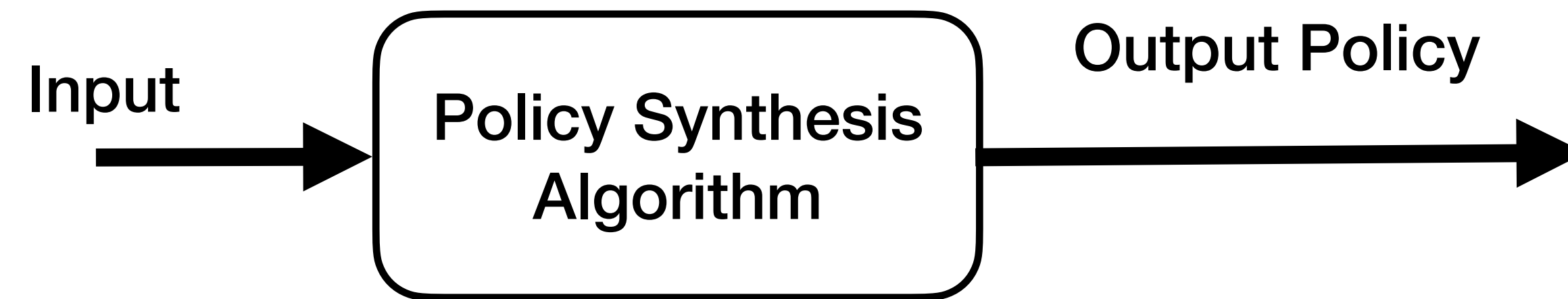
Towards trustworthy decision making leveraging Formal Methods and RL

Specification Inference
+ RL

Verification
‘Ensure controller satisfies the
goal’

Adaptive Reactive
Synthesis
+ RL

Shielding



-
1. M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, and U. Topcu (AAAI'18): Safe reinforcement learning via shielding
 2. B. Könighofer, J. Rudolf, A. Palmisano, M. Tappler, and R. Bloem (ISSE'23): Online shielding for reinforcement learning