Toward Automated Scenario Generation with Deep Reinforcement Learning in GIFT

Jonathan Rowe[1] Andy Smith[1] Bob Pokorny[2]

Bradford Mott[1] James Lester[1]

[1] North Carolina State University

[2] Intelligent Automation, Inc.

Simulation-Based Training

Challenges

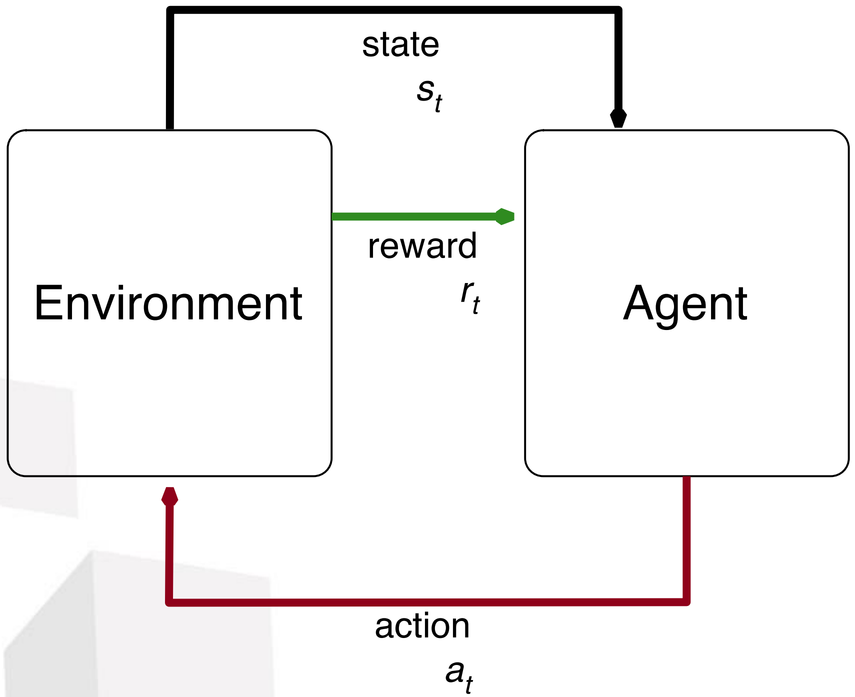
- Training scenarios are resource-intensive to create
- Authoring tools require specialized knowledge
- Training scenarios are often not reusable
- Finite set of training scenarios available to learners
- Training scenarios are often one-size-fits-all

Project Overview

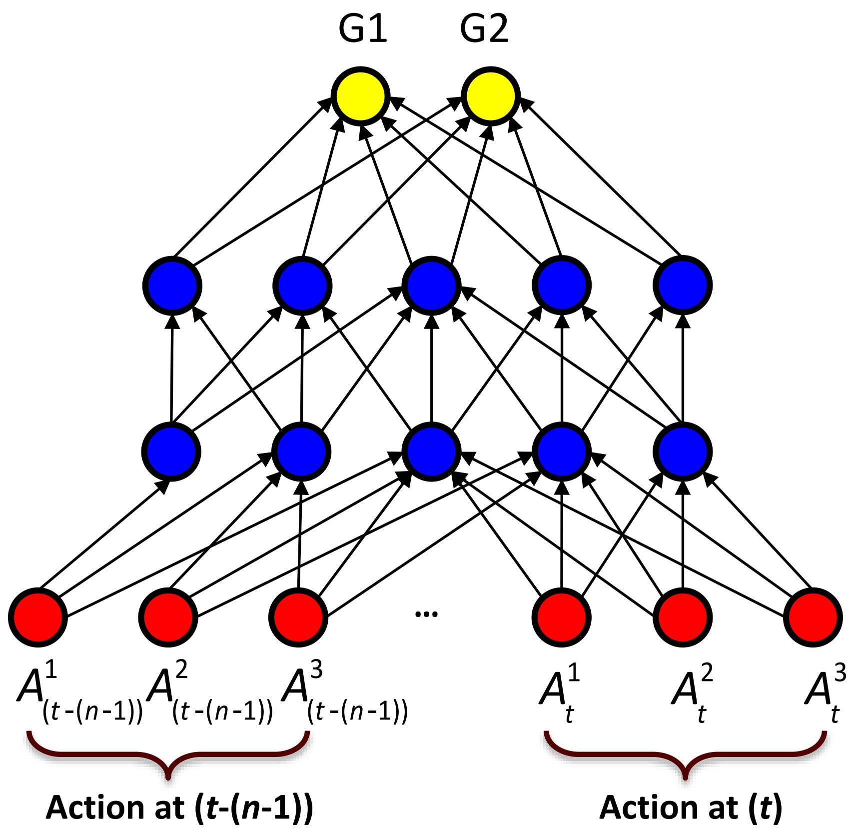
Research program objective: Devise generalized, data-driven scenario generation models that dynamically adapt training events to achieve target learning objectives in simulation-based virtual training environments.

Deep Reinforcement Learning

Reinforcement Learning



Deep Neural Networks



Outline

- Background and Related Work
- Deep RL-Based Scenario Generation
- Virtual Battlespace 3 Testbed Environment
- Preliminary Results
- Demo
- Conclusions and Future Work

Outline

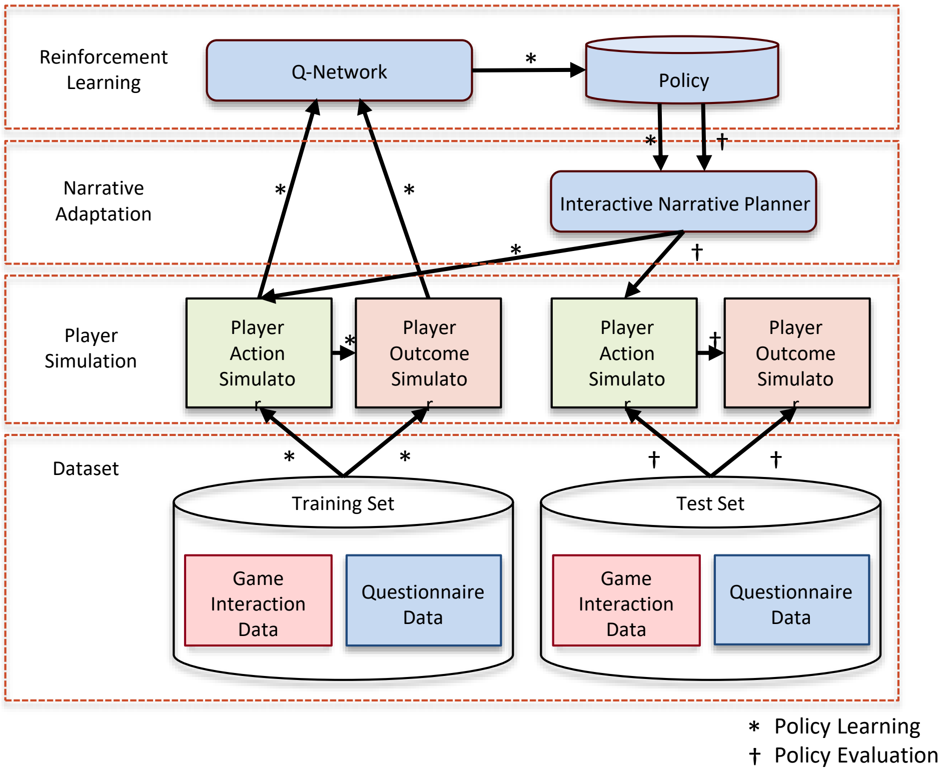
- **Background and Related Work**
- Deep RL-Based Scenario Generation
- Virtual Battlespace 3 Testbed Environment
- Preliminary Results
- Demo
- Conclusions and Future Work

Research Context: Interactive Narrative Planning

- Character-centric interactive narrative planning
  (Cavazza, Charles & Mead 2002; Si, Marsella & Pynadath, 2006; McCoy et al. 2014)
- Story-centric interactive narrative planning
  - Search-based drama management (Weyhrauch 1997; Nelson & Mateas, 2005)
  - STRIPS-style planning (Riedl & Young, 2010; Porteous, 2017)
  - Reactive planning (Mateas & Stern 2005; Barber & Kudenko 2007)
  - Case-based reasoning (Fairclough 2004; Ontanon & Zhu, 2011)
  - Decision-theoretic models (Mott & Lester, 2006)
  - Information retrieval (Swanson & Gordon, 2012)

Research Context: RL-Based Interactive Narrative

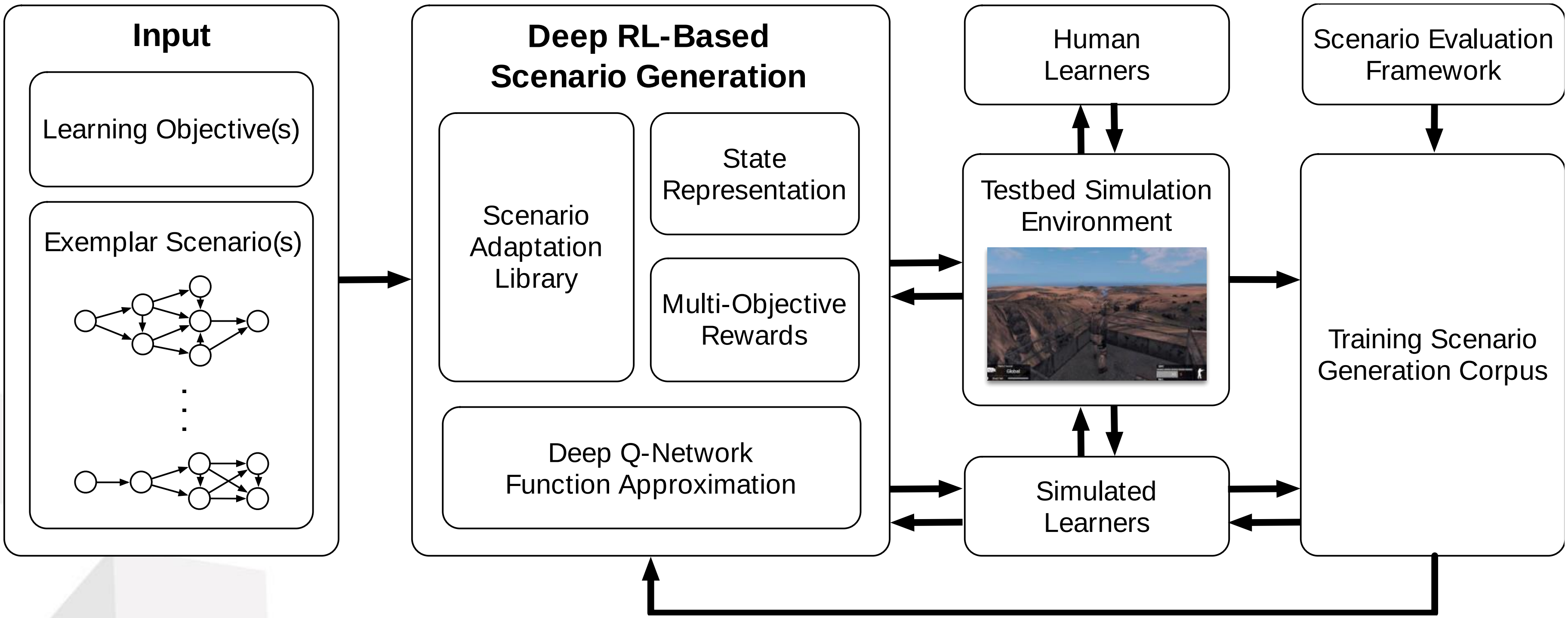
- Modular RL-based interactive narrative (Rowe, Mott, & Lester, 2014; Rowe & Lester, 2015)
- Representation optimization in modular RL-based interactive narrative (Wang, Rowe, Mott, & Lester, 2016)
- Multi-objective reinforcement learning (Sawyer, Rowe, & Lester, 2017)
- Deep RL-based interactive narrative personalization (Wang, et al., 2017a; b; 2018)



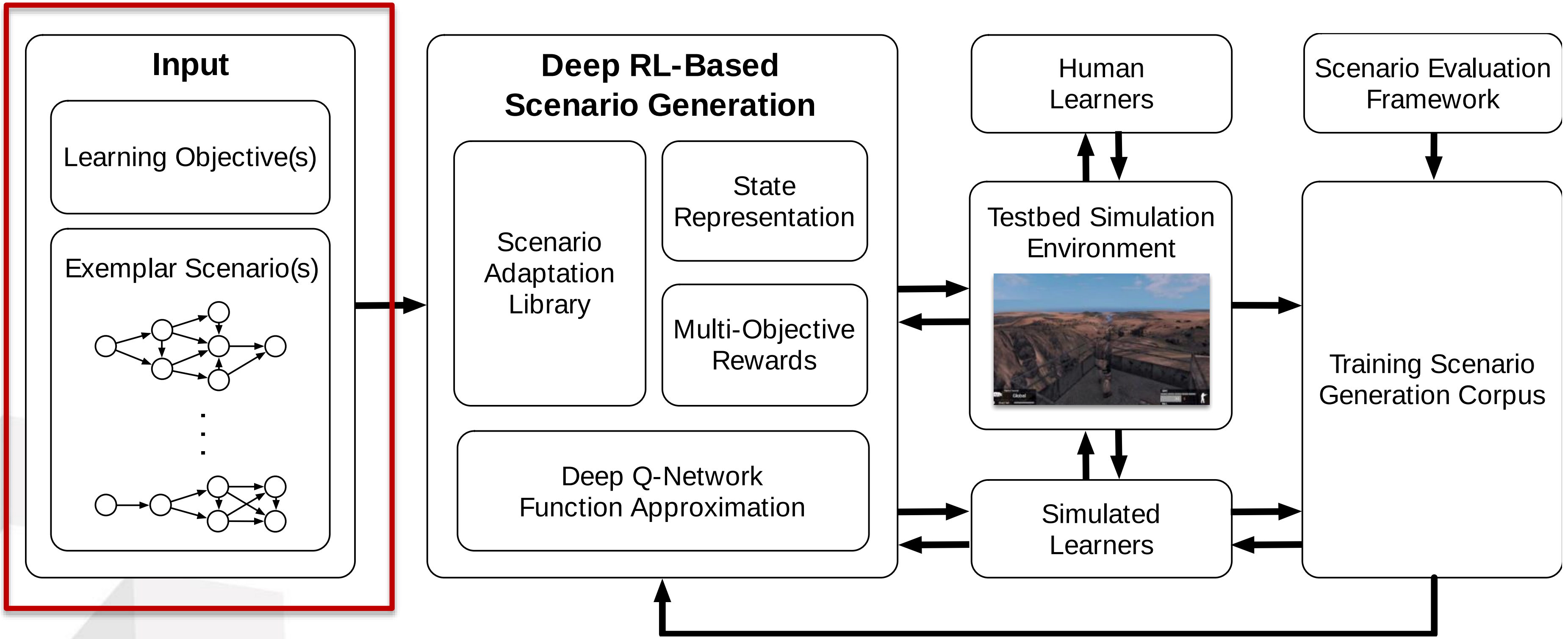
Outline

- Background and Related Work
- **Deep RL-Based Scenario Generation**
- Virtual Battlespace 3 Testbed Environment
- Preliminary Results
- Demo Video
- Conclusions and Future Work

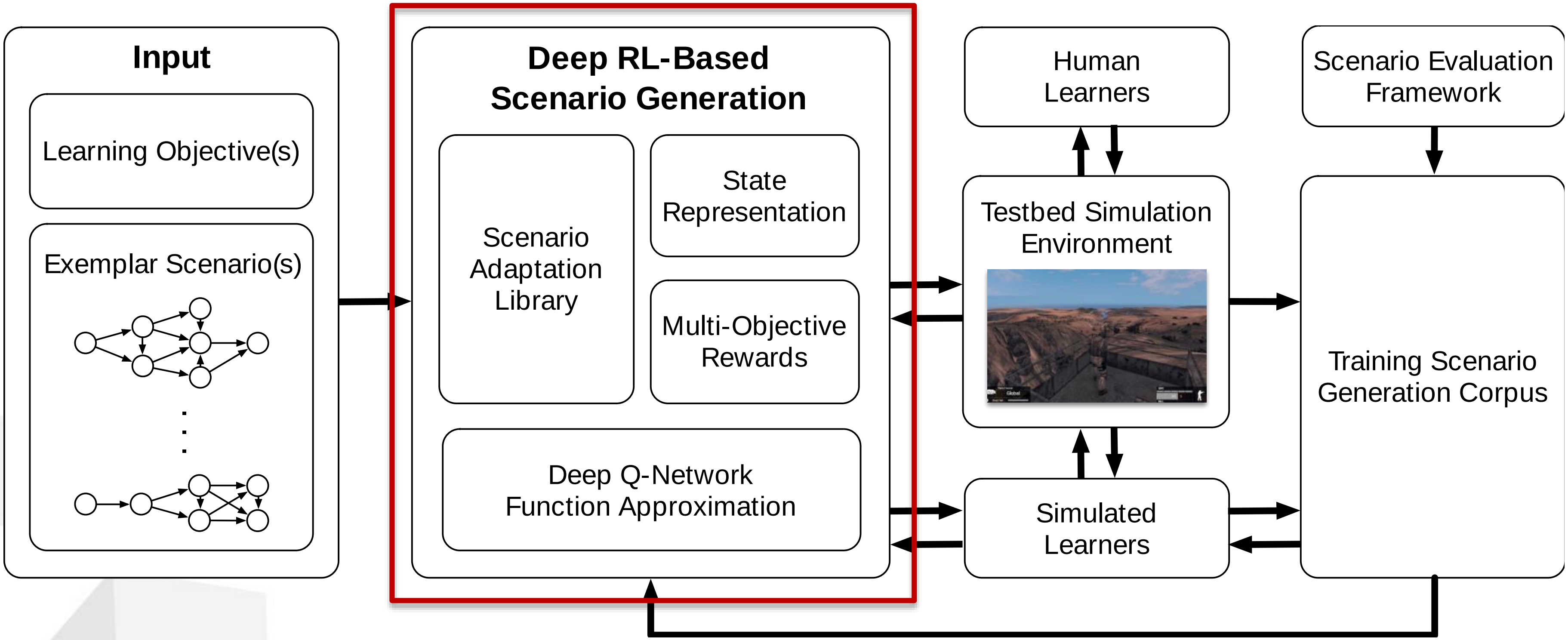
Deep RL-Based Scenario Generation Framework



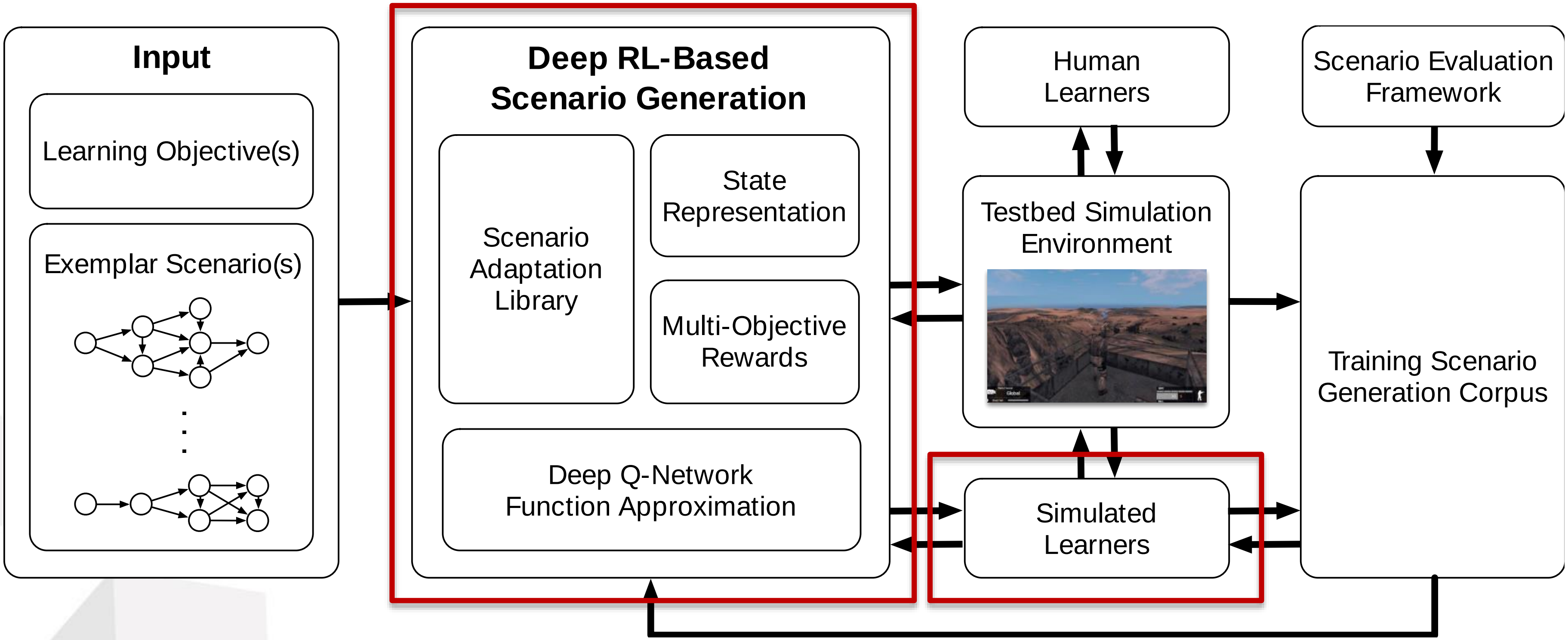
Deep RL-Based Scenario Generation Framework



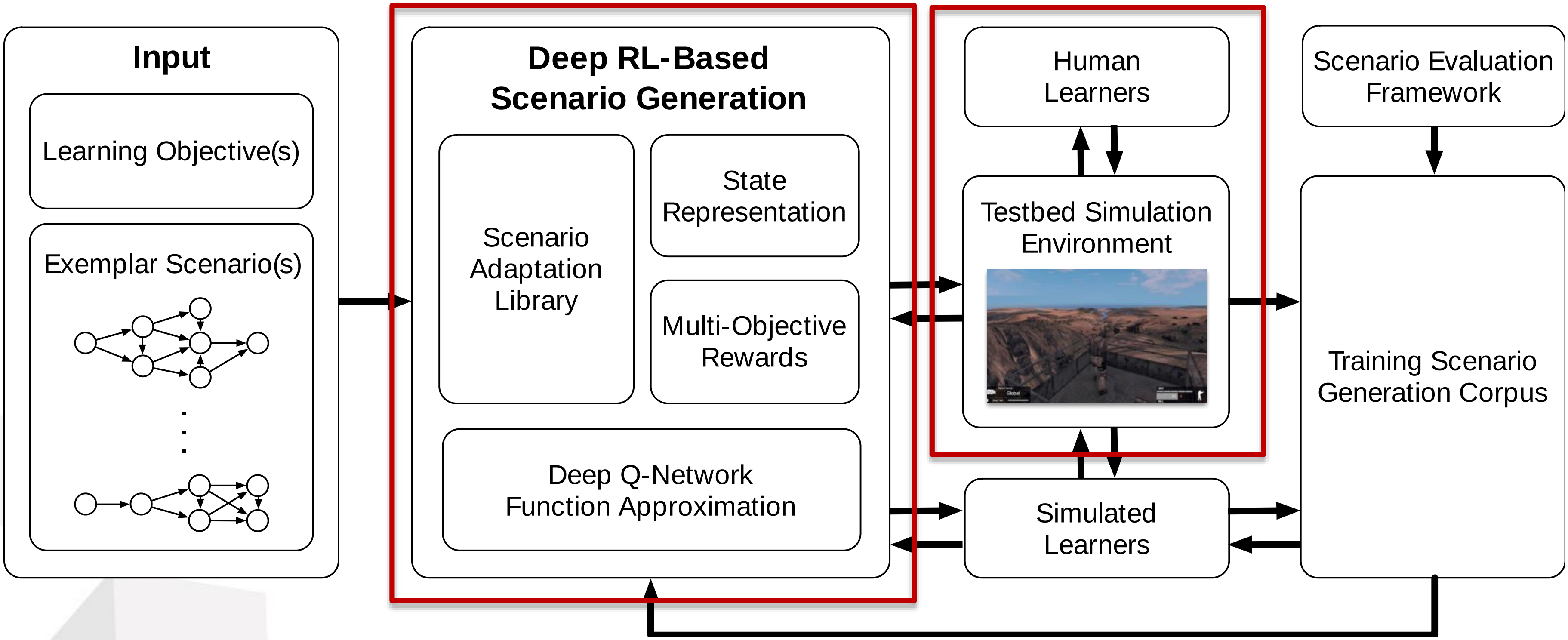
Deep RL-Based Scenario Generation Framework



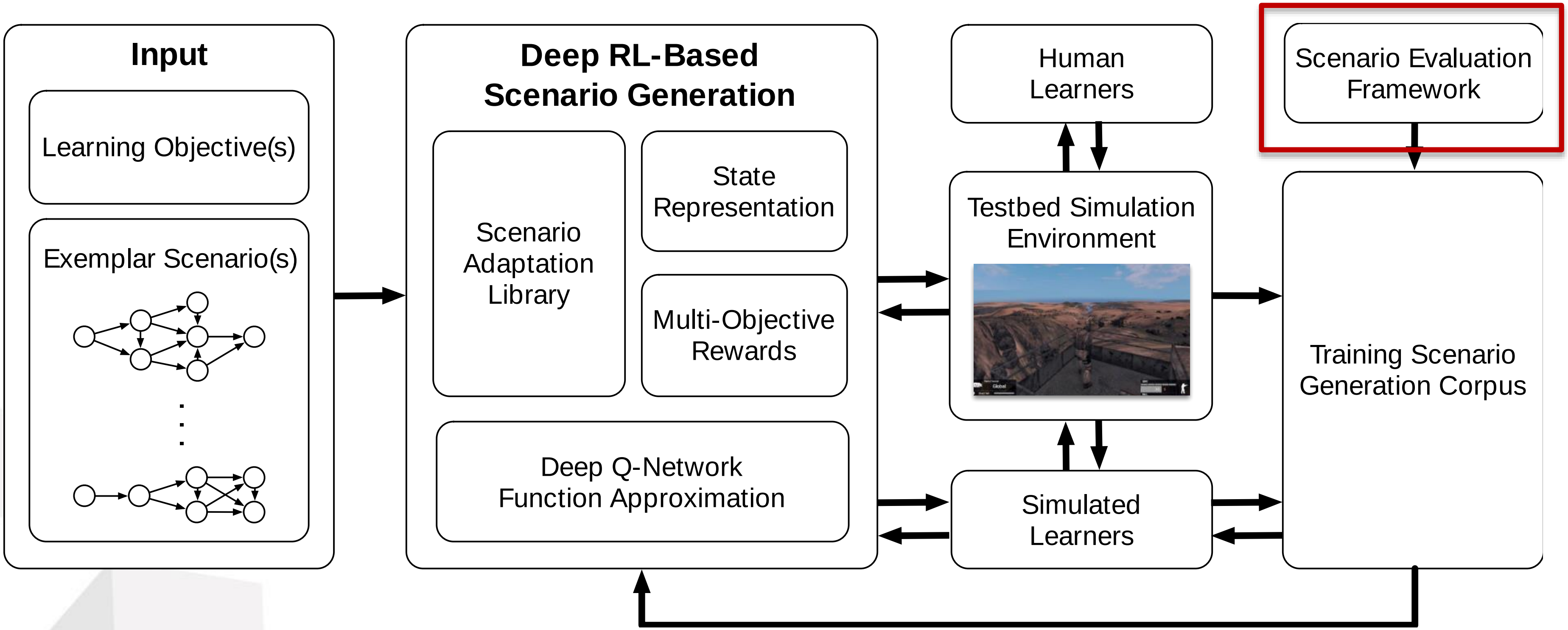
Deep RL-Based Scenario Generation Framework



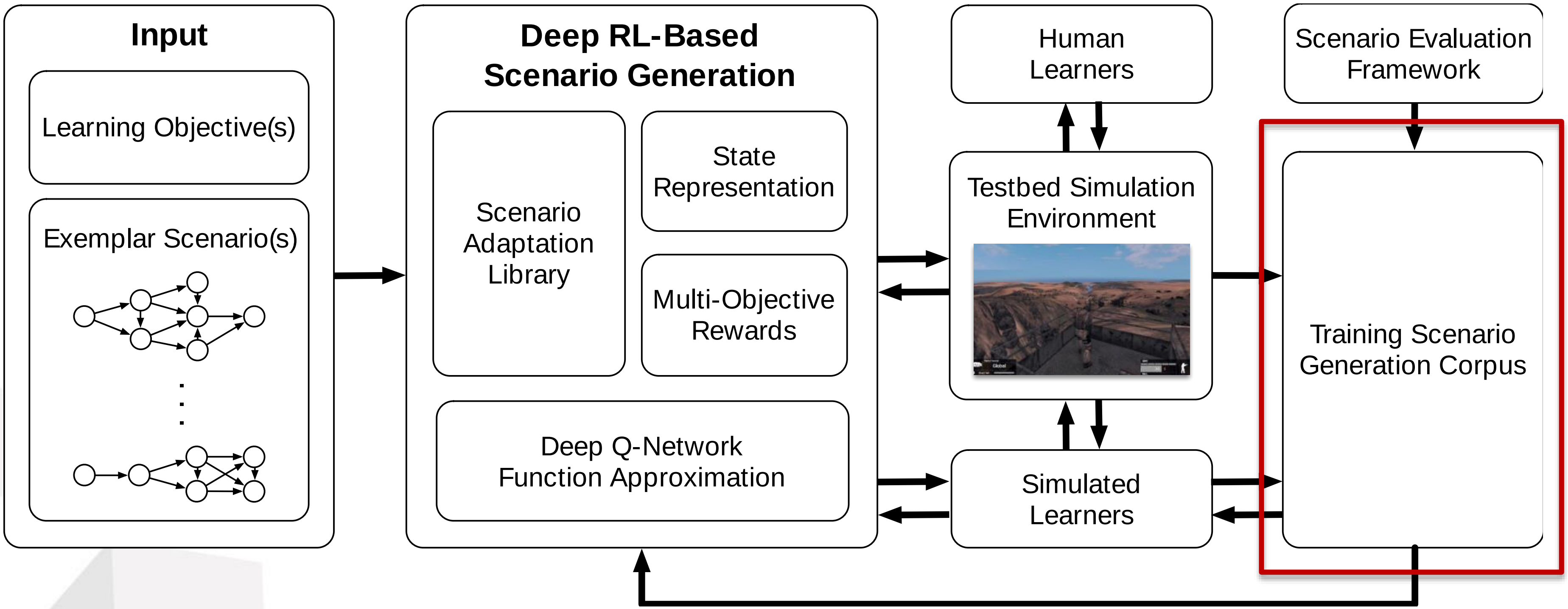
Deep RL-Based Scenario Generation Framework



Deep RL-Based Scenario Generation Framework

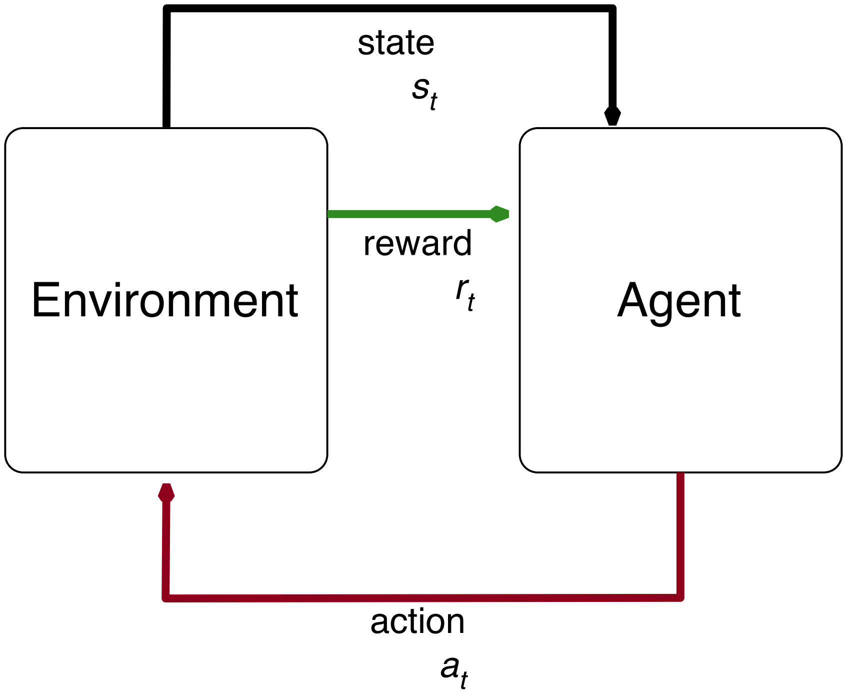


Deep RL-Based Scenario Generation Framework



Reinforcement Learning

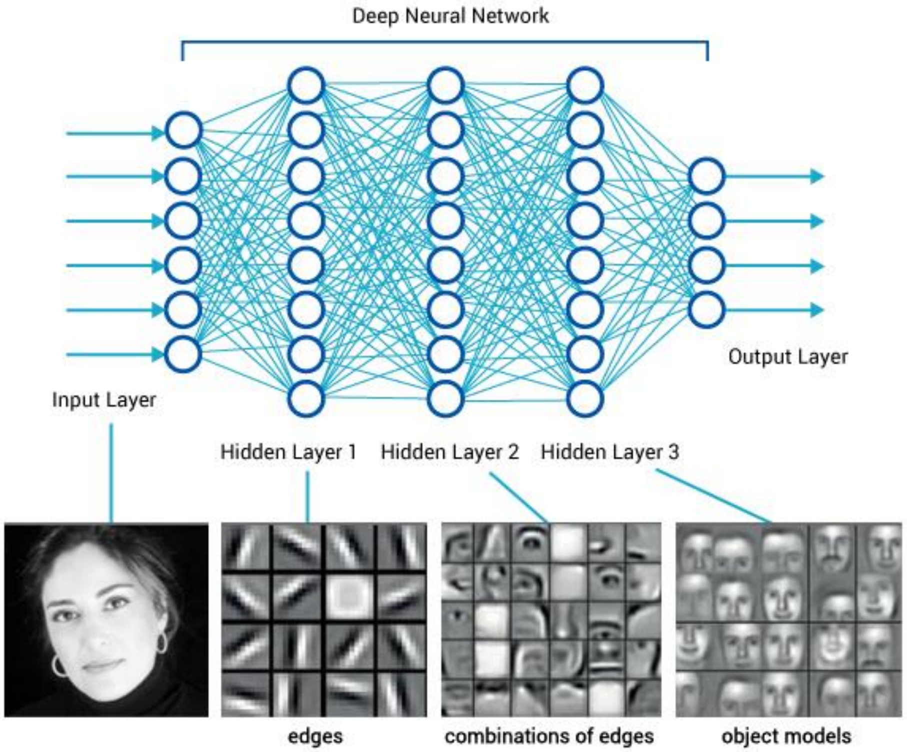
- **Problem:** Devise software agent that learns how to behave in order to maximize numerical reward
- No external supervision
- Delayed rewards



Adapted from Sutton & Barto (1998)

Deep Learning

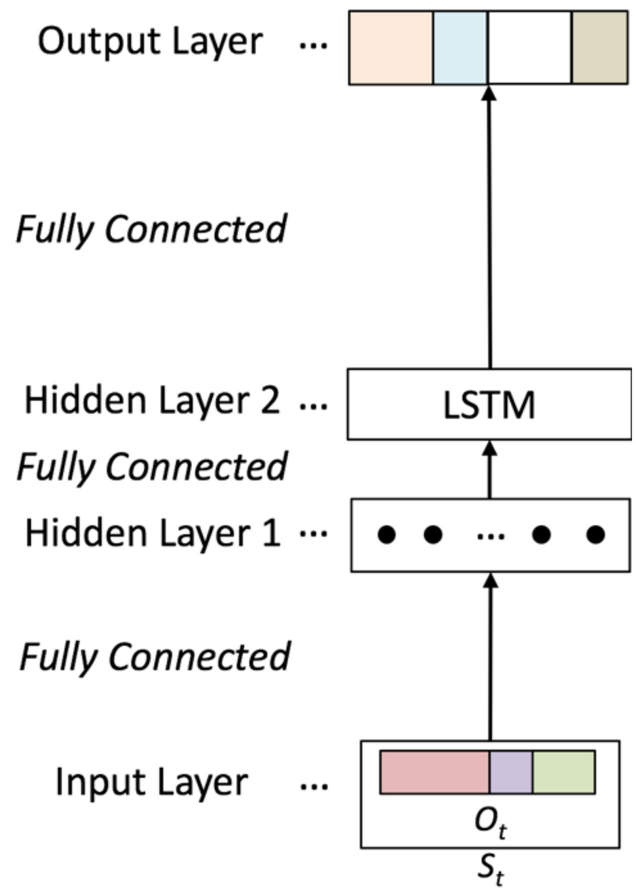
- Family of machine learning techniques for modeling hierarchical representations of data
- Used for multi-level feature learning and pattern classification
- Deep neural networks
  - Multi-layer perceptrons
  - Recurrent neural networks
  - Convolutional neural networks



https://catalystsecure.com/blog/2017/07/deep-learning-in-e-discovery-moving-past-the-hype/

Deep Q-Network Based Scenario Generation

- Derive scenario generation policy that maximizes accumulated rewards $R_t = \sum_{\tau=t}^{\infty} \gamma^{\tau-t} r_\tau$
- Input is the state $s_t \in S = (o_{t-n+1}, \cdots, o_t)$
- Actions defined by scenario adaptation library
- Multi-objective reward signal
- Output is $Q_\pi(s_t, a_t)$



Interactive Narrative Personalization with Deep Reinforcement Learning (IJCAI-2017)

Synthetic Training Data

- Large amount of data is ideal for deep RL
- Generate synthetic data for training deep RL-based scenario generation models using simulated learners
- Leverage multi-dimensional design framework for simulated learners (Rowe et al., 2017)
    - Representational granularity
    - Computational framework
    - Model complexity
    - Learning process
    - Model validity
- Configure a range of simulated learners to investigate how to tailor run-time scenario generation

Outline

- Background and Related Work
- Deep RL-Based Scenario Generation
- **Virtual Battlespace 3 Testbed Environment**
- Preliminary Results
- Demo
- Conclusions and Future Work

Virtual Battlespace 3

- Popular simulation platform for small-unit training
- Developed by Bohemia Interactive Simulations
- Integrated with GIFT
- Provides developer tools for scenario/mission editing
- **Initial Task Domain:**
  Call for Fire Training

Call for Fire Training

- Infantry calls for indirect fire from supporting artillery:
  - Forward Observer(FO) identifies target
  - FO signals artillery battery (AB)
  - AB fires at target location
  - FO sends targeting adjustments (repeated)
  - FO ends firing mission
- Realized in VBS3 using SimCentric VBSFires Add-on

Outline

- Background and Related Work
- Deep RL-Based Scenario Generation
- Virtual Battlespace 3 Testbed Environment
- **Preliminary Results**
- Demo
- Conclusions and Future Work

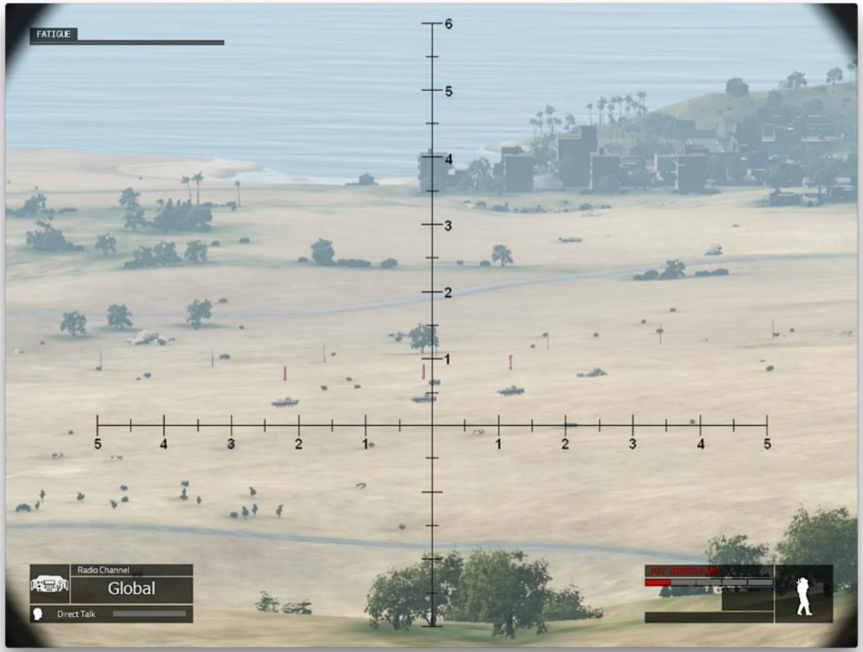
Multi-armed Bandit

Stochastic multi-armed bandit problem (Agrawal & Goyal, 2012)

- Given a slot machine with *N* arms
- At each time step, one of the *N* arms is pulled
- Each arm, when played, yields a random real-valued reward according to a fixed unknown distribution [0, 1]
- The random rewards are i.i.d. and independent of other arms
- Reward is observed immediately after pulling an arm

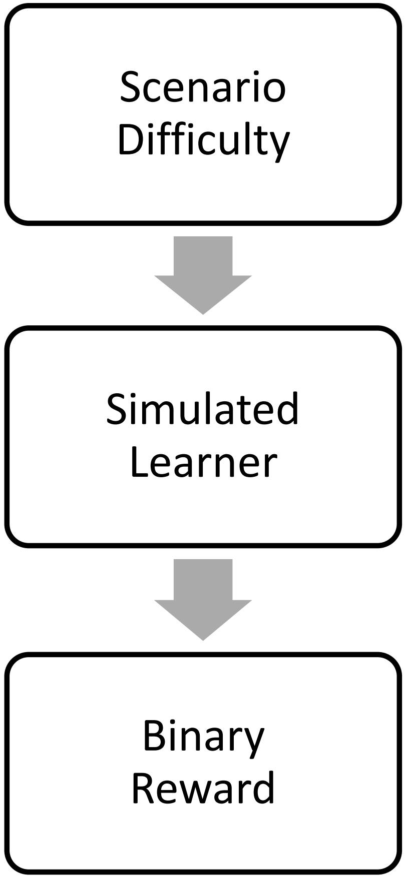
Scenario Adaptation Library

- Specify initial conditions of Call for Fire task:
  - **Weather**: Clear, Cloudy, Raining
  - **Time of day**: Day, Dusk, Night
  - **Type of target**: Stationary, Moving
- Conditions affect difficulty of CFF task
- Each combination of conditions is an arm of the MAB (18 possible scenarios)



Simulated Learner

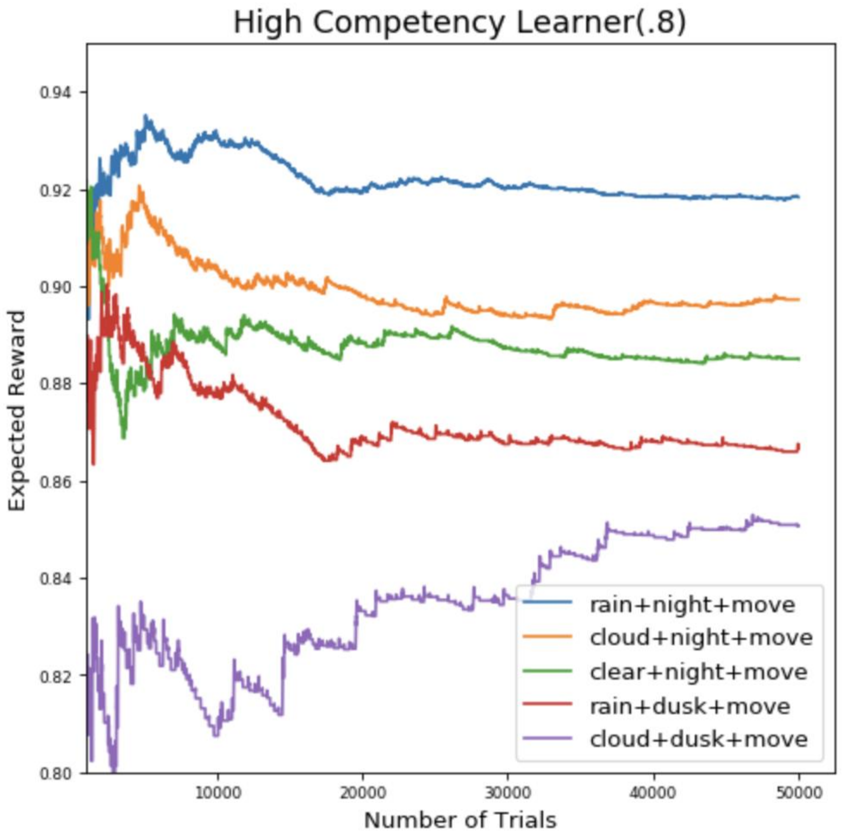
- **Competency Score:** Simulated learner's prior knowledge is represented on the range [0,1]
- **Difficulty Level:** Scenarios are assigned a difficulty [0,1] based on combined initial conditions
- **Reward:** Generated stochastically based on scenario difficulty and learner competency
  - Binary indicator of effectiveness of training scenario for learner
  - Sampled according to binomial distribution

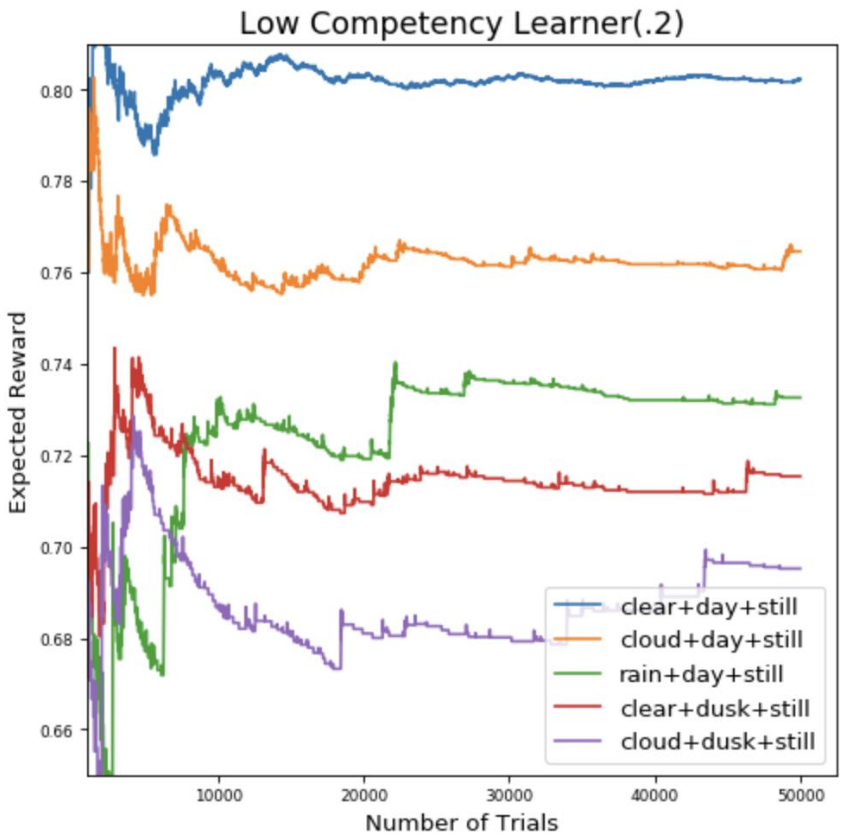


Simulation Setup

- An 18-armed bandit was constructed to cover all possible combinations of initial conditions
- Two simulations were run with different populations of simulated learners
  - High competency learners (*M=.8, SD = .1)*
  - Low competency learners (*M=.2, SD = .1)*
- For each simulation 50,000 trials were run with a different randomly generated learner for each trial
- UCB1 algorithm was used to determine which arm to pull (i.e. manage exploitation/exploration)

Simulation Results





Video

SCENARIO GENERATION DEMO

VBS3 Call for Fire

Discussion

- Approach provides ranking of scenarios based on population of users
- Leverage data from human learners to induce and validate simulated learner models
- Plan to investigate methods for incorporating multiple objectives into the reward function
- Investigate alternative algorithmic techniques to improve efficiency (fewer trials)

Conclusions

- Automated scenario generation will play a central role in simulation-based training.
- Deep RL shows significant promise for data-driven automated scenario generation.
- We are investigating deep RL-based scenario generation for Call for Fire training in the VBS3 simulation environment.
- Preliminary results with multi-armed bandits provide an initial demonstration of data-driven scenario generation that produces a ranked ordering of generated training scenarios.

Future Directions

- Expand scenario adaptation library to capture a broader range of transformations to exemplar training scenarios.
- Expand formulation of automated scenario generation beyond initial scenario conditions and explore deep RL techniques.
- Investigate richer simulated learner models to serve as a bootstrapping mechanism for scenario generation.
- Investigate multi-objective rewards to account for tradeoffs between learning objectives during scenario generation.

Acknowledgements

Collaborators

- Bita Akram (*Computer Science*)
- Kirby Culbertson (*Art & Design*)
- Amanda Dean (*Art & Design*)
- Andrew Emerson (*Computer Science*)
- Oriana Johnson (*Education*)
- Dawoda Kah (*Art & Design*)
- Matthew Lee (*Army SME*)
- Barry Liu (*Computer Science*)
- Robert Taylor (*Computer Science*)
- Wookhee Min (*Computer Science*)
- Pengcheng Wang (*Computer Science*)
- Robert Sawyer (*Computer Science*)

Support

ARL cooperative agreement W911NF-15-2-0030

Contact

jprowe@ncsu.edu