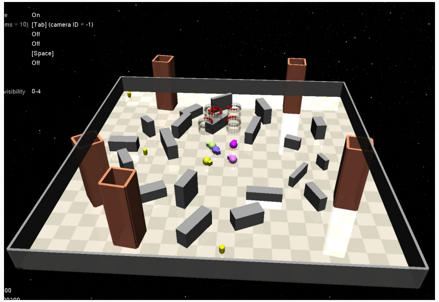
Abdullah Khaled Weekly Status Report

Weeks of June 29, 2026
& July 6, 2026

BOSTON UNIVERSITY

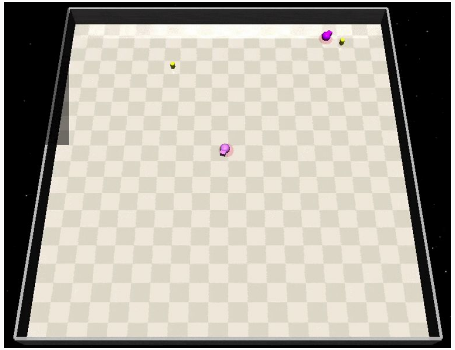
Environment Creation

- Conducted background research on different SAR environments, and used that information to create a mock-up of my own
- Created new objects for the custom env and new spawning behaviors for these objects
- Implemented pseudo-occluded lidar so agent is aware of object locations, but not if masked
- Made multiple levels that increase in difficulty (for initial testing purposes)



PPO Training

- Imported Stable-Baselines3 into project and translated step returns into SB3-usable formats
- Started reward engineering to improve model capabilities
- Adjusted rewards based on model "reward hacking" (e.g. circling casualties)

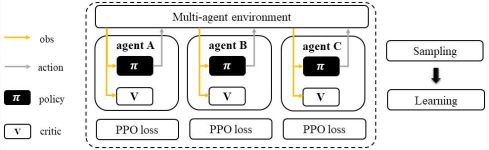


Challenges

- Creating new lidar_obs type meant refactoring large chunks of code
- Steep initial learning curve, but I am learning quickly
- Agent doesn't go for rewards, stays in relative safety

Future Direction

- Separate agent trajectories (currently combined for SB3 PPO integration) i.e. IPPO
- Train model on more complex environments, adjusting params along the way
- Use LTL or other high-level planner to coordinate and update goals



Abdullah Khaled Weekly Status Report Ctd.

Weeks of June 29, 2026
& July 6, 2026

BOSTON UNIVERSITY

Environment Creation - Daily Tasks

06/30 - Obtained approval to start working on SpecRLBench for SAR env
07/01 - Set up SpecRLBench on local machine
07/02 - Read 7 papers before creating mock-up of custom environment
07/03 - Created walls, ringed placements, buildings, border placements
07/04 - Incorporated vision, created casualties
07/05 - Added custom name for environment
07/06 - Designed agent spawn params, added building walls, and did some debugging
07/06 - Overall environment completed
07/07 - Added keyboard movements for debugging, added pseudo-occluded vision type
07/09 - Added three levels to SAR tasks (+1 existing) for N=5 SAR environments total

PPO Training - Daily Tasks

07/07 - Started thinking about/implementing rewards/punishments for buildings, casualties, collisions
07/08 - Migrated code to DCL GitHub repo, set up SpecRLBench in remote desktop
07/08 - Set up PPO example project and learned about PPO and MA variations (e.g. IPPO, MAPPO) and previous implementations in SAR environments
07/09 - Integrated SAR environment into SB3's PPO model
07/09 - Started adjusting rewards/punishments based on model quality
07/10 - Added object-lidar pairs for casualty targeting in the future

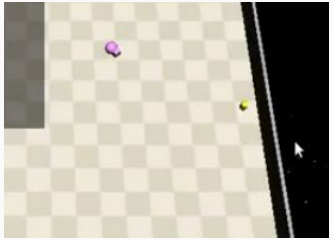
Abdullah Khaled Weekly Status Report Ctd.

Week of July 13, 2026

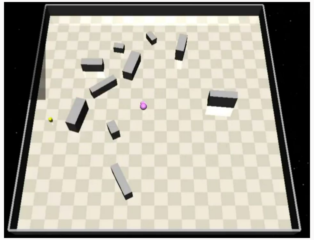
BOSTON UNIVERSITY

Increasing Task Complexity

Single Agent and Surface Casualty (SASC)
100% Rescue Rate



SASC with Obstacles (N=10)



Single Agent Entrapped Casualty (SAEC)

SAEC with Distractor Building

Task Construction Rationale

- **SASC with Obstacles**: This was the simplest route to increasing the complexity of the environment, but was POMDP
- **SAEC**: Gave agent information akin to fully observed MDP
- **SAEC with Distractor**: Increase complexity while still having an environment closer to a MDP

Intrinsic Rewards - RND PPO

- After Dr. Li talked to me about exploration-driven heuristics and intrinsic rewards, I read papers about different intrinsic rewards with PPO and found RND
- I implemented it based on previous research with the help of AI and tested on SASC with Obstacles and achieved 96.0% success
- Tested on SAEC with Distractor and achieved 80.0% success

Safe Environments - Wall Collisions

- Once I saw policies with high accuracy, I implemented wall collisions (WC)
- After showing the results (see table) to Zijian, I plan to try implementing Lagrangian PPO so policy emphasizes costs too, since Vanilla PPO

Abdullah Khaled Weekly Status Report

Week of July 13, 2026

BOSTON UNIVERSITY

Testing Results (n = 1)

Environment	Vanila PPO	RND PPO
SASC w/Obstacles	54%	96%
SAEC	92%	90%
SAEC w/ Distractor	36%	80%
SASC w/ Obstacles/WC	92%	82%
SAEC w/ WC	62%	66%
SAEC w/ Distractor/WC	0%	22%

Challenges

- At first, I had a lot of difficulty understanding the different hyperparameters, but the more trials I tried on the SASC with Obstacles environment (N=56), I learned a lot more
- Implementing more complex PPO variations and researching them has also been a fun challenge, and I feel like I have been exposed to many different types of models

Future Directions

- Implement Lagrangian PPO using pre-built library and test on WC tasks
- Use these variations with IPPO to incorporate multi-agent systems
- Continue testing/improving model performance through hyperparameter tweaks

Milestones

06/30 - SpecRLBench SAR env approval
07/06 - Overall environment completed
07/09 - Integrated SAR environment into SB3 PPO
07/12 - First successful PPO policy training run
07/15 - Pivot to entrapped casualty tasks
07/16 - Implemented and tested RND-PPO
07/17 - Added accurate wall collisions

Abdullah Khaled Weekly Status Report

Week of July 20, 2026

BOSTON UNIVERSITY

Continued Model Testing

- Continued testing hyperparameters to improve accuracy of different models (i.e. PPO, PPO-Lagrangian, RND)
- Added additional models to testing based on OpenAI's Safety Starter Agents GitHub repository: **TRPO, TRPO-Lagrangian, SAC, SAC-Lagrangian**
- Lagrangian variations had to be custom combination of SB3 base models with Lagrangian logic added on top
- Results of testing are shown to the right. For simplicity, naming conventions are:

SASC: L0
SASC w/Obstacles: L1
SAEC w/ Obstacles: L2
SAEC w/ Obstacle & Distractor: L3

Note: When wall collisions enabled, 'WC' is appended to task name

Algo	L0	L1	L1WC	L2	L2WC	L3	L3WC
PPO	100	98	86	98	68	94	×
PPO-Lag	×	×	94	×	74	×	12
TRPO	100	96	88	98	78	92	52
TRPO-Lag	×	×	84	×	54	×	18
SAC	98	74	×	×	×	×	×
SAC-Lag	×	×	×	×	×	×	×

SB3 Emigration to Safe PO

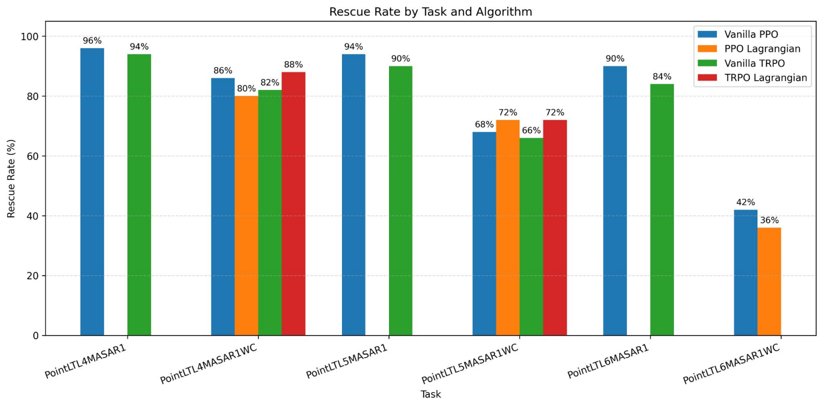
- Currently, SB3 had integrity issues:
    - Custom implementation for Lagrangian variations, RND
    - Would require more custom code for multi-agent algorithms
- Therefore, I chose to switch to Safe PO (developed by the creators of Safety-Gymnasium) because of its native Lagrangian and multi-agent implementations
- No RND implementation, but RND module only adds intrinsic reward without large code diffs like Lagrangian required

Abdullah Khaled Weekly Status Report Ctd.

Week of July 20, 2026

BOSTON UNIVERSITY

Current Safe PO Progress



- After a lot of hyperparameter tuning, achieved comparable results to SB3

Milestones

07/21 - Safe PO Migration
07/24 - Single-agent algorithms validated and evaluated, multi-agent evaluation begins

06/30 - SpecRLBench SAR env approval
07/06 - Overall environment completed
07/09 - Integrated SAR environment into SB3 PPO
07/20 - Final SB3 training runs

Challenges

- Deciding between SB3, OmniSafe, and Safe PO required a lot of comparison between strengths/weaknesses of each
- Safe PO migration meant refactoring large portions of code, but less difficult because of Safety-Gymnasium compatibility
- Had to add SB3-only hyperparameters (e.g. entropy coefficient, learning rate decrease factor) for similar performance

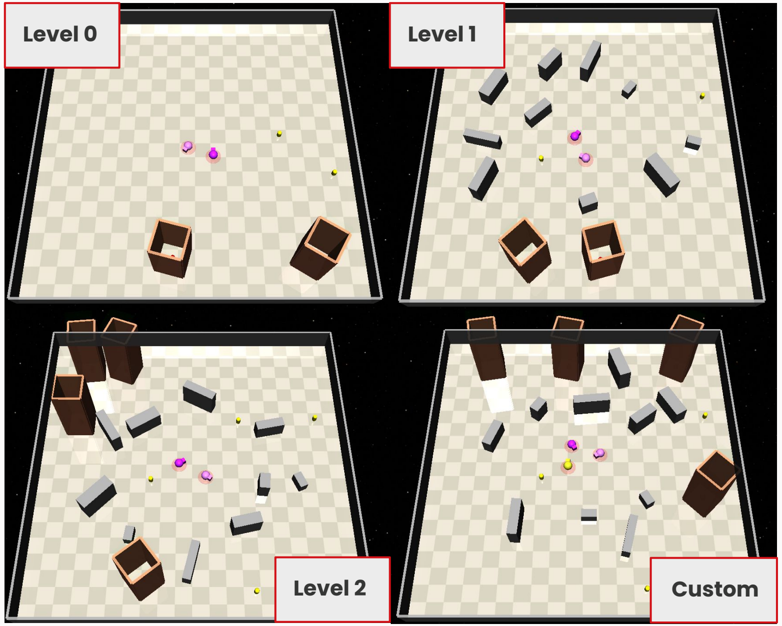
Future Directions

- Create and test multi-agent tasks in Safe PO
- Evaluate different MA algorithms on environments
- Begin abstract and poster

Abdullah Khaled Weekly Status Report

Weeks of July 27, 2026
& Aug 3, 2026

BOSTON UNIVERSITY



Repository Organization

- Used submodules for SpecRLBench, Safe-Policy-Optimization, and GenZ-LTL that link to personal forks to isolate changes

Environment Refactoring

- Realized that previously, **MultiGoalSAR** was actually single goal (i.e. entrapped or surface casualty only)
- Refactored L0 and L1 to **SingleGoalSAR** to reflect task
- Added three MultiGoalSAR levels (per-agent numbers):
    - **Level 0**: One surface casualty and one entrapped casualty
    - **Level 1**: Level 0, but with walls
    - **Level 2**: Level 1, but with two surface/entrapped casualties
- Created YAML file **CustomizedSAR** environment that allows user to customize major features of environment

GenZ-LTL Addition to Project

- Now that I had demonstrated that Safe PO works, Zijian recommended me to start using GenZ-LTL
- Propositions for environment: **{entrapped, surface, walls}**
- Specification for environment: **"(!(walls | surface) U entrapped) & (!surface U entrapped)**

Abdullah Khaled Weekly Status Report Ctd.

Weeks of July 27, 2026
& Aug 3, 2026

BOSTON UNIVERSITY

Pre-Poster Results

	PPO			PPO-Lagrangian			GenZ-LTL		
	$\eta_s \uparrow$	$\eta_v \downarrow$	$\mu \downarrow$	$\eta_s \uparrow$	$\eta_v \downarrow$	$\mu \downarrow$	$\eta_s \uparrow$	$\eta_v \downarrow$	$\mu \downarrow$
Level 0	$0.070_{\pm 0.102}$	$0.34_{\pm 0.31}$	$1157.83_{\pm 181.63}$	$0.272_{\pm 0.139}$	$0.16_{\pm 0.06}$	$905.78_{\pm 72.78}$	$\mathbf{0.778}_{\pm \mathbf{0.045}}$	$\mathbf{0.09}_{\pm \mathbf{0.07}}$	$\mathbf{805.24}_{\pm \mathbf{54.03}}$
Level 1	$0.004_{\pm 0.005}$	$\mathbf{0.39}_{\pm \mathbf{0.30}}$	$\mathbf{631.00}_{\pm \mathbf{8.49}}$	$0.008_{\pm 0.013}$	$0.62_{\pm 0.14}$	$1714.35_{\pm 770.25}$	$\mathbf{0.132}_{\pm \mathbf{0.068}}$	$0.59_{\pm 0.09}$	$1083.06_{\pm 199.89}$

η_s= Success Rate, η_v= Violation Rate, μ=Mean Successful Episode Length

- GenZ-LTL significantly outperforms traditional models in both environments, though all models struggle in Level 1.
- GenZ-LTL had higher μ than PPO.
- GenZ-LTL has lower η_v on Level 0 but highest on Level 1.
- η_v and μ likely skewed due to highly infrequent successful episodes.

Milestones

06/30 - SpecRLBench SAR env approval
07/06 - Overall environment completed
07/09 - Integrated SAR environment into SB3 PPO
07/21 - Safe PO Migration

Challenges

- Although success seen at Level 0, there appears to be a gap that exists at Level 1 that prevents even GenZ-LTL from learning effectively (likely env bug)
- η_v unusually high for GenZ-LTL; will look into

Future Directions

- Investigate Level 1 for bugs and/or modify to improve observability
- Investigate violation rate and how to improve environment/policy to prevent it

Final Remarks

I want to thank Dr. Li and Zijian for their incredible support, and I look forward to continuing our collaboration!

07/27 - Safe PO validation, GenZ-LTL start
08/01 - GenZ-LTL success on L0
08/04 - Poster submission!