PHD PROPOSAL DEFENSE

Uncertainty-Aware Reinforcement Learning for Safe Autonomous Field Robots

A Proposal for Doctoral Research

Maya Chen

PhD Candidate, Department of Computer Science, Meridian University

Committee: Prof. Elena Vasquez (chair) ▪ Prof. Daniel Osei ▪ Prof. Priya Nandakumar

March 12, 2027

Agenda

Roadmap for This Proposal

Motivation

Gap and Related Work

Methodology

Preliminary Results

Timeline

Contributions and Risks

The Problem

Learned Policies Meet Unstructured Terrain

The Failure Mode

Policies trained in the **Nexa** simulator encounter terrain statistics at deployment that the training distribution never covered.

Where It Bites

Agricultural inspection and disaster-response rovers built on the **Vertex** platform must act under this mismatch *without* a human in the loop.

The Cost

Overconfident actions on out-of-distribution terrain produce near-misses and, in field trials, outright rollovers.

Central claim of this proposal

A field robot's policy should know *when its own dynamics model is guessing* – and slow down, not merely act on the mean prediction.

Research Aims

Three Aims, One Deployable System

Aim 1

Design a **calibrated, onboard-feasible** uncertainty estimator for learned terrain-interaction dynamics models.

Aim 2

Develop a **risk-aware policy optimization** framework that trades exploration against a formal safety budget.

Aim 3

Validate in the field on the Vertex rover across three terrain classes at the Meridian Outcrop Test Range.

Aims 1–2 are trained and benchmarked in simulation before any hardware commitment.

Aim 3 closes the loop: the same uncertainty estimate that shapes the policy in simulation is the one logged on the physical rover.

Where Existing Work Falls Short

No Prior Approach Clears All Three Bars

Approach	Uncertainty	Onboard	Field
Model-free RL [1]	×	✓	✓
Ensemble-based UQ [2]	✓	×	×
Constrained policy opt. [3]	×	✓	×
Proposed approach	✓	✓	✓

Reading the table

Ensemble methods estimate uncertainty well but need 5–10 forward passes per action – too slow for the rover's onboard compute. Constrained policy methods run onboard but treat the dynamics model as exact, so the safety margin is only as good as an unmodeled sim-to-real gap.

Related Work

Three Threads This Proposal Combines

- **Sim-to-real transfer.** Vasquez [4] documents the gap this proposal targets: policies tuned in **Nexa** degrade sharply once terrain friction and compliance leave the training envelope.
- **Uncertainty quantification.** Garcia [2] shows evidential deep learning gives single-pass, calibrated variance estimates – the property Aim 1 builds on, absent the ensemble's compute cost.
- **Safe policy optimization.** Osei [3] and Nandakumar [1] formalize constrained MDPs for field robots, but both assume the risk signal is already given – Aim 2 supplies that signal from Aim 1's estimator.

The gap

No published system couples a *calibrated, single-pass* uncertainty estimate directly into the constraint of a field-deployed policy, then validates the coupling outdoors. Preliminary work [5] took the first step in simulation only.

Research Design

Three Phases, Each Gated on the Last

Phase 1

Calibrated UQ. Train an evidential dynamics model on Nexa rollouts; calibrate against held-out terrain shift.

Phase 2

Risk-aware policy. Fold the Phase 1 estimate into a CVaR-constrained policy optimizer; benchmark in simulation.

Phase 3

Field validation. Deploy on the Vertex rover at the Meridian Outcrop Test Range across three terrain classes.

Each phase has a pre-registered go/no-go metric (Phase 1: calibration error < 0.05; Phase 2: constraint violation rate $< 2\%$ in simulation) before hardware time is committed.

Phases 1–2 share the same codebase and log format as Phase 3, so no reimplementation happens at the sim-to-real boundary.

Phase 1: Calibrated Uncertainty Quantification

Single-Pass, Onboard-Feasible

Replace the point-estimate dynamics head with an **evidential** output layer (Normal-Inverse-Gamma), giving mean, aleatoric, and epistemic variance in one forward pass.

Train on 40k Nexa rollouts spanning gravel, mud, and debris contact models.

Calibrate epistemic variance against a held-out *terrain-shift* split, not the training distribution.

Distill to a 3-layer MLP sized for the rover's Jetson-class onboard compute – target < 5 ms per query.

Go / no-go

Expected calibration error < 0.05 on the shift split, at < 5 ms/query on target hardware.

Phase 2: Risk-Aware Policy Optimization

A Formal Safety Budget

CVaR-constrained objective

$$\max_{\theta} \; \mathbb{E}_{\tau \sim \pi_\theta}\left[\sum_{t=0}^{T} \gamma^t \, r_t\right] \quad \text{s.t.} \quad \mathrm{CVaR}_\alpha\big[\, C(\tau) \,\big] \;\le\; \delta$$

$C(\tau)$ is a per-episode cost built from Phase 1's epistemic variance: the policy pays for acting confidently where the model is unsure.

δ is set from the rover's mechanical tolerance (maximum permissible roll/pitch before a stability margin is lost).

Solved with a Lagrangian relaxation on top of a trust-region policy update, reusing the actor-critic backbone from [5].

Phase 3: Field Validation Protocol

Meridian Outcrop Test Range

Terrain class	Primary hazard	Trials
Loose gravel incline	Loss of traction	30 runs
Flooded furrow field	Submerged obstacles	30 runs
Post-storm debris field	Unmapped occlusions	30 runs

Each terrain class is run under both the Phase 2 policy and a constrained-but-uncertainty-blind baseline, same random seeds where the simulator permits matched conditions.

Safety officer holds a hardware e-stop for every run; no trial is scored if the e-stop is used.

Outcomes logged as the *Outcrop-3* benchmark, released with the dissertation [6].

Pilot Study

Simulation Results, Nexa Terrain Suite, $n = 200$ Episodes/Cell

Environment	Method	Success (%)	Near-miss (%)	Latency (ms)
Loose gravel incline	Baseline	71.2	18.4	42
	Proposed	89.6	6.1	47
Flooded furrow field	Baseline	64.8	24.0	39
	Proposed	85.3	8.7	45
Post-storm debris field	Baseline	58.0	31.5	41
	Proposed	80.9	11.2	46

Baseline: constrained policy optimization without uncertainty-shaped cost (reimplementation of [3]). Proposed: Phase 1 estimator feeding the Phase 2 CVaR constraint.

What the Pilot Suggests

And What It Cannot Yet Tell Us

Encouraging

Near-miss rate drops by more than half in every environment, largest in the debris field where terrain shift is greatest.

Latency overhead of the evidential head is 5–6 ms – inside the Phase 1 go/no-go budget.

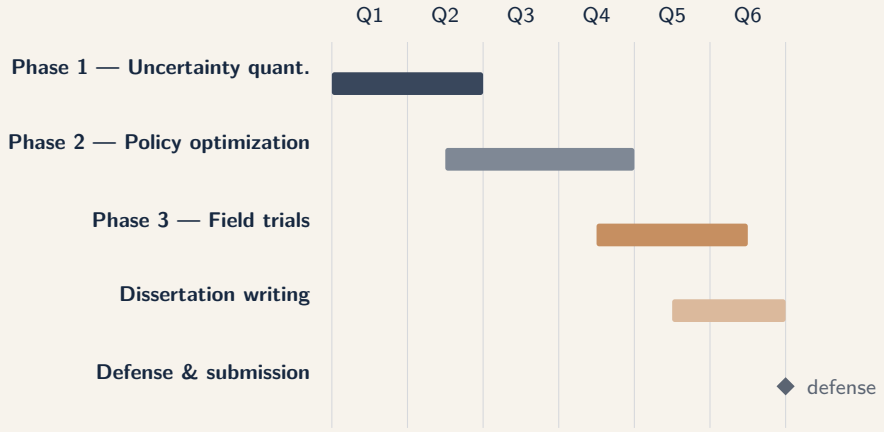
Still open

These numbers are simulation-only; Aim 3 is precisely the test of whether the gain survives the sim-to-real gap.

Pilot used a single seed family – the full study pre-registers five seeds per cell before Phase 2 is called complete.

Proposed Timeline

Six Quarters to Defense



Milestones and Deliverables

What Ships at Each Gate

- **End of Q2:** Phase 1 calibration report; go/no-go review with committee.
- **End of Q4:** Phase 2 simulation benchmark submitted to the Apex Journal of Autonomous Systems.
- **End of Q5:** Outcrop-3 field data collection complete; benchmark suite frozen.
- **Q6:** Dissertation draft to committee four weeks before the scheduled defense date.

Expected Contributions

Three Claims, Three Targets

C1

A calibrated, onboard-feasible uncertainty estimator for learned terrain-interaction dynamics.

Target: Nimbus Robotics Conference

C2

A risk-constrained policy optimization framework with a formal CVaR safety bound.

Target: Apex Journal of Autonomous Systems

C3

The Outcrop-3 field benchmark and validation protocol, released openly.

Target: Synapse Workshop on Safe Learning

Risks and Mitigations

Named Now, Not Discovered in Q5

Risk	Likelihood	Mitigation
Vertex rover fleet time is shared across labs	Medium	Quarterly blocks reserved with the Meridian Core Facility
Evidential estimator miscalibrates under domain shift	Medium	Conformal recalibration layer added before Phase 3
Weather windows at the Outcrop Test Range	High	Three-week buffer per trial block; indoor proxy course as fallback
Onboard compute budget on the Jetson-class unit	Low	Early profiling; distilled 3-model ensemble as fallback

Summary

What This Proposal Asks the Committee to Approve

A dynamics model that reports *when it is guessing*, cheap enough to run onboard.

A policy that spends a formal, mechanically-grounded safety budget instead of an ad hoc heuristic.

A field test at the Meridian Outcrop Test Range that either confirms the simulation gains transfer, or tells us precisely where they do not.

Thank you.

Questions and discussion welcome.

References I

[1] Priya Nandakumar. “Model-Free Reinforcement Learning under Terrain Uncertainty: A Survey”. In: *Apex Journal of Autonomous Systems* 14.3 (2022), pp. 112–134.

[2] Sofia Garcia. “Evidential Deep Learning for Robust Dynamics Models in Field Robotics”. In: *Proceedings of the Nimbus Robotics Conference*. 2024, pp. 210–221.

[3] Daniel Osei. “Constrained Policy Optimization for Risk-Bounded Field Robots”. In: *Synapse Workshop on Safe Learning for Robotics*. 2023, pp. 45–58.

[4] Elena Vasquez. “Sim-to-Real Transfer Gaps in Unstructured Terrain Navigation”. In: *Proceedings of the Nimbus Robotics Conference*. 2021, pp. 301–312.

[5] Maya Chen and Elena Vasquez. “Toward Calibrated Onboard Uncertainty for Field Robot Dynamics”. In: *Synapse Workshop on Safe Learning for Robotics*. 2025, pp. 88–97.

References II

[6] Meridian Robotics Lab. *The Outcrop Test Range: A Field Benchmark for Terrain-Aware Autonomy*. Tech. rep. Meridian University, 2026.