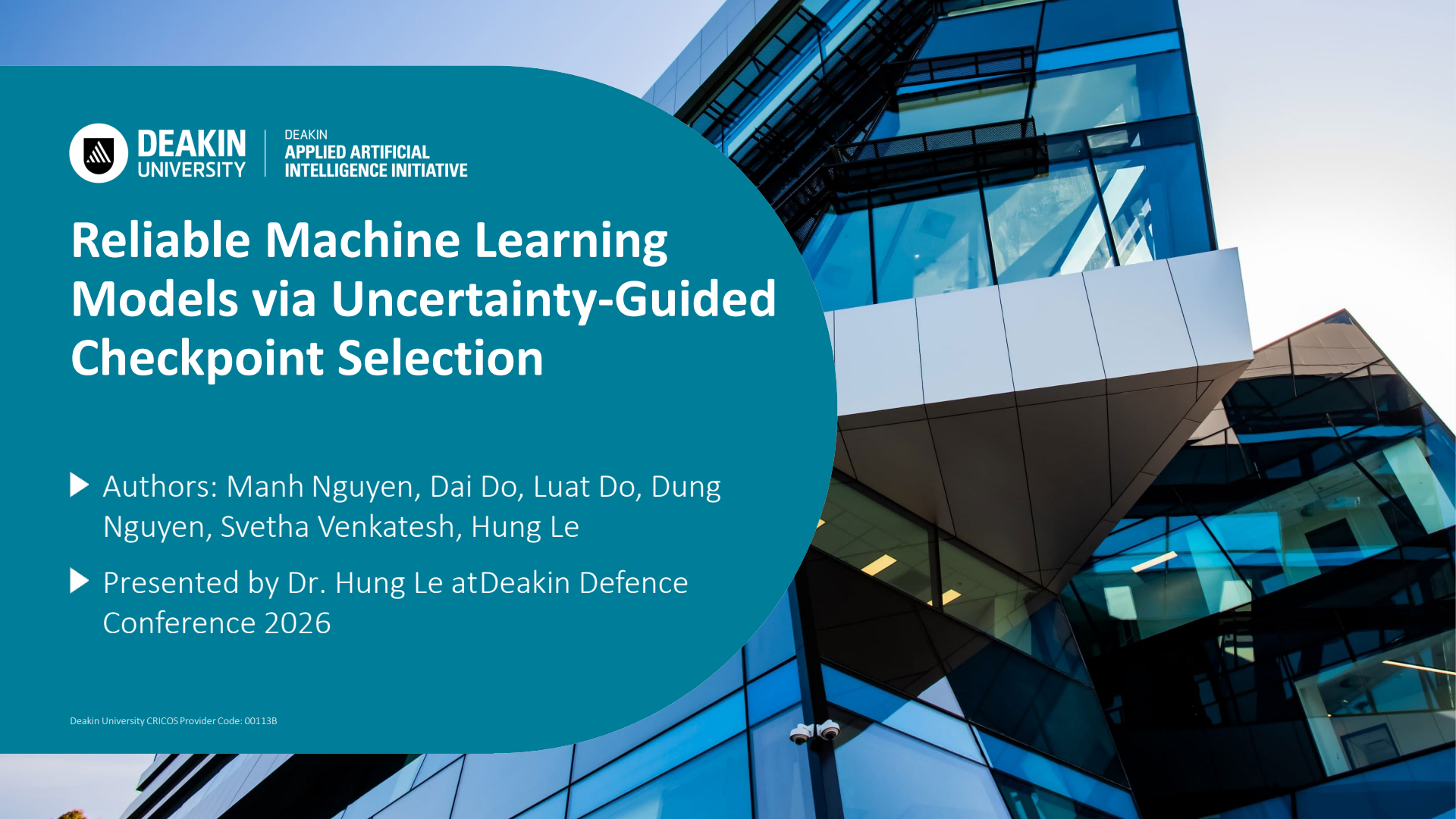


DEAKIN UNIVERSITY | DEAKIN APPLIED ARTIFICIAL INTELLIGENCE INITIATIVE

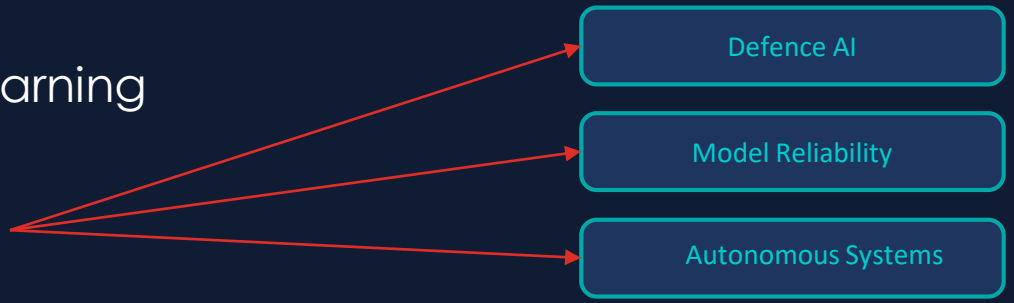
Reliable Machine Learning Models via Uncertainty-Guided Checkpoint Selection

- Authors: Manh Nguyen, Dai Do, Luat Do, Dung Nguyen, Svetha Venkatesh, Hung Le

- Presented by Dr. Hung Le atDeakin Defence Conference 2026

About Us

- Our lab: Applied AI Initiative, Deakin University
- Hung Le is a DECRA fellow, senior lecturer at Deakin University, leading research on deep sequential models and reinforcement learning
- For many years, we have conducted research on the following topics:



The Problem: Selecting the Wrong Model Can Fail Missions

Defence ML Deployment Context

- Autonomous systems & decision support
- Intelligence analysis & cyber operations
- Contested, uncertain environments
- Performance failures introduce operational risk

Unreliable model selection → mission risk

Why Model Selection is Hard

Final checkpoint

May overfit or degrade late in training

Validation-based

Needs labelled data , often unavailable

Manual inspection

Slow, expensive, not scalable

Performance on training average

Masks failure on hard/OOD cases

Our Approach: Uncertainty-Guided Checkpoint Selection (UGCS)

Key Insight: A model that performs well on high-uncertainty, difficult samples is more robust and generalises better to unseen operational conditions.

1. Identify Hard Samples

Rank training instances by predictive uncertainty during training (no extra passes needed)

2. Score Checkpoints

Average performance on “hard” (uncertain) samples over a short training window

3. Select Best Checkpoint

Stable, discriminative ranking that avoids overfitting noise or transient training spikes

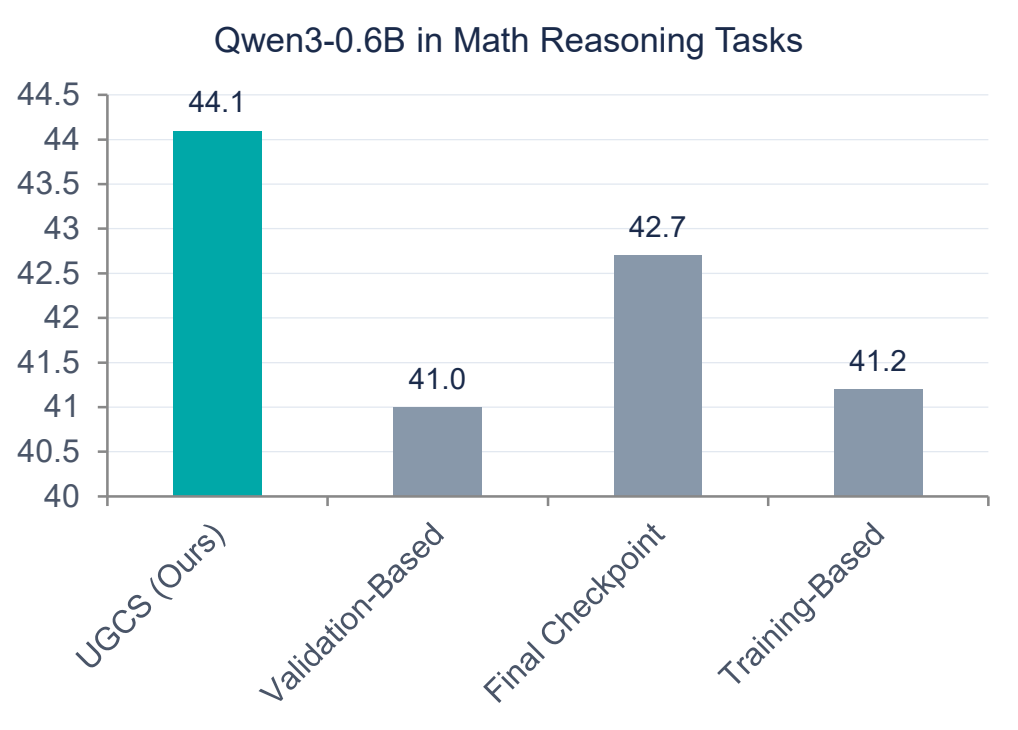
✓ No validation data required

✓ No extra forward passes

✓ Model-agnostic

✓ Negligible overhead

Results: UGCS Consistently Outperforms Baseline Strategies



Key Findings

- Best generalisation across ALL architectures and datasets tested
- Reduces variance in selection → more predictable, safer deployment
- Avoids checkpoints that look good on average but fail on hard cases
- Works with reinforcement finetuning of Large Reasoning Models

Uncertainty-based performance is a reliable proxy for overall model robustness

Relevance for Defence

Trustworthy Autonomy

Reliable model selection reduces risk of autonomous system failures in contested environments

Decision Support

Models selected by UGCS are more robust to edge cases that arise in intelligence and cyber analysis

Resource Constrained Ops

No validation data or extra compute needed → viable under time, data, and hardware constraints

Rapid Deployment Readiness

Model-agnostic approach integrates with existing training pipelines → low barrier to adoption

Open to collaboration with Defence, industry, and government — contact: thai.le@deakin.edu.au

Summary

01

Better checkpoints = better reliability

UGCS selects models that generalise to hard cases, not just average performance

02

Practical for operational contexts

No validation data, no extra compute, low overhead, high impact

03

Ready for Defence translation

Model-agnostic, evaluated on LLMs and classical ML, applicable across autonomy and decision support