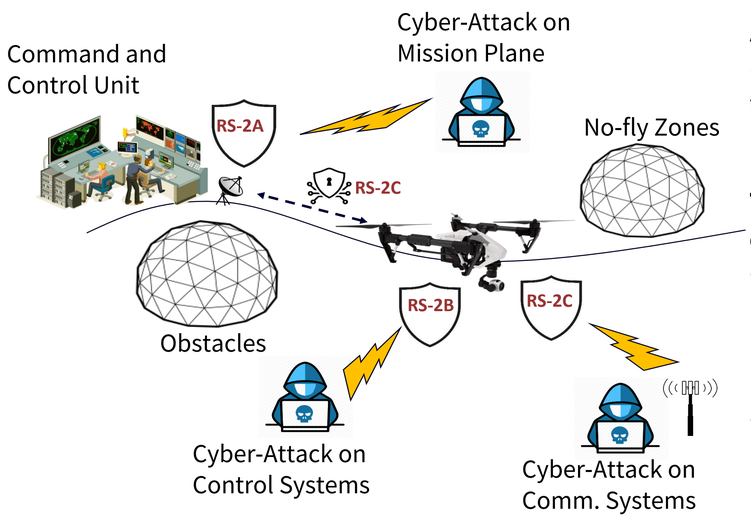
Trustworthy Autonomous Systems
RS-2B Securing the Control Surface

Cranfield University
School of Aerospace, Transport and Manufacturing

Research Fellow: Dr. Burak Yuksek, *burak.yuksek@cranfield.ac.uk*
Investigator: Prof. Gokhan Inalhan, *inalhan@cranfield.ac.uk*

Engineering and Physical Sciences Research Council

RS2: Secure Operations of Autonomous Systems



Autonomous systems face numerous challenges in their operation, due to the uncertain and dynamic multi-layer attack surfaces

TAS-RS2 aims to solve following challenges

- Modelling & addressing potential attacks in discrete mission, control and communication layers
- Study & address hybrid cascaded cross-layer threats in the dynamic AS space

- **RS-2A** • Mission & Operation Plane
  RS-2A: Exposure to cyber-physical attacks by characterizing the attack surfaces, i.e., entry points and likelihoods across the mission surface in a technology & mission-invariant manner.
- **RS-2B** • **Control and Navigation Plane**
  RS-2B: Provide quantifiable safety and feedback to the mission surface when the limits of secure controllability are compromised within a time horizon under current policies and adversarial situations.
- **RS-2C** • Information and Communications Plane
  RS-2C: Provide secure communications across the different layers in the informatics plane from detection of signals to networking.

RS-2B: Securing the Control Surface

Autonomous Systems rely on the ability to conduct run time adaptations of control decisions over attacks or "perceived" attacks:

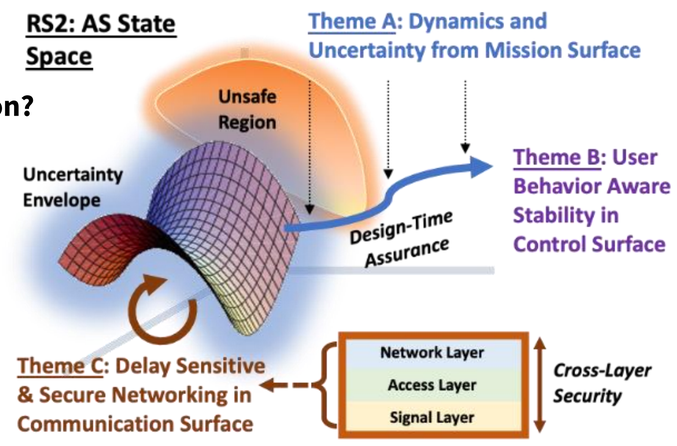
- Adversaries
- Environment uncertainties
- Degraded performance

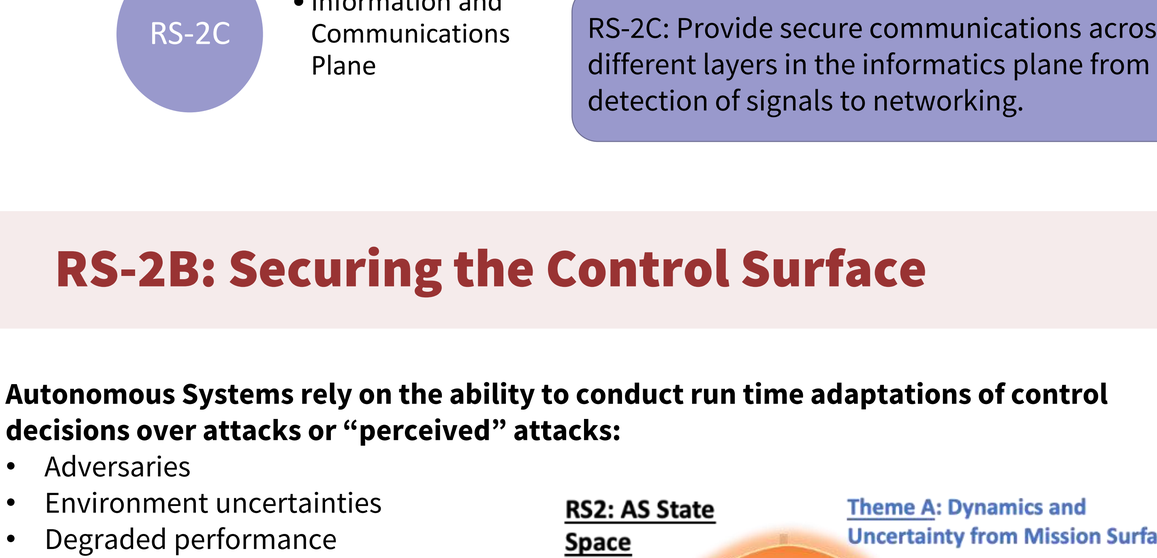
How to do this in a "trustworthy" fashion?

- Safe,
- Secure,
- Reliable

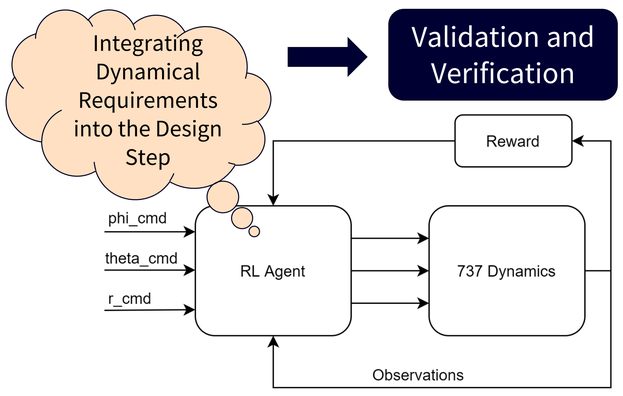
Attack Definitions

- Sensing and Communication Errors
- Loss of an actuator
- Environmental conditions
- Electronic attacks
- Electromagnetic deception
- Injecting false pattern into data





A. Dynamical V&V of the AI-Based FCS

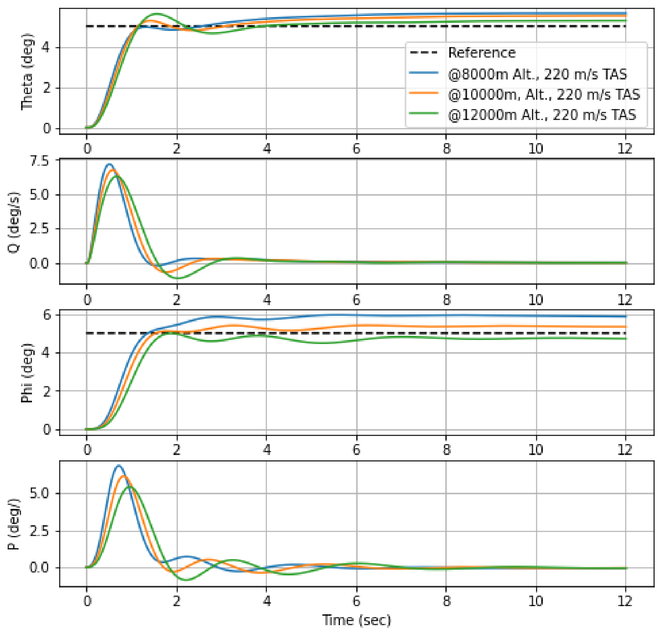


JSBSim Open Source Flight Dynamics · python · Keras

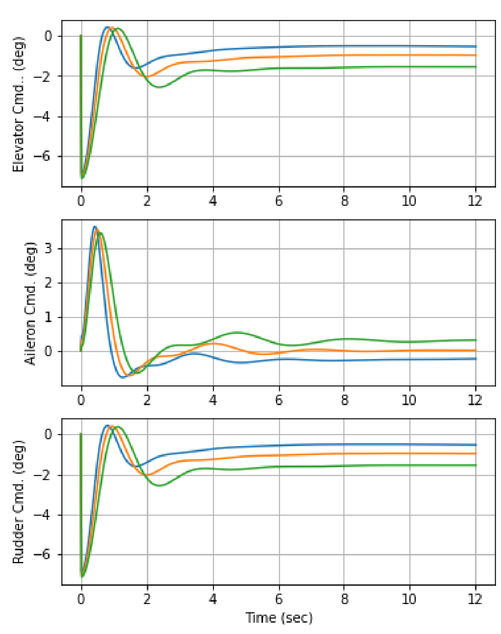
Mathematical Model of Boeing 737 with;

- 6 degrees-of-freedom nonlinear dynamics
- Propulsion system model
- Turbulence, wind and gust effects
- Sensor models

Attitude and Body Rate Responses of Nonlinear Boeing 737 Model with AI-based Flight Control System



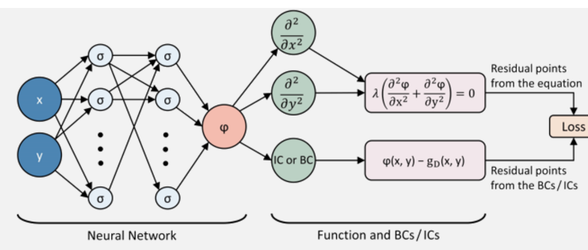
Control Surface Deflection Commands

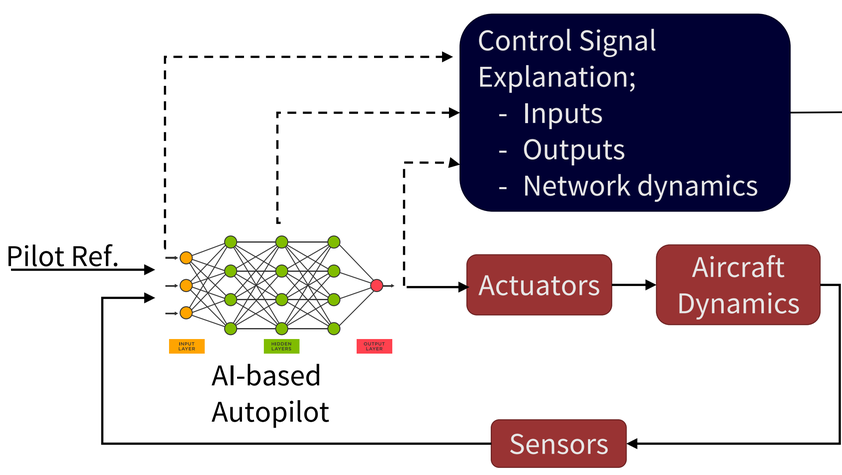


B. Explainability of the AI-Based FCS

Interpretability => Explainable and Trustworthy AI

- Physics Informed Deep Learning
- Ability to identify system behaviour
- Generalization capability
- Anomaly detection/classification



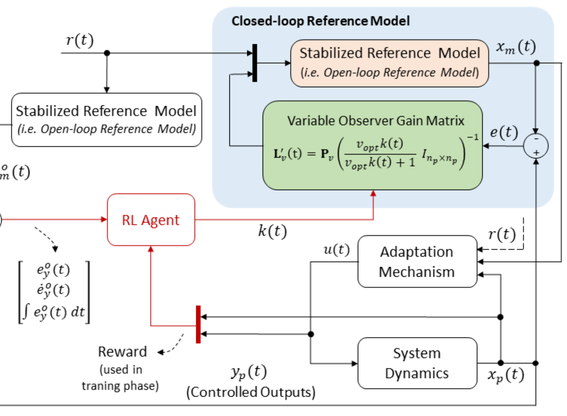


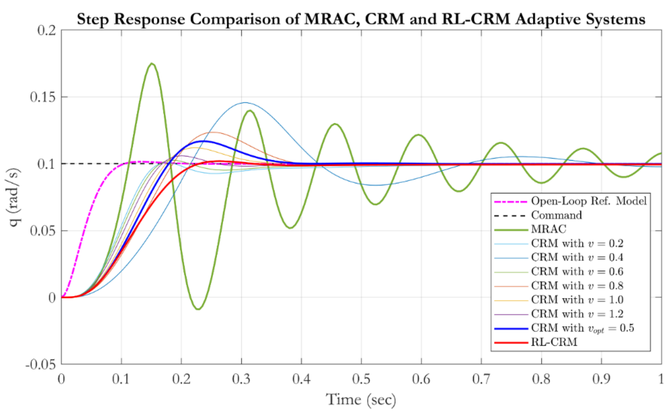
Explainable Flight Control Systems

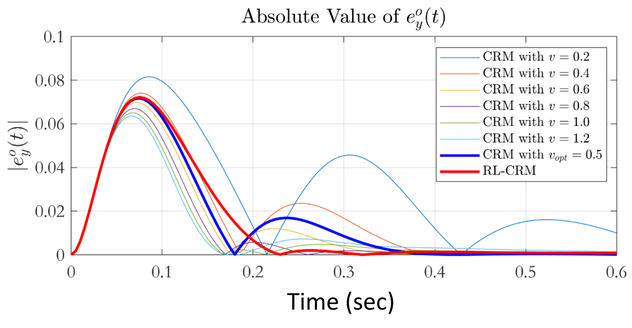
- Input-output relationship
- Inform pilots about the control action

C. Adaptive Security Strategies

- Deep Reinforcement Learning Based Adaptive Controls
  - Learn adaptation strategy through observation between reference model and the reality







Performance Metrics	MRAC	CRM	Improvement (%)	RL-CRM	Improvement (%)
$\overline{\|\|\dot{K}_x\|\|}$	15.2114	3.7341	75.4520	2.4489	83.9008
$\overline{\|\|\dot{K}_r\|\|}$	18.4647	7.8298	57.5958	5.5146	70.1344
$\overline{\|\|\dot{\hat{\theta}}\|\|}$	0.0888	0.0338	61.9369	0.0207	76.6892
$\overline{\|\|y_m\|\|_\infty}$	0.2	0.2064	-3.2	0.2	-
$\overline{\|\|e_y\|\|}$	0.4616	0.1957	57.6039	0.1379	70.1256
$\overline{\|\|e_y^o\|\|}$	0.4616	0.3928	14.9047	0.3886	15.8145
$\overline{\|\|\dot{u}\|\|}$	6.5704	2.0811	68.3262	1.4163	78.4290

This work is supported, in part, by the Engineering and Physical Sciences Research Council [grant number EP/V026763/1]