RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

ENAC

ICO
Institut Cybersécurité Occitanie

UNIVERSITÉ PERPIGNAN VIA DOMITIA

Cybersécurité pour les systèmes embarqués critiques à base d'IA

DOCTORANTE : CÉLINE BELLANGER

DIRECTEURS DE THÈSE : PIERRE-LOÏC GAROCHE, MATTHIEU MARTEL

ENCADRANTE DE THÈSE : CELIA PICARD

Cadre de la thèse

→ Thèse débutée le 1er octobre 2022

→ ENAC (Toulouse) / Université de Perpignan

→ Co-financement ICO / ENAC

Contexte

Contexte

Domaine d’application : l’aéronautique

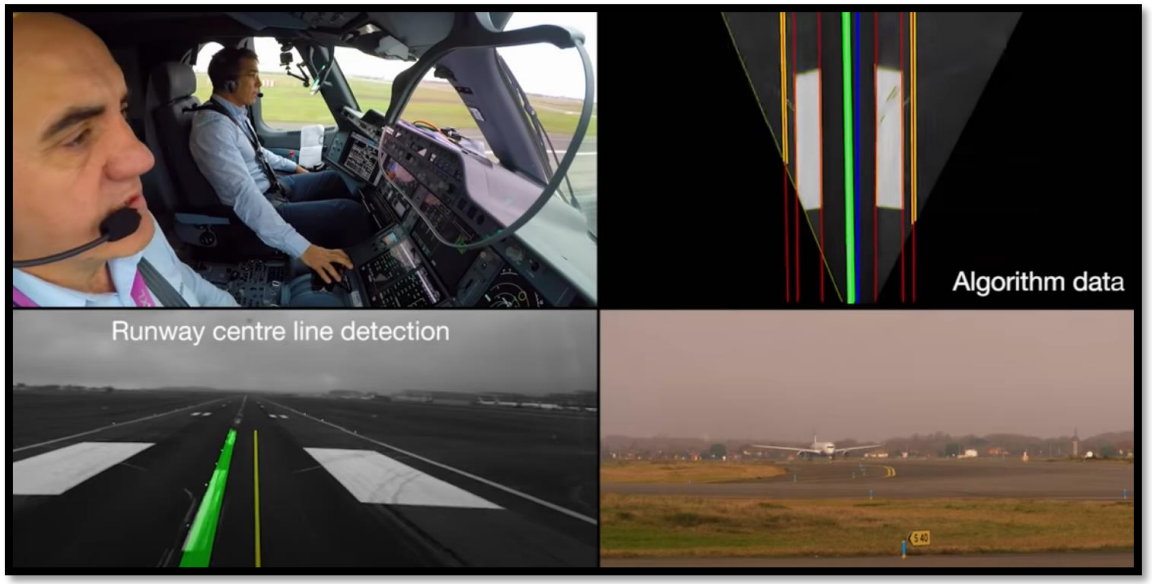
L’industrie aéronautique s’intéresse de plus en plus à l’IA :

Contexte

Domaine d'application : l'aéronautique

L'industrie aéronautique s'intéresse de plus en plus à l'IA :

- Projet ATTOL Airbus



Essai de vol autonome – Airbus A350

https://www.youtube.com/watch?v=9TIBeso4abU&ab_channel=Airbus

Contexte

Domaine d'application : l'aéronautique

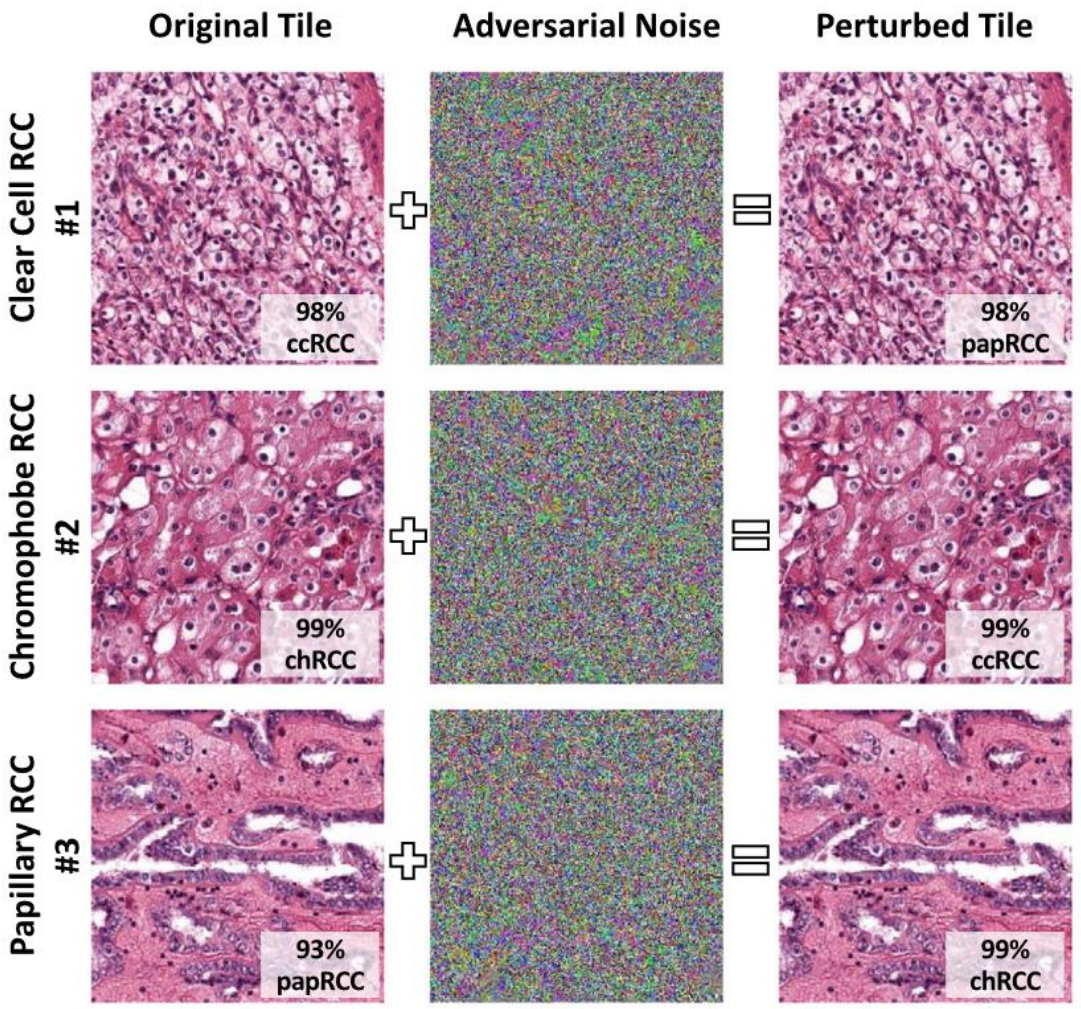
L'industrie aéronautique s'intéresse de plus en plus à l'IA :

- Projet ATTOL Airbus
- Utilisation de l'IA dans les contrôleurs (guidage, stabilisation, robustesse)

Contexte

L'Intelligence Artificielle est sensible à de nouveaux types de cyberattaques

➔ L'exemple des *adversarial attacks*



Ghaffari Laleh, N., Truhn, D., Veldhuizen, G.P. *et al.* Adversarial attacks and adversarial robustness in computational pathology. *Nat Commun* **13**, 5711 (2022). https://doi.org/10.1038/s41467-022-33266-0

Contexte

Notre problématique :

Assurer la sûreté des contrôleurs à base de réseaux de neurones utilisés dans l'aéronautique

Notre approche

I. Identification de propriétés d’intérêt

II. Vérification de propriétés

I. Identification de propriétés d’intérêt

I. Identification de propriétés d'intérêt

Propriétés de base liées aux contrôleurs et à leur fonctionnalités, potentiellement sensibles aux cyberattaques :

- Stabilité
- Guidage
- Robustesse
- Performances
- …

I. Identification de propriétés d'intérêt

Propriétés fonctionnelles des contrôleurs basés sur des réseaux de neurones :

- Liées à l'environnement
- Liées aux données d'entraînement et de validation
- Liées à la structure du réseau de neurones

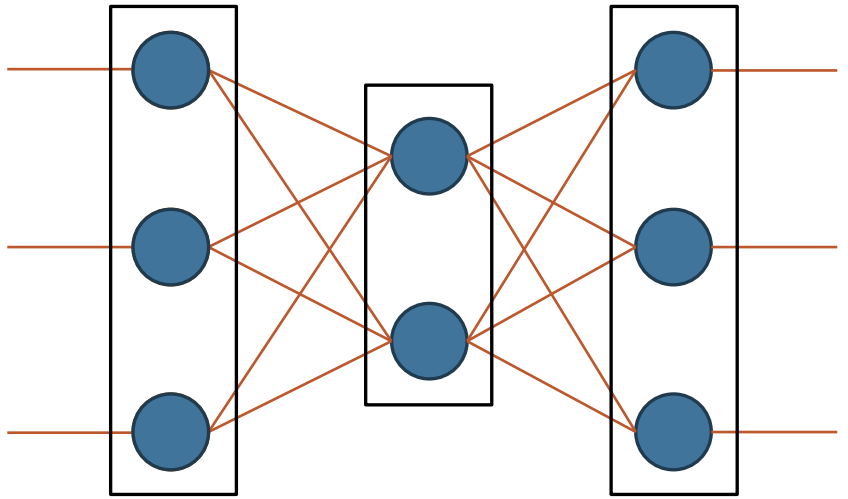


Schéma d'un réseau de neurones

I. Identification de propriétés d'intérêt

Propriétés liées aux cyberattaques :

1. Identification des cyberattaques impactant les réseaux de neurones dans les systèmes embarqués

2. Détermination de propriétés caractéristiques de ces cyberattaques

I. Identification de propriétés d'intérêt

Étude des systèmes dynamiques :

- Réaction à l'environnement ➔ notion temporelle
- La plupart des propriétés peuvent être définies en fonction du temps

  ➔ Signal Temporal Logic (STL)

- Simulink : outil très utilisé pour la conception de systèmes dynamiques

  ➔ Nécessité de s'adapter à Simulink

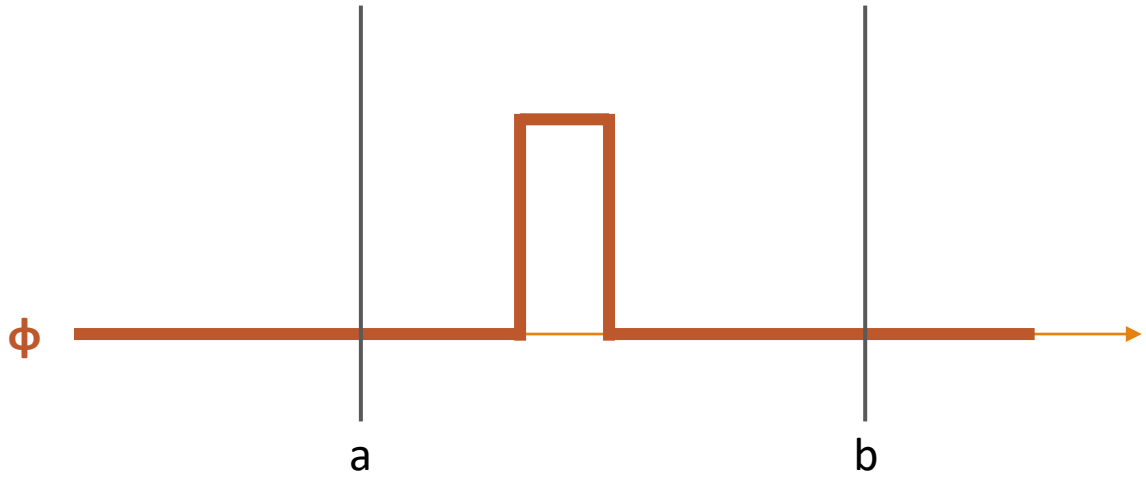
II. Vérification de propriétés STL

II. Vérification de propriétés STL

- Signal Temporal Logic (STL)
  - $\Diamond_{[a,b]}\, \phi$: *Eventually*
  - $\Box_{[a,b]}\, \phi$: *Always*
  - $\phi_1\, U_{[a,b]}\, \phi_2$: *Until*

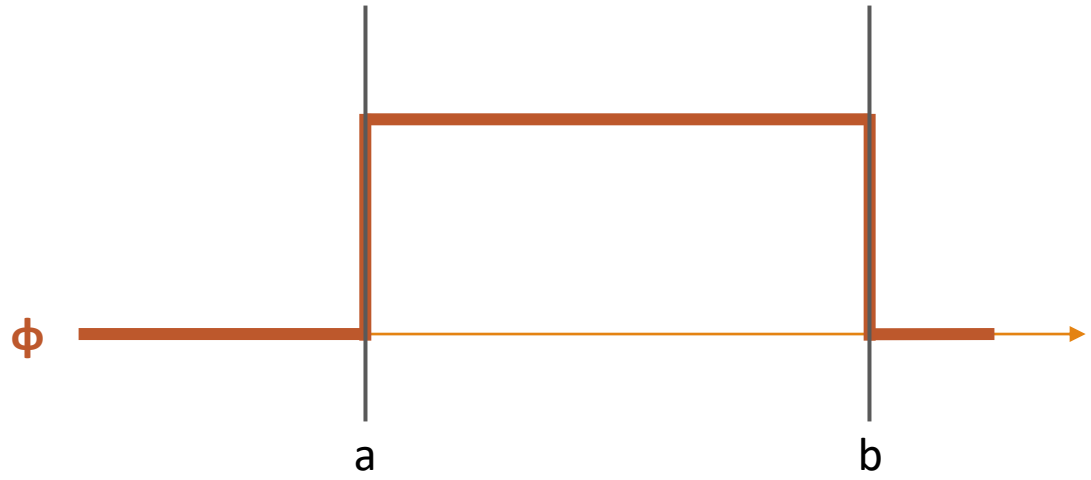
II. Vérification de propriétés STL

- Signal Temporal Logic (STL)
  - $\Diamond_{[a,b]}\, \phi$: ***Eventually***
  - $\Box_{[a,b]}\, \phi$: *Always*
  - $\phi_1\, U_{[a,b]}\, \phi_2$: *Until*



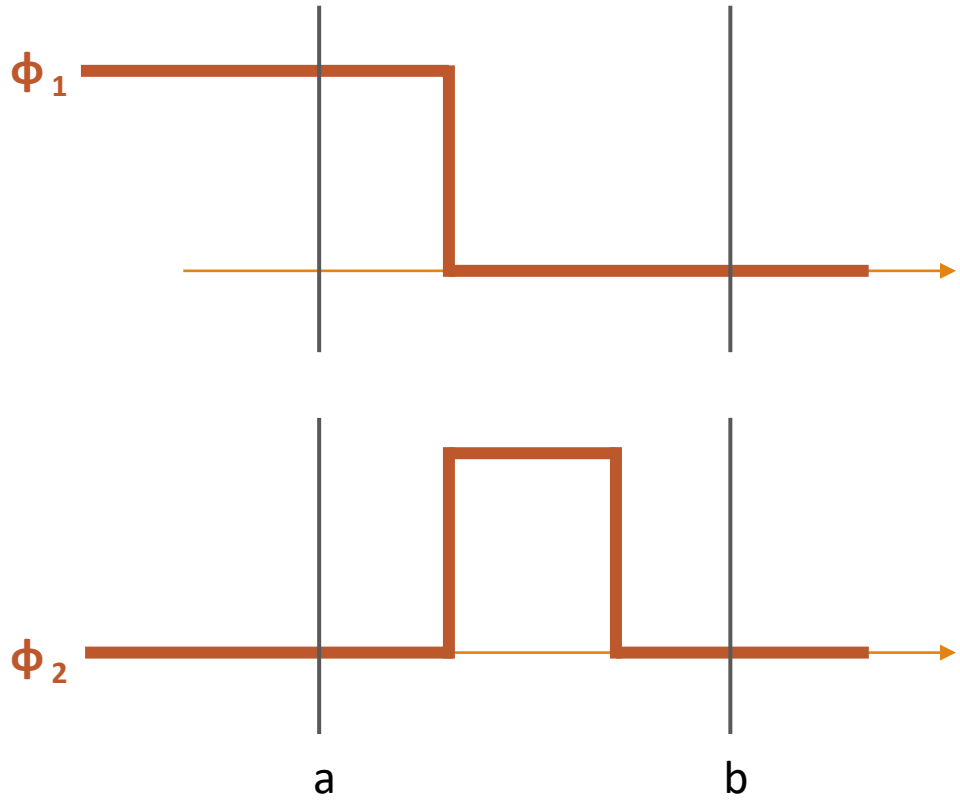
II. Vérification de propriétés STL

- Signal Temporal Logic (STL)
  - $\Diamond_{[a,b]}\, \phi$: *Eventually*
  - $\Box_{[a,b]}\, \phi$: ***Always***
  - $\phi_1\, U_{[a,b]}\, \phi_2$: *Until*



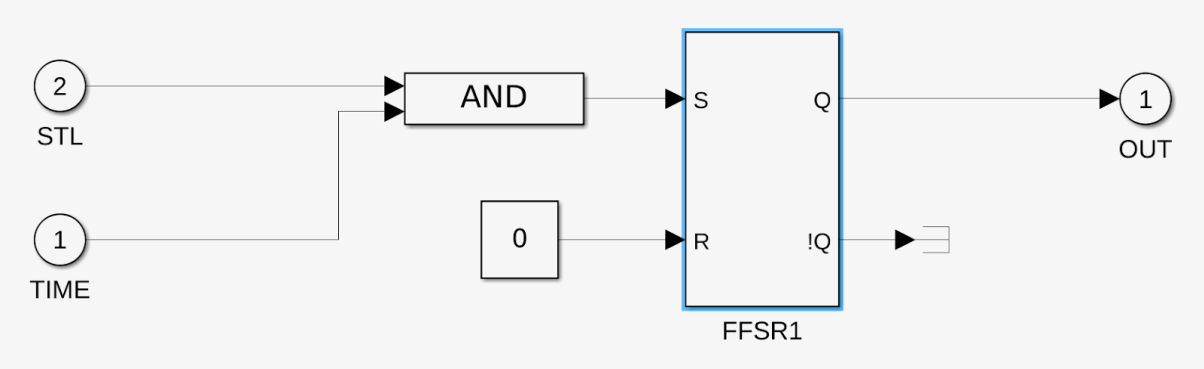
II. Vérification de propriétés STL

- Signal Temporal Logic (STL)
  - $\Diamond_{[a,b]}\, \phi$: *Eventually*
  - $\Box_{[a,b]}\, \phi$: *Always*
  - $\boldsymbol{\phi_1\ U_{[a,b]}\ \phi_2}$ **:** ***Until***



II. Vérification de propriétés STL

- Implémentation d'opérateurs STL – Balsini

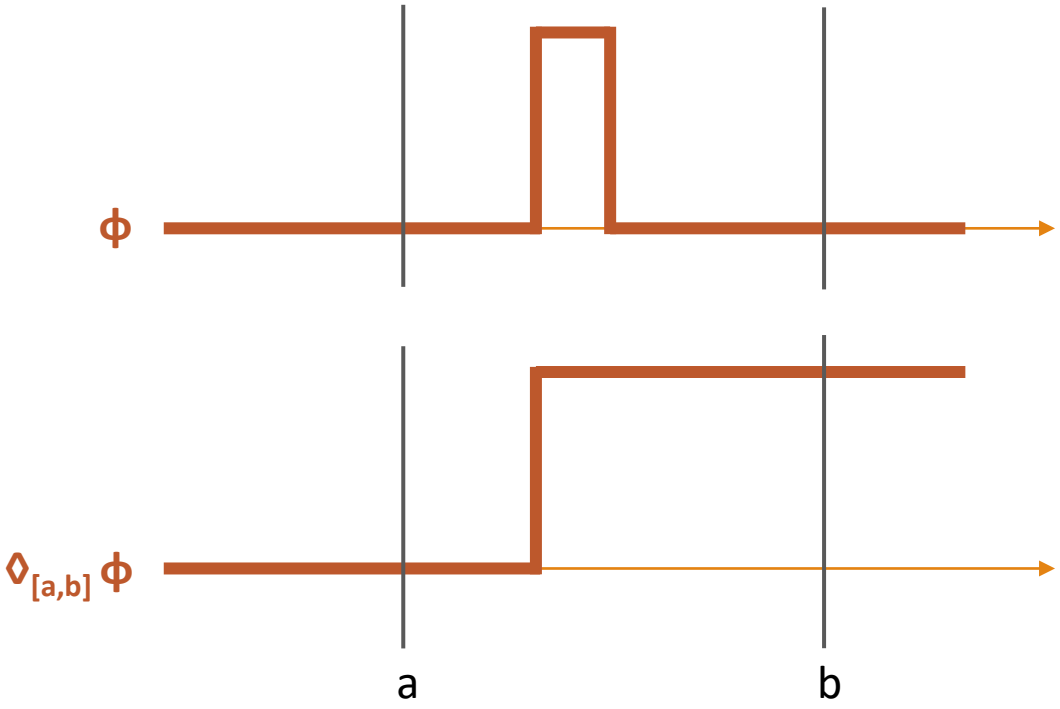


Bloc STL pour représenter l'opérateur *Eventually*

Absence de preuve formelle du modèle

II. Vérification de propriétés STL

- Implémentation d'opérateurs STL – Balsini



Absence de distinction entre *False* et *Unknown* (logique trivaluée de Kleene[1])

[1]Stephen Cole Kleene. Introduction to Metamathematics. North-Holland. Amsterdam, 1952.

II. Vérification de propriétés STL

- Implémentation d'opérateurs STL – Balsini

$$\Diamond_{[0,T]} (p1 \wedge \Box_{[0,\tau]} (\neg p2))$$

Restrictions sur les opérateurs imbriqués

II. Vérification de propriétés STL

Notre proposition :

- Utilisation de la logique trivaluée de Kleene : spécifications et implémentation
- Opérateurs implémentés en Lustre et prouvés formellement
- Possibilité de faire de l'imbrication d'opérateurs sans limite

II. Vérification de propriétés STL

Notre proposition :

- Utilisation de la logique trivaluée de Kleene : spécifications et implémentation
- Opérateurs implémentés en Lustre et prouvés formellement

1ère itération

- Possibilité de faire de l'imbrication d'opérateurs sans limite

2ème itération

STATUS

Ce qui a été fait :

- Spécification des opérateurs STL en logique trivaluée
- Implémentation des opérateurs en Lustre (sans imbrication)

En cours :

- Preuves formelles de la correction des opérateurs

À venir :

- Implémentation des opérateurs avec imbrication d’opérateurs
- Identification des propriétés d’intérêt

MERCI POUR VOTRE ATTENTION

- Objectif : Assurer la sûreté des contrôleurs à base de réseaux de neurones utilisés dans l'aéronautique
- Détermination de propriétés d'intérêt sur ces systèmes
- Étude de ces propriétés à l'aide d'outils STL