# FalsifAI: Falsification of AI-Enabled Hybrid Control Systems Guided by Time-Aware Coverage Criteria

Zhenya Zhang[1], **Deyun Lyu**[1], Paolo Arcaini[2], Lei Ma[3], Ichiro Hasuo[2], Jianjun Zhao[1]

[1] *Kyushu University, Fukuoka, Japan*
[2] *National Institute of Informatics, Tokyo, Japan*
[3] *University of Alberta, Edmonton, Canada*

## Motivation

Cyber-Physical Systems (CPS) are combinations of computing units and mechanical systems. Nowadays, Artificial Intelligent (AI) components are increasingly deployed on CPS to perform complex control tasks under safety-critical conditions.

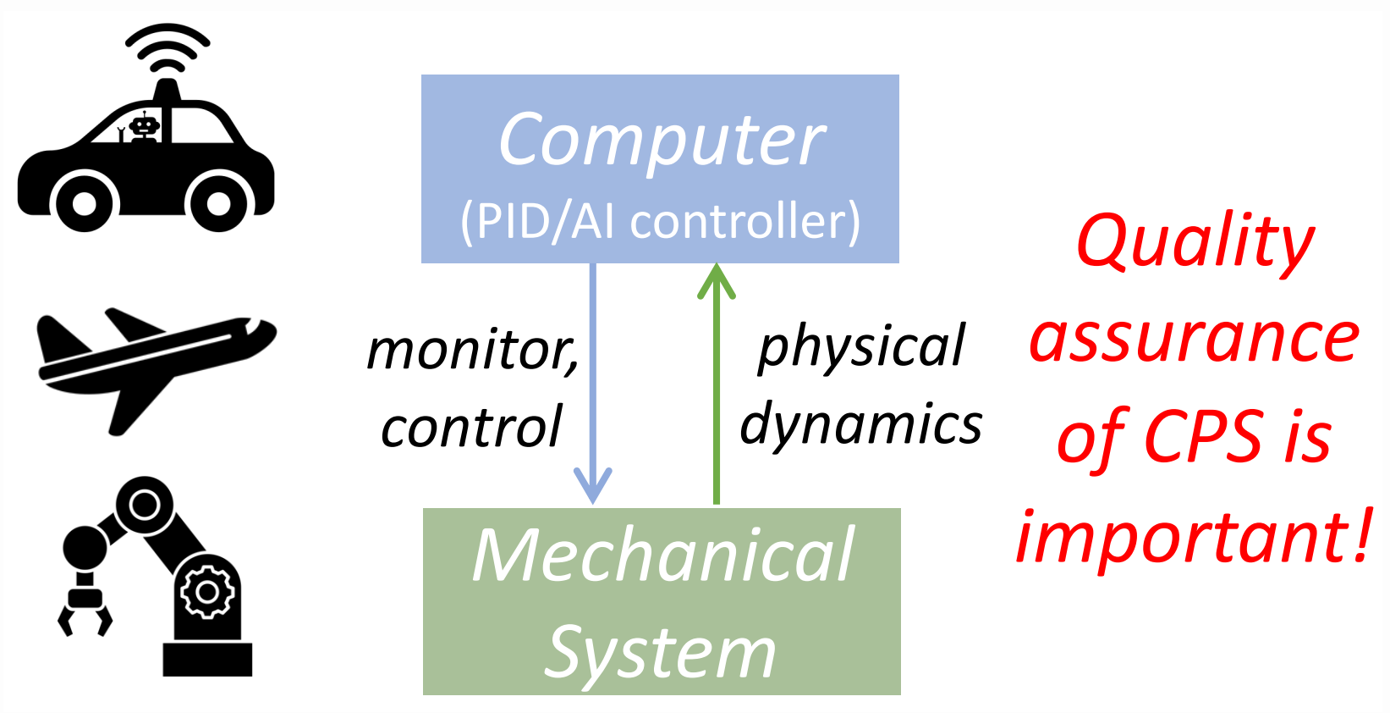


**Figure 1:** The workflow of CPS

### Classic falsification and its weakness

Falsification is a well-known validation method for quality assurance in CPS domain.

(1) Aim: Find a breach of the given specification;

(2) Method: Hill climbing algorithm;

(3) Guidance: Robustness provided by quantitative robust semantics.

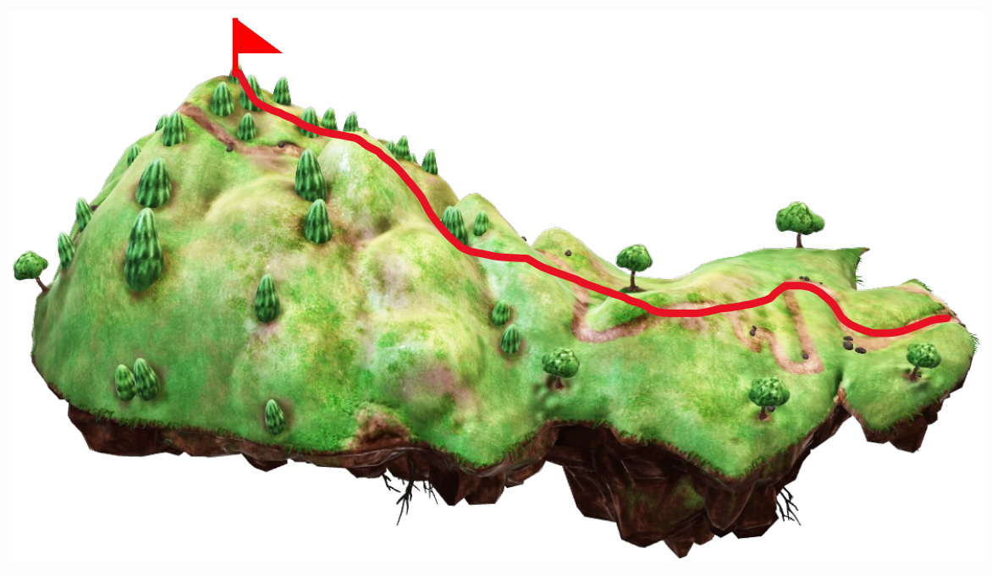


**Figure 2:** Hill climbing algorithm

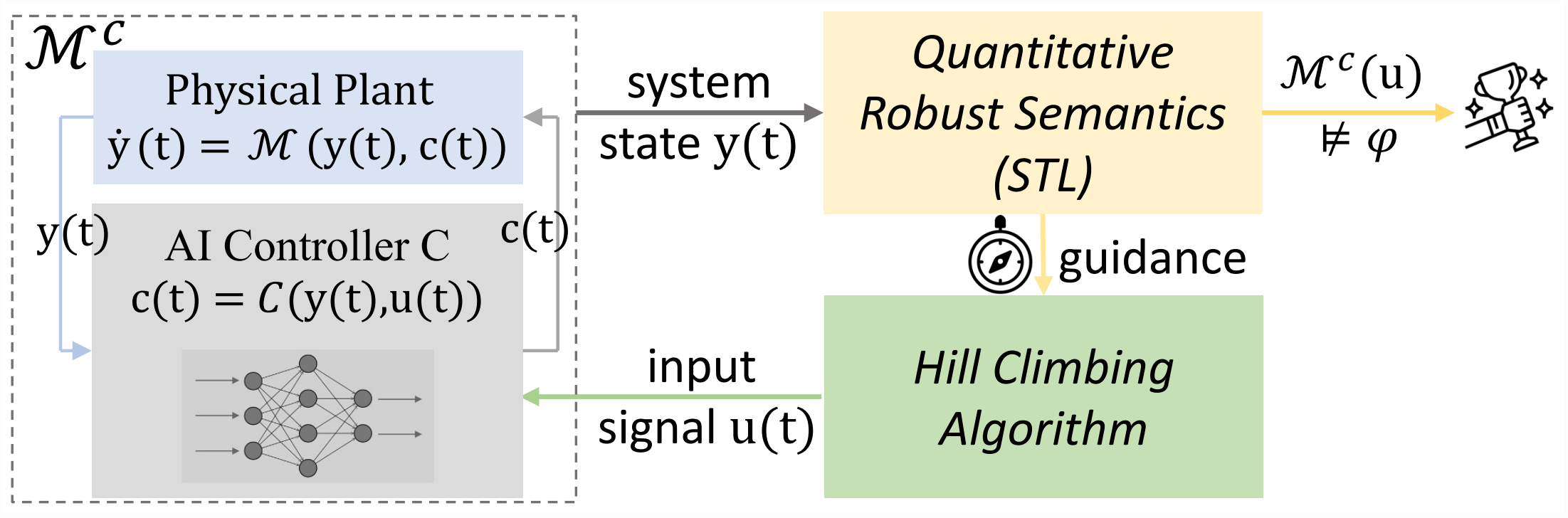


**Figure 3:** Classic falsification of AI-enabled CPS guided by robustness

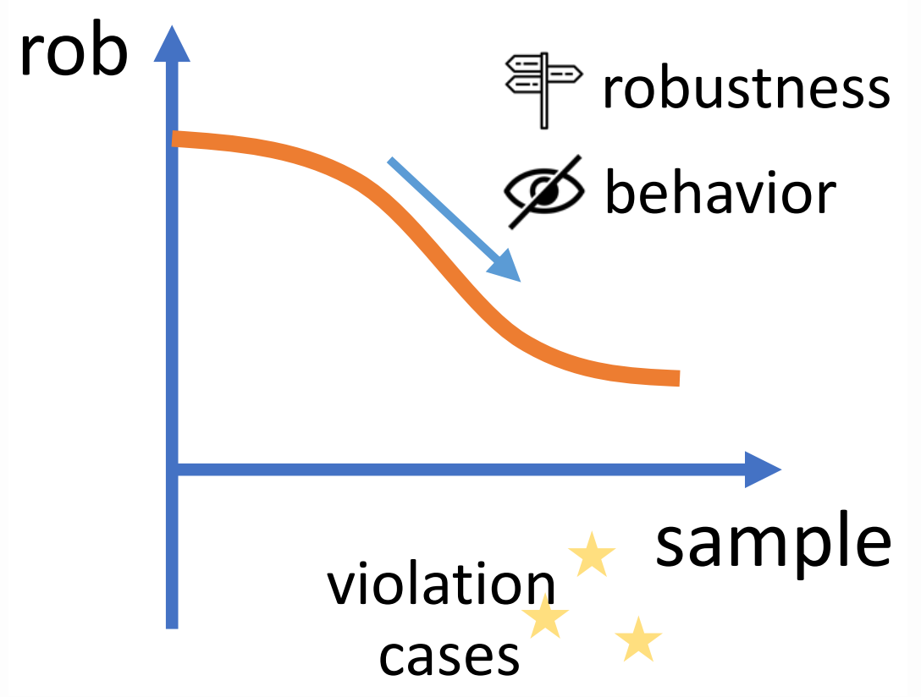
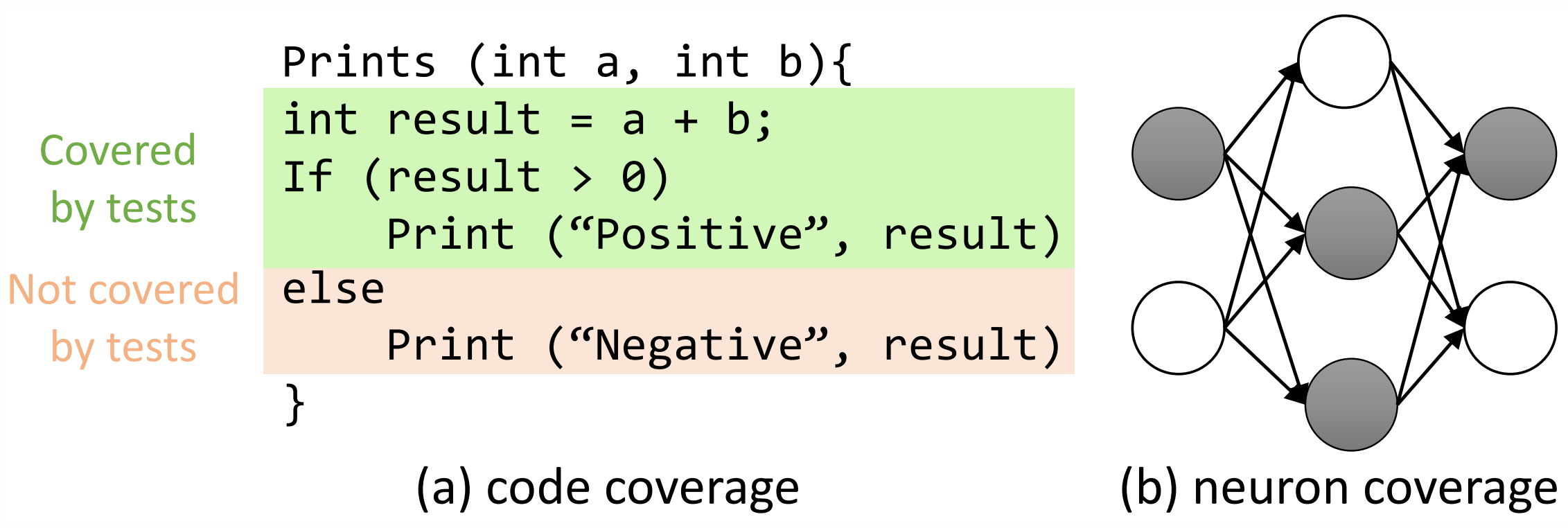


**Figure 4:** Weakness of classic falsification

## A Possible Solution

### Coverage Criteria

Coverage criteria are measures of test adequacy for programs and DNN models.



```
Prints (int a, int b){
int result = a + b;
If (result > 0)
    Print (“Positive”, result)
else
    Print (“Negative”, result)
}
```

(a) code coverage

(b) neuron coverage

**Figure 5:** Coverage criterion in traditional software and deep neural network

### Why using coverage criteria as guidance?

(1) Describe the test requirements;

(2) Fully explore system behaviors;

(3) Guide the generation of new test cases.

## A Coverage-Guided Falsification Framework

The workflow of our proposed framework FalsifAI is shown below [1].

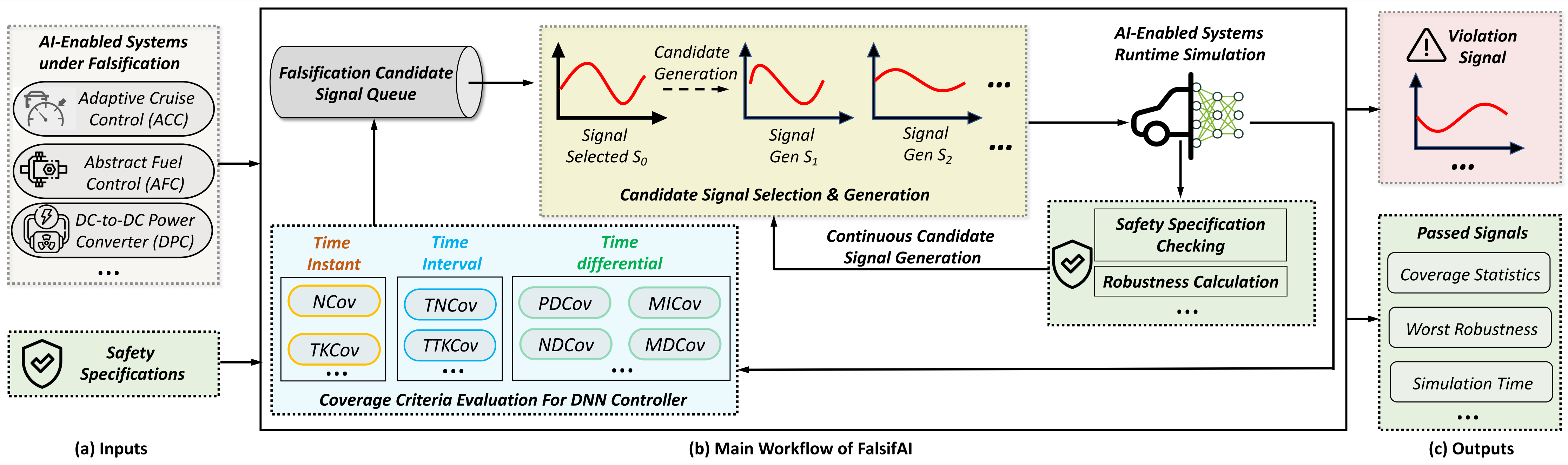

AI-Enabled Systems under Falsification
Adaptive Cruise Control (ACC)
Abstract Fuel Control (AFC)
DC-to-DC Power Converter (DPC)
...
Safety Specifications
(a) Inputs
Falsification Candidate Signal Queue
Candidate Generation
Signal Selected $S_0$
Signal Gen $S_1$
Signal Gen $S_2$
...
Candidate Signal Selection & Generation
AI-Enabled Systems Runtime Simulation
Continuous Candidate Signal Generation
Safety Specification Checking
Robustness Calculation
...
Time Instant
Time Interval
Time differential
NCov
TKCov
TNCov
TTKCov
PDCov
MICov
NDCov
MDCov
Coverage Criteria Evaluation For DNN Controller
(b) Main Workflow of FalsifAI
Violation Signal
Passed Signals
Coverage Statistics
Worst Robustness
Simulation Time
(c) Outputs

**Figure 6:** The workflow of FalsifAI

### Time-Aware Coverage Criteria

1. Time Instant Coverage Criteria 2. Time Interval Coverage Criteria 3. Time Differential Coverage Criteria

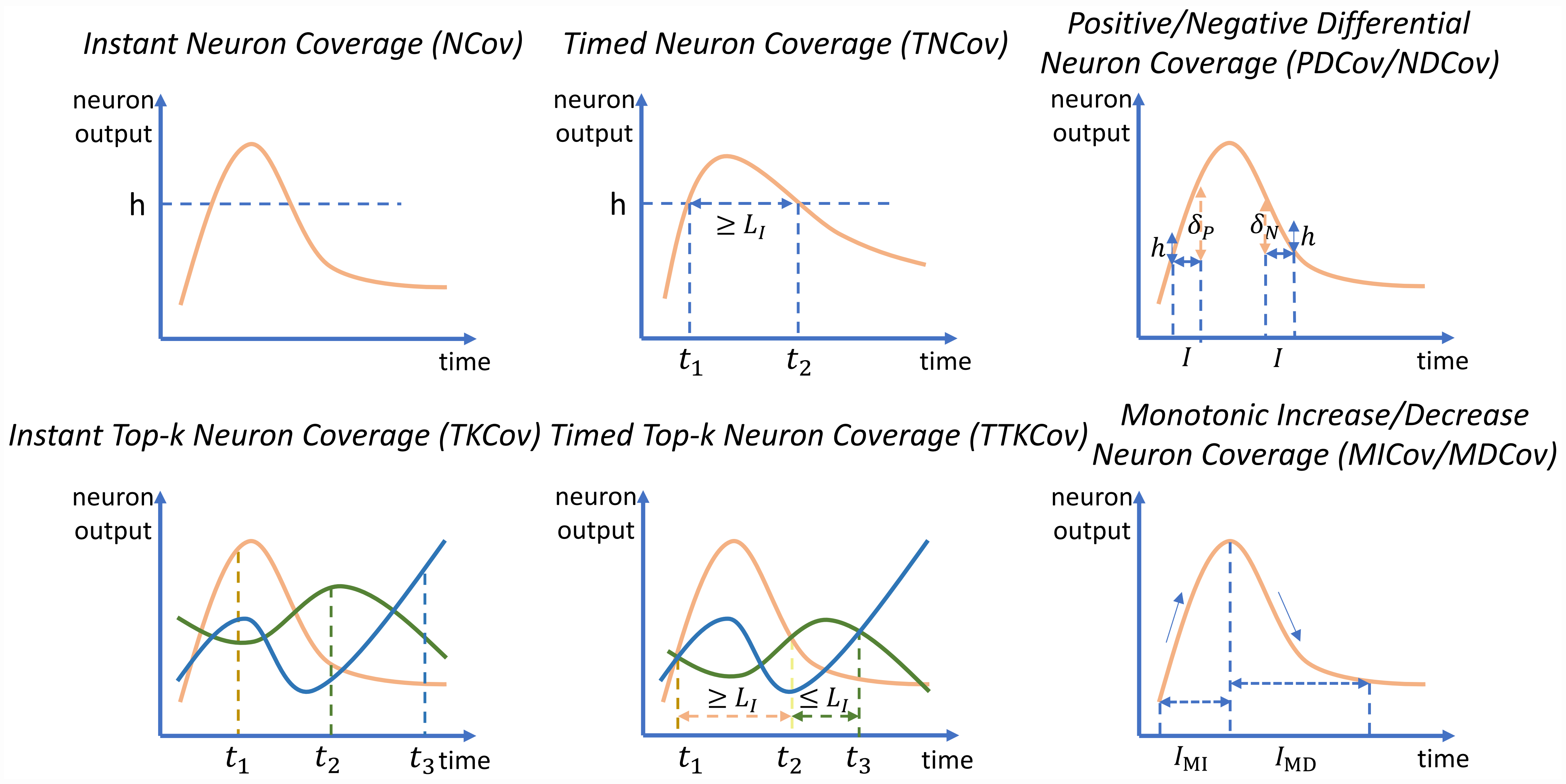
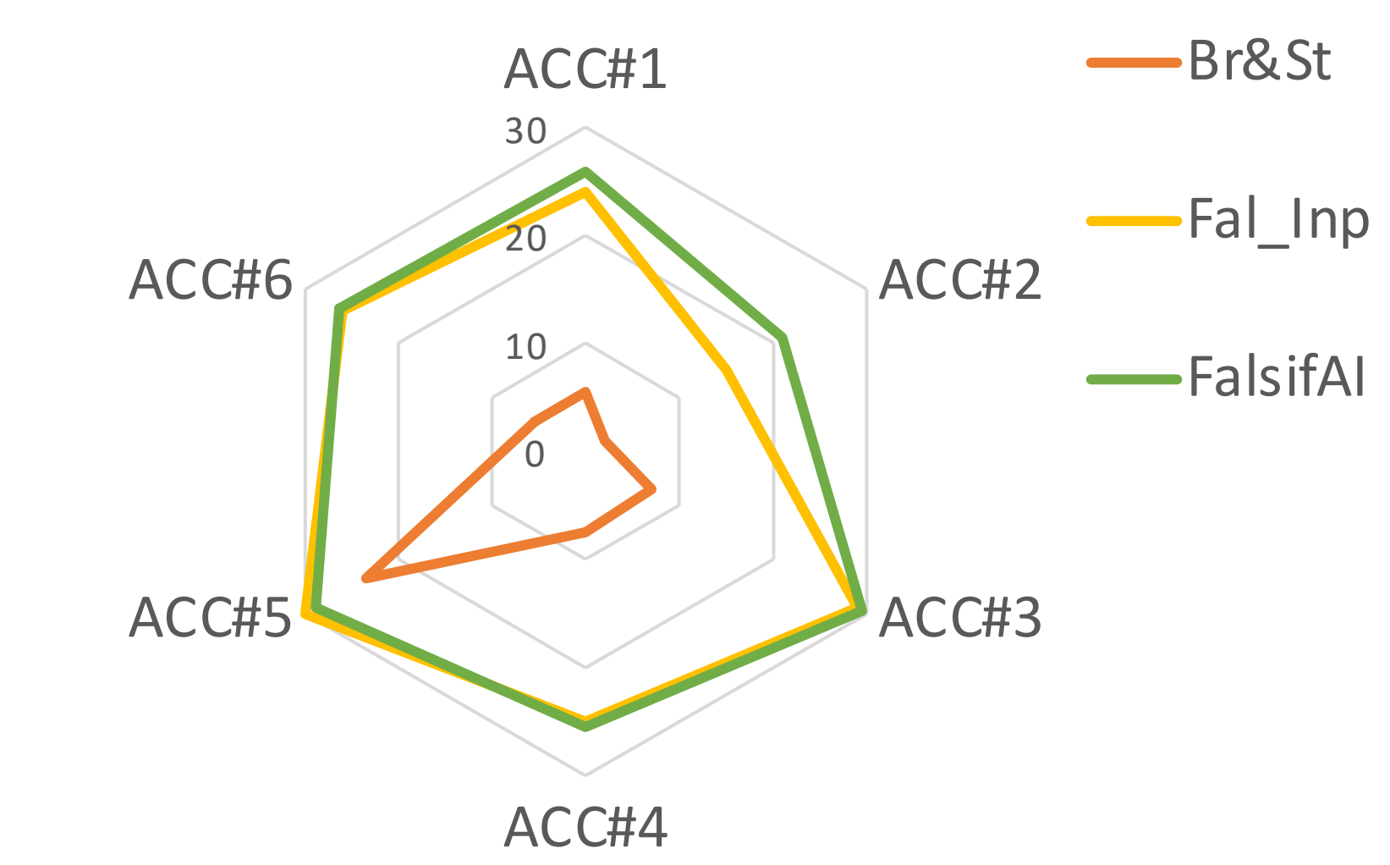


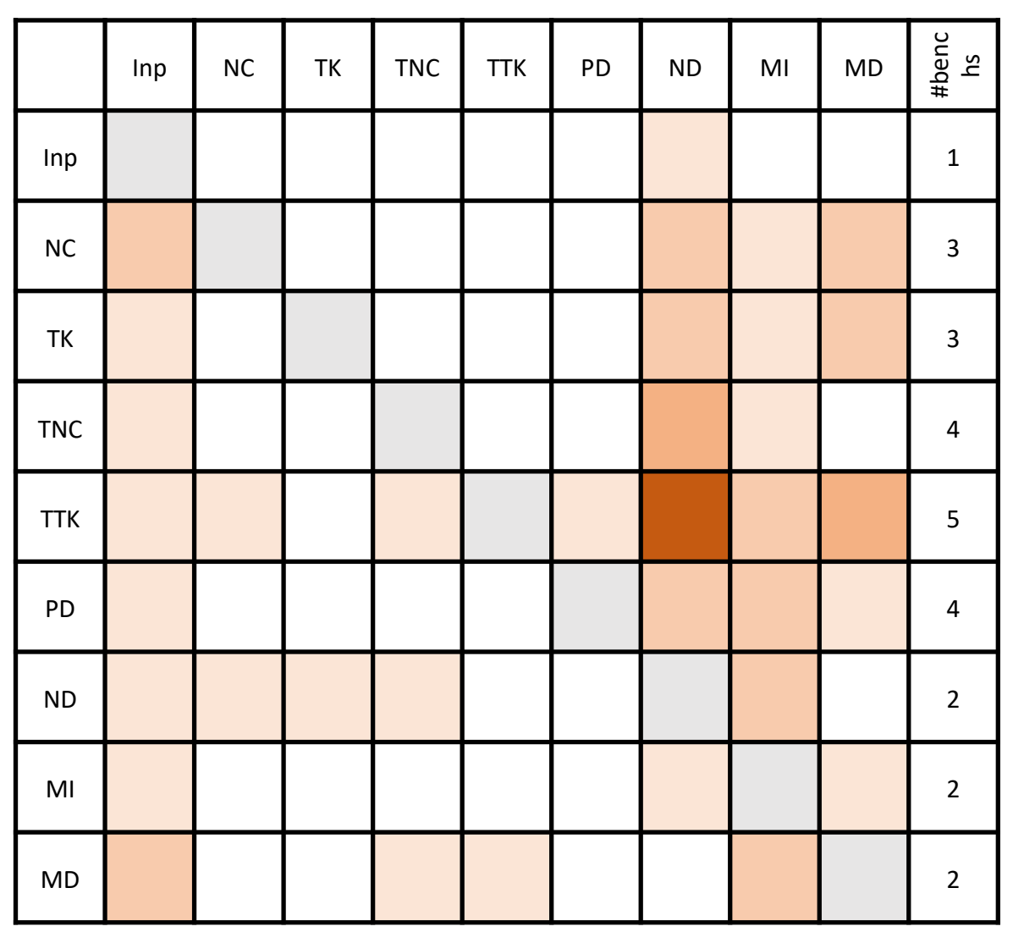
**Figure 7:** Eight time-aware coverage criteria

## Experiments

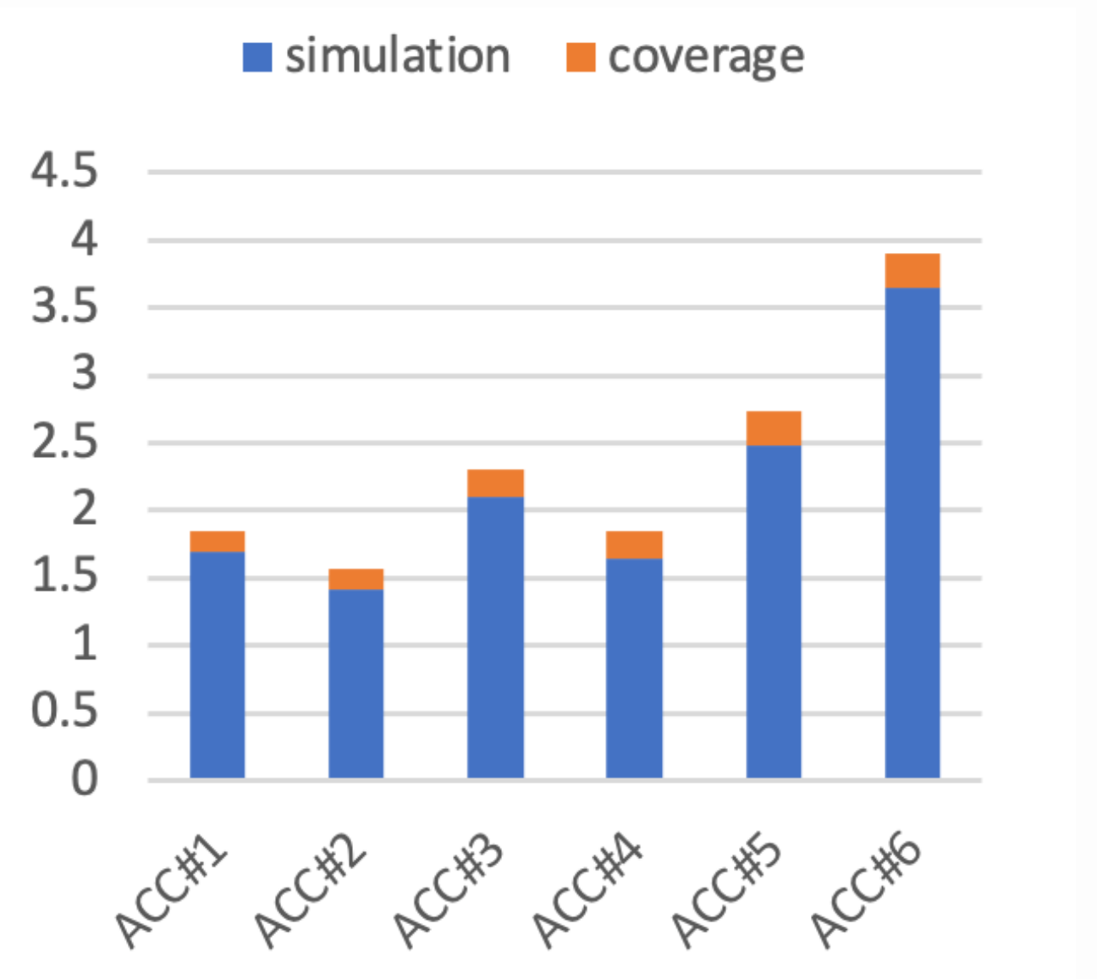
Our evaluation was performed on 3 subject CPS with 6 specifications and 18 well-trained DNN controllers. Refer to our paper [1] for more details.



RQ1: Falsification success rate of ACC with spec#1



RQ2: Effectiveness as a guidance



RQ3: Overhead of FalsifAI

Conclusion:

FalsifAI significantly outperforms Br&St, but is not always better that Fal_Inp.

Guidance provided by eight coverage criteria is better than the one provided by the coverage of the input space.

FalsifAI performs effectively for CPS with different sizes of DNN controllers.

## Conclusion and Future Work

In this paper, we proposed a coverage-guided falsification framework FalsifAI, which exhaustively employs eight time-aware coverage criteria as guidance to explore the temporal behaviors of AI-enabled CPS. These coverage criteria aim to capture different time-series features of DNN controller and provide better guidance to find violation cases to the system specification. The large-scale evaluation demonstrates the effectiveness of FalsifAI and our proposed coverage criteria. In the future, we will extent this work to other types of neural network controllers and design different coverage criteria for falsification.

## References

[1] Z. Zhang, D. Lyu, P. Arcaini, L. Ma, I. Hasuo, and J. Zhao, “FalsifAI: Falsification of AI-Enabled Hybrid Control Systems Guided by Time-Aware Coverage Criteria,” *IEEE Transactions on Software Engineering*, pp. 1–17, 2022.