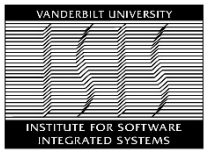


Tools for Assurance-based Learning-enabled Cyber-Physical Systems: An Experience Report

Gabor Karsai (PI)

contributions by

Ted Bapty, Abhishek Dubey, Taylor Johnson, Xenofon Koutsoukos, Janos Sztipanovits and many others

Supported by DARPA Assured Autonomy Program

DARPA Assured Autonomy
MBSE with LECs

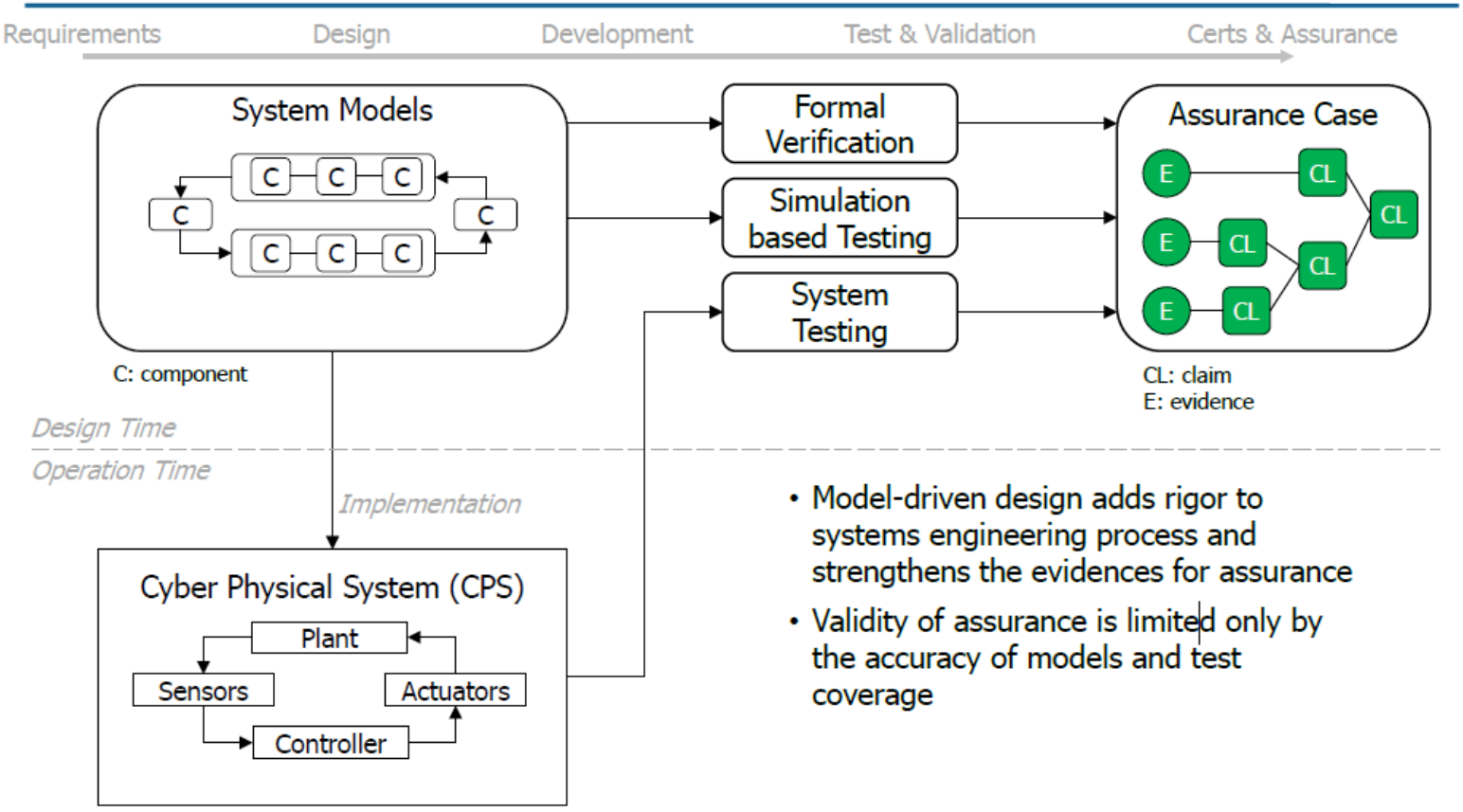
- To develop rigorous design and analysis technologies for the continual assurance of learning-enabled autonomous systems, in order to guarantee safety properties in adversarial environments

Glossary

AA:	Assured Autonomy
CPS:	Cyber-Physical System
GSN:	Goal Structuring Notation
LEC:	Learning-Enabled Component
MBSE:	Model-based Systems Engineering

Source: DARPA AA Program Presentation

Model-driven Design for Safety Assurance State of Art

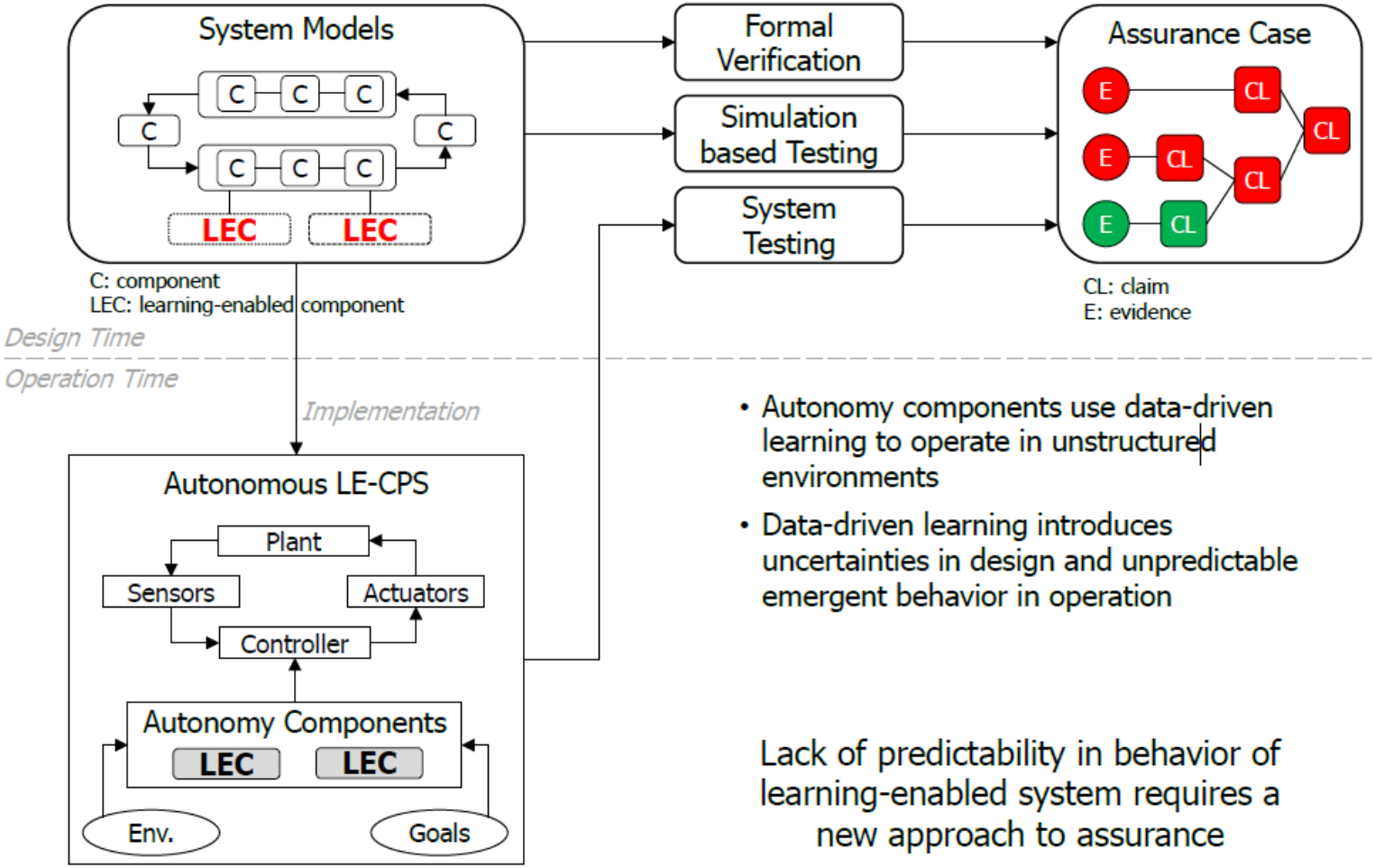


- Model-driven design adds rigor to systems engineering process and strengthens the evidences for assurance
- Validity of assurance is limited only by the accuracy of models and test coverage

Applicable only to non-learning systems operating in well-characterized environments

Source: DARPA AA Program Presentation

Learning-Enabled Autonomous Systems Lack Safety Assurance



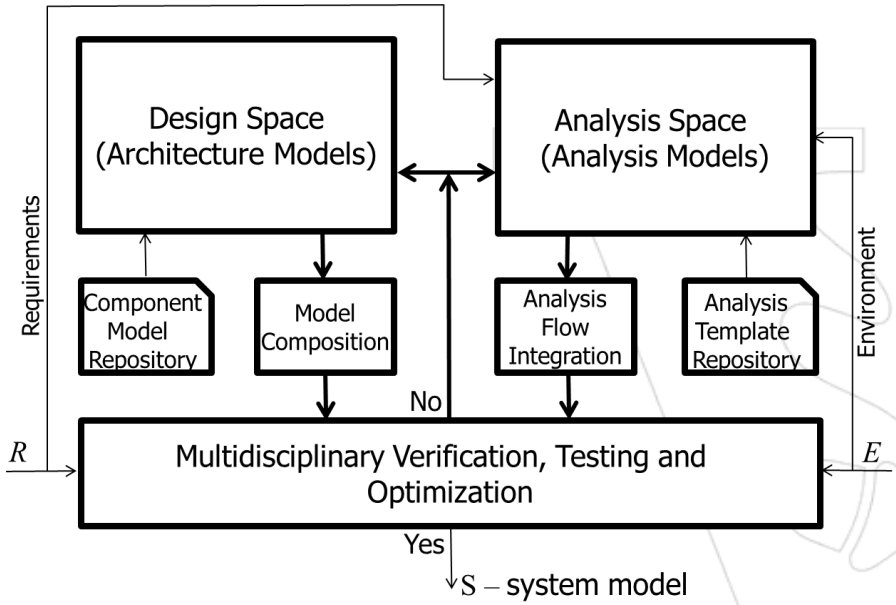
- Autonomy components use data-driven learning to operate in unstructured environments
- Data-driven learning introduces uncertainties in design and unpredictable emergent behavior in operation

Lack of predictability in behavior of learning-enabled system requires a new approach to assurance

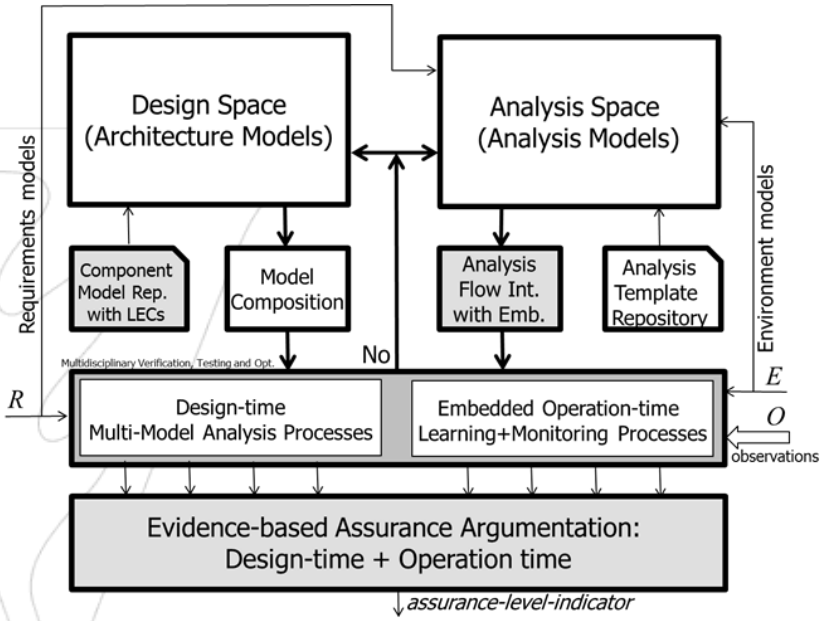
Source: DARPA AA Program Presentation

ALC Project vision

Model-driven design flow



Model-driven design flow with LEC-s



Challenges:

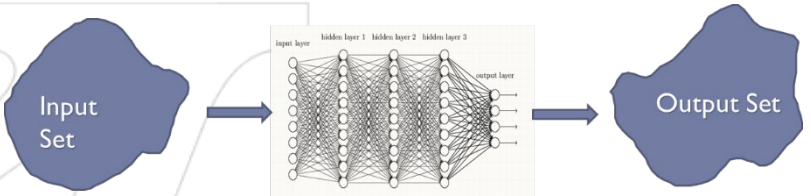
- What are the 'semantically rigorous and integrated high-level abstractions' for CPS with LECs?
- What systems engineering tools are needed to support assured development of such systems?

MBSE Design Process with LECs

1. Component models for LECs
   - → LEC's are software components with well-defined interfaces
   - → They are 'valid' only for the *data set* they were trained on
2. Verification of LECs
   - → For some cases, formal verification (model checking) is feasible for LECs LEC's are software components with well-defined interfaces
3. Assurance monitoring of LECs
   - → For some other cases, we can give an 'estimate' for the confidence in the output of a LEC (given the training set)
4. Design toolsuite for CPS with LECs
   - → The systems engineering toolsuite must provided integrated support for system + software engineering, LEC training/testing/verification, and system assurance

1. Component Model for LECs

- LECs are …
  - Computational (cyber) components, with causal interfaces
    - ➔ 'signal' interfaces
  - Engineered via 'training'
    - ➔ are 'correct' (?) only for the bounded set the training data was drawn from
  - Not to be trusted if input is not from same distribution as the training data
    - ➔ run-time monitoring of 'assurance metric' is needed



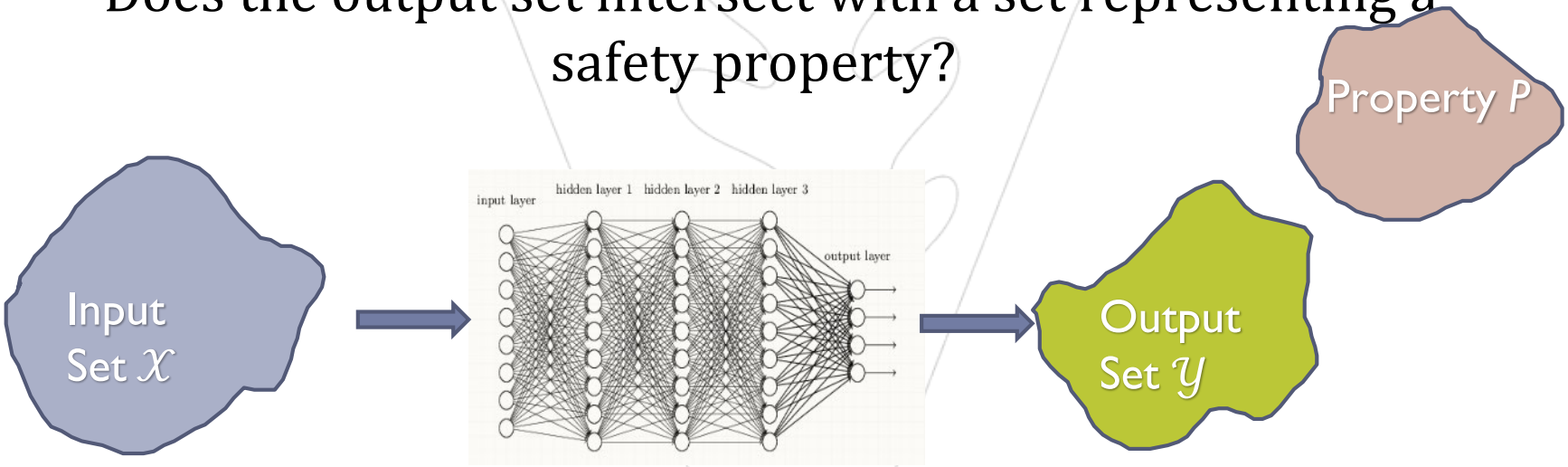
Legend: ➔ *Lessons learned*

2. Verification of LECs*

- Verification of BIBO property

  Given a NN F & an input set $\mathcal{X}$, the **output reachable set** of F is $\mathcal{Y} = \{y \mid y = F(x), x \in \mathcal{X}\}$.

  Does the output set intersect with a set representing a safety property?



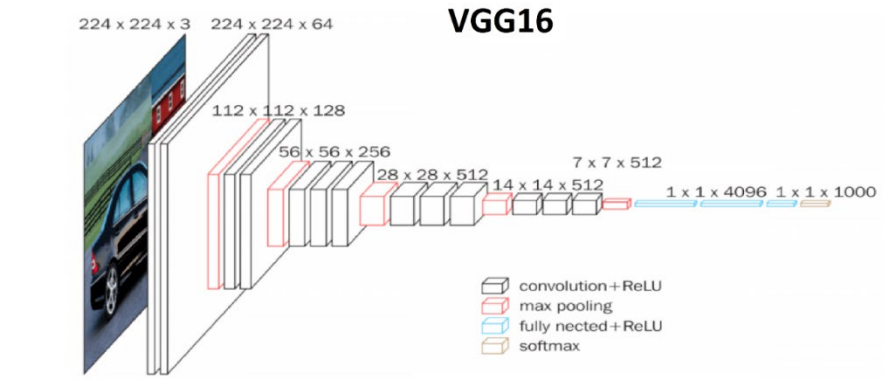
➔ *Works if the sets are known and can be expressed as polytopes.*

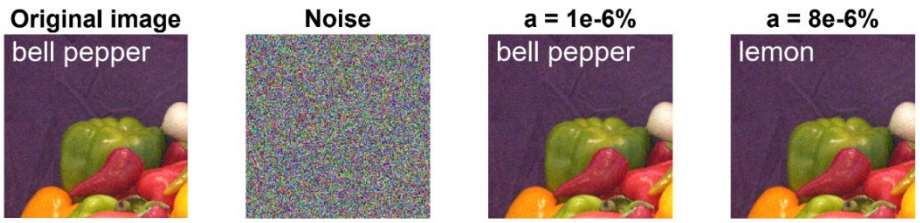
NNV: The Neural Network Verification Tool for Deep Neural Networks and Learning-Enabled Cyber-Physical Systems (https://arxiv.org/abs/2004.05519)

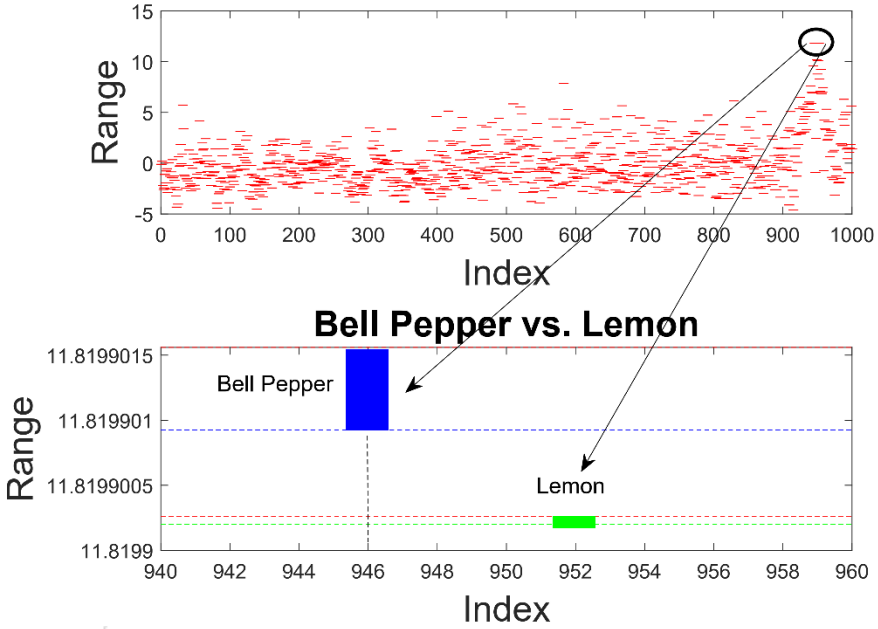
*: Work of Taylor Johnson and team

2. Verification of LECs

- Robustness verification [Classifier LECs]







Disturbed images = Original image + *a* x Noise

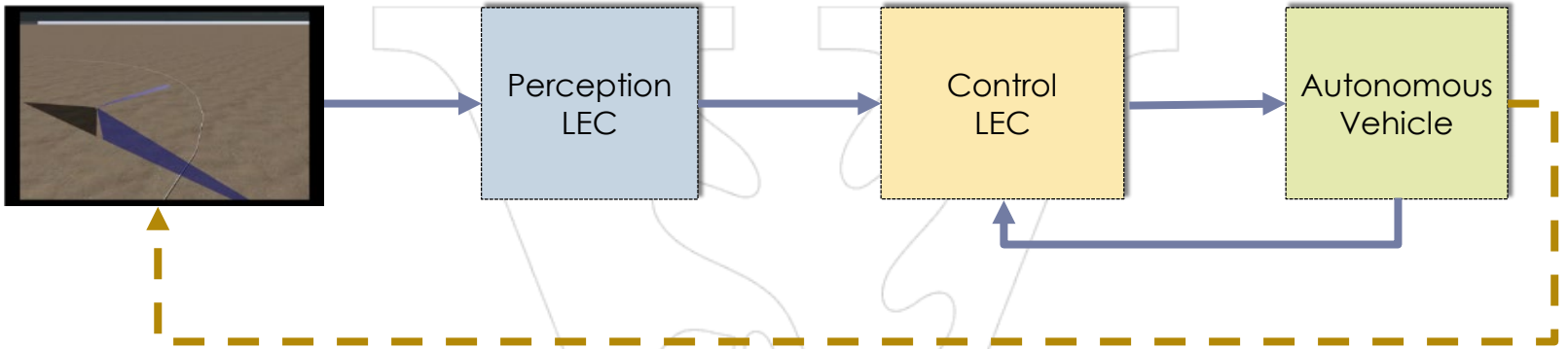
Is VGG16 robust wr.t. FGSM attack for $a \leq 2 \times 10^{-8}$?

➔ *Works if the noise can be bounded.*

H.-D. Tran, S. Bak, W. Xiang, and T. T. Johnson, Verification of Deep Convolutional Neural Networks Using ImageStars, in Computer Aided Verification, Cham, 2020, pp. 18–42.

3.Assurance monitoring of LECs*

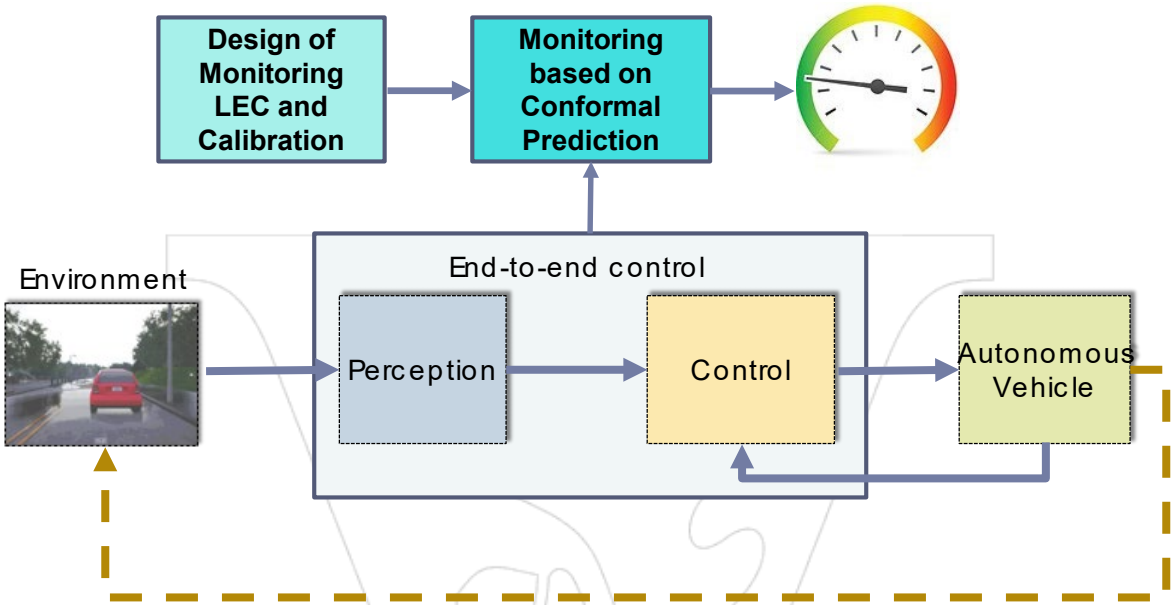
Can we trust the output of the LEC?



- **Assurance Monitoring Based on Inductive Conformal Prediction**
  - Characterize how close the LEC behavior is to a model that represents the *expected safe behavior* obtained during the training phase.
  - Compute measures of confidence associated with predictions from LECs
  - Nonconformity measure is used to evaluate the degree to which a new example disagrees from a set of examples
  - Confidence is computed based how different is a test example compared to a set of calibration examples

*: Work of Xenofon Koutsoukos and team

3.Assurance monitoring of LECs



Assurance monitoring based on inductive conformal anomaly detection

- Variational autoencoder (VAE)
- VAE for regression/classification
- Adversarial Autoencoder (AAE)
- Deep support vector description (SVDD)

- Evaluation
  Self-driving simulator (and open datasets)
  Autonomous underwater vehicle

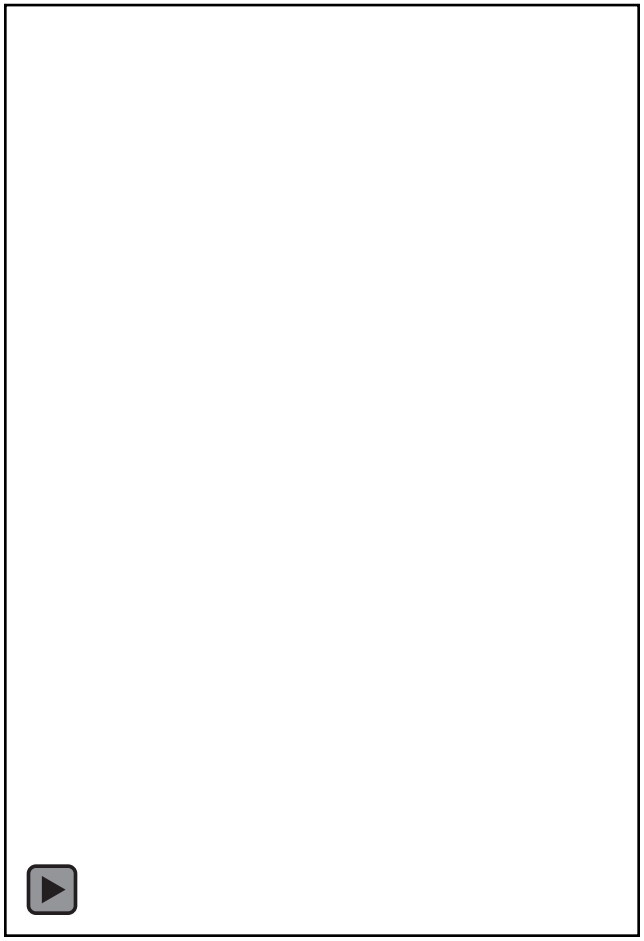
➔ *Several techniques are available, but a supervisory controller has to decide what to do.*

F. Cai and X. Koutsoukos, Real-time out-of-distribution detection in learning-enabled cyber-physical systems, in 2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS), 2020, pp. 174–183.

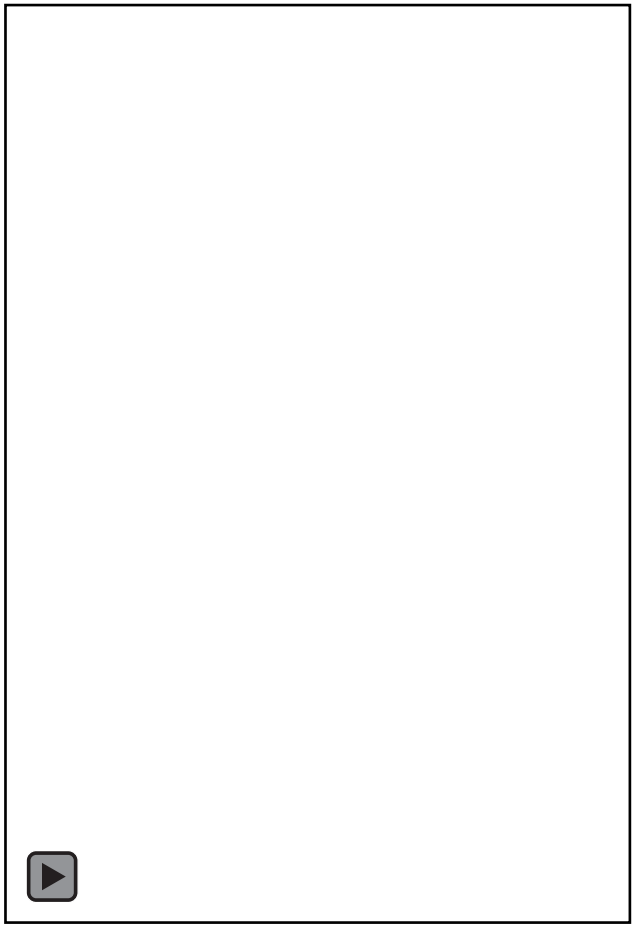
3.Assurance monitoring of LECs

Simulation results - Adversarial input

No attack



Attack



4. Design toolsuite for CPS+LECs

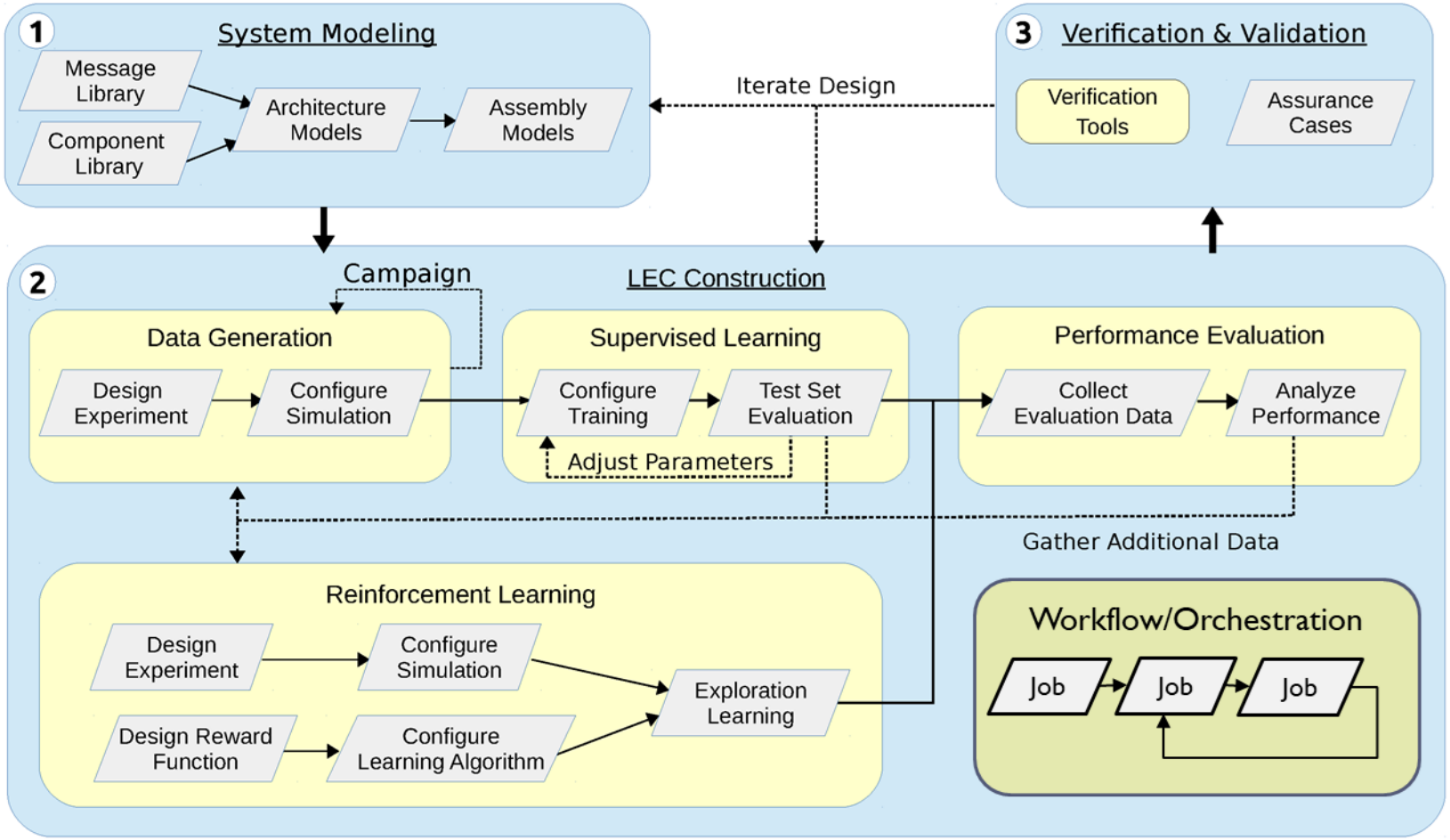
- In addition to MBSE tasks, design tools must support
  - All aspects of LEC construction: training, testing, evaluation, assurance, …
    - ➔ Automation for typical workflows
  - Integration of assurance activities
    - ➔ Assurance provenance: manage and preserve assurance artifacts (including training data)

➔ *Classic MBSE must be extended to support (1) LEC construction and (2) assurance.*

C. Hartsell, N. Mahadevan, S. Ramakrishna, A. Dubey, T. Bapty, T. T. Johnson, X. D. Koutsoukos, J. Sztipanovits, and G. Karsai, Model-based design for CPS with learning-enabled components, in Proceedings of the Workshop on Design Automation for CPS and IoT, DESTION@CPSIoTWeek 2019, Montreal, QC, Canada, 2019, pp. 1–9.

4. Design toolsuite for CPS+LECs

- MBSE extended for LEC development



4. Design toolsuite for CPS+LECs

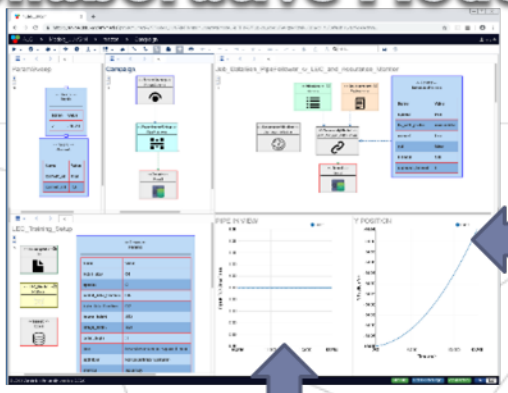
VANDERBILT UNIVERSITY
INSTITUTE FOR SOFTWARE INTEGRATED SYSTEMS

ALC Workflows

Collaborative Modeling

System Integrator

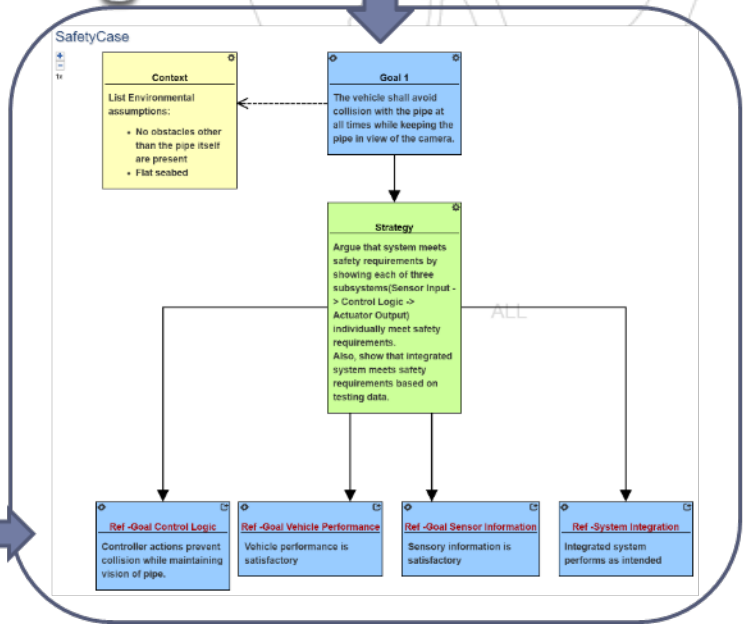
LEC Developer



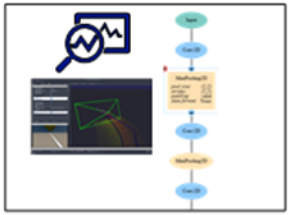
Design Time

Assurance

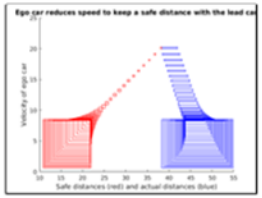
- *The model driven toolchain supports training, verification and design-time assurance of learning enabled components.*
- *Toolchain helps with developing safety assurance cases for the system using collected evidence.*
- *Complete provenance tracking of experimental runs and data collection is supported.*



Execution, Training, Data Collection, Verification

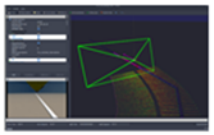


{d} Run System with LEC



{c} Verify Closed-Loop System With LEC

{b} Train LECs

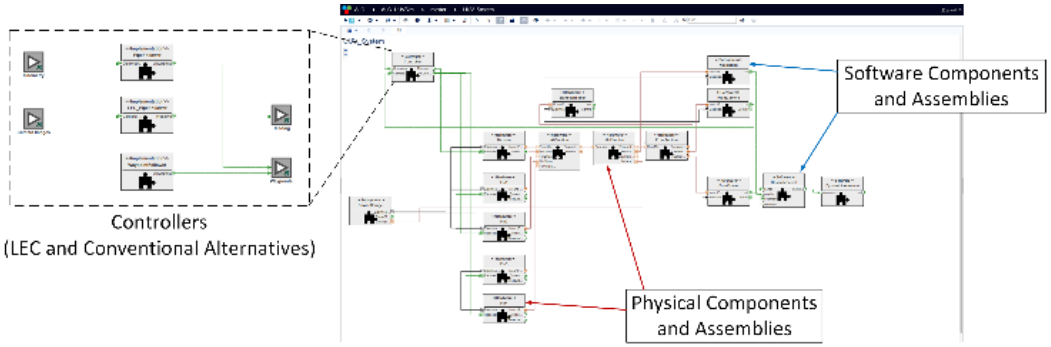


{a} Run Scenarios To Collect Data

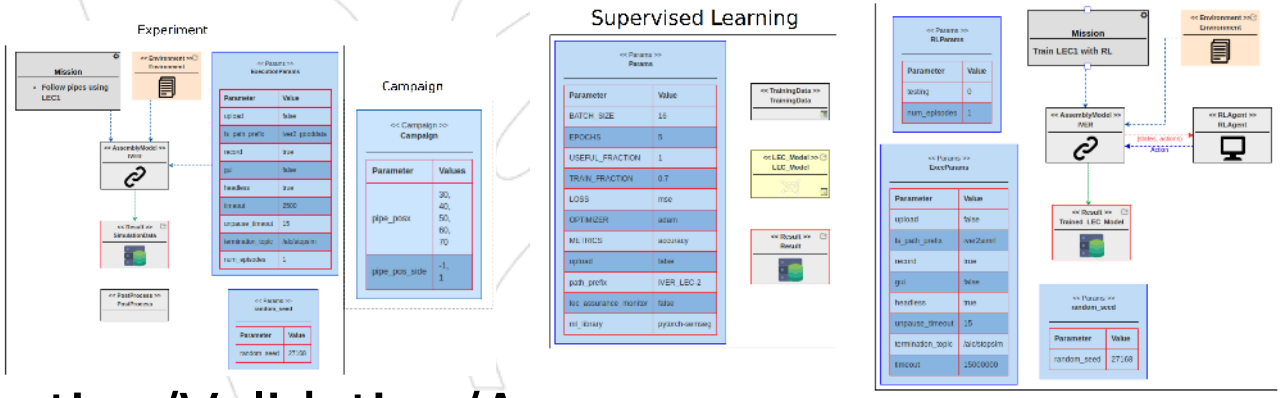
Typical Workflow Sequence

4. Design toolsuite for CPS+LECs

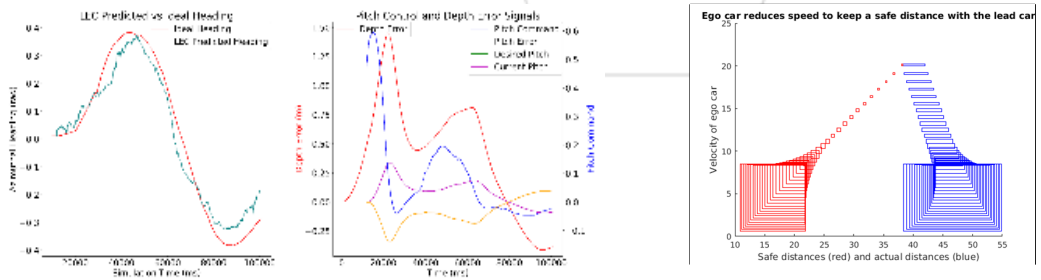
- Modeling
  - System Architecture / Sys

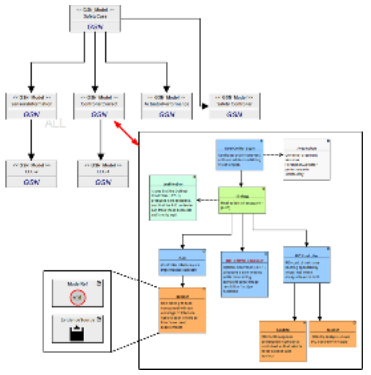


- LEC Construction
  - Data collection
  - Training
  - Evaluation



- Testing -- Verification/Validation/Assurance

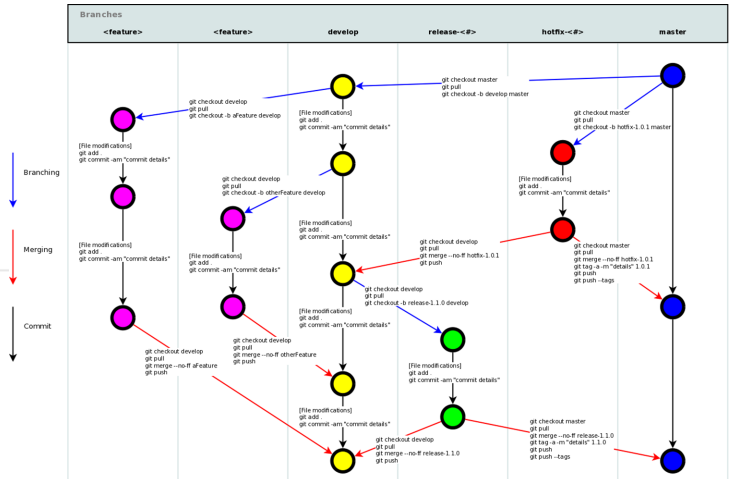




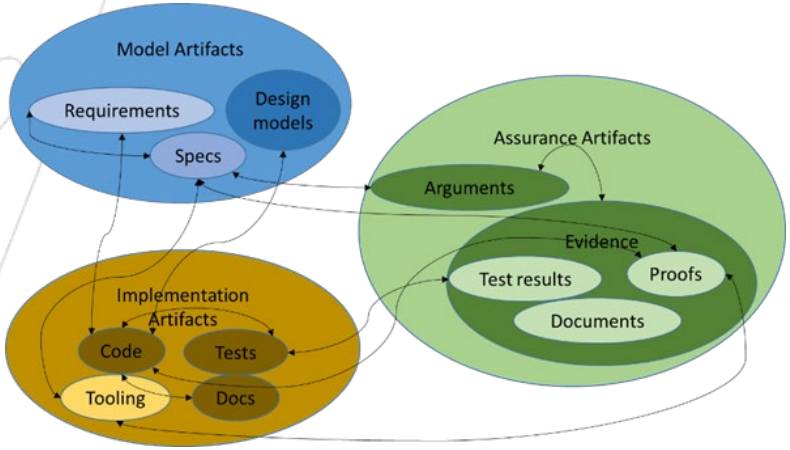
4. Design toolsuite for CPS+LECs

The need for assurance provenance

1. The artifacts are produced (and maintained) in a continuous development process
   - Version controlled, continuous development and integration



2. The artifacts are in complex dependency relationships
   - Explicit representation and management of these dependencies is inevitable



➔ *The toolchain must support Continuous Integration/Assurance/Deployment.*

Karsai, Gabor, and Daniel Balasubramanian. "Assurance provenance: the next challenge in software documentation." In Leveraging Applications of Formal Methods, Verification and Validation. Software Engineering: 11th International Symposium, ISoLA 2022, Rhodes, Greece, October 22–30, 2022, Proceedings, Part II, pp. 90-104. Cham: Springer Nature Switzerland, 2022.

Summary

- LEC-s are becoming part of the CPS 'toolbox'
  - Their verification and assurance is a challenge
- LEC-s are dependent on the training data
  - Tooling is needed for automation and maintaining artifact provenance
- Assurance of individual components is necessary but not sufficient
  - Ideas and tooling are needed for compositional, system-level assurance
  - Design/implementation/verification artifacts must be linked to assurance case(s) to provide continuous assurance provenance

More information:

- https://assured-autonomy.isis.vanderbilt.edu/
- https://assured-autonomy.org/
- https://github.com/AbLECPS/
- https://ablecps.github.io/