Research Directions: Cyber-Physical Systems

www.cambridge.org/cbp

Question

Cite this article: Paoletti N and Woodcock J (2023). How to ensure safety of learning-enabled cyber-physical systems? *Research Directions: Cyber-Physical Systems*. **1**, e2, 1–2. https://doi.org/10.1017/cbp.2023.2

Received: 2 February 2023
Accepted: 2 February 2023

Author for correspondence:
Nicola Paoletti,
Email: Nicola.paoletti@kcl.ac.uk

How to ensure safety of learning-enabled cyber-physical systems?

Nicola Paoletti[1] and Jim Woodcock[2]

[1]Department of Informatics, King's College London, UK and [2]Department of Computer Science, University of York, UK

Context

Modern cyber-physical systems (CPS) integrate machine learning and deep learning components for a variety of tasks, including sensing, control, anomaly detection and learning process dynamics from data. Formal verification of CPS is paramount to ensure their correct behaviour in many safety-critical application domains (including robotics, avionics, health and automotive). However, traditional CPS verification methods are designed to work with mechanistic CPS models, and hence cannot deal in general with data-driven components. Therefore, how to guarantee the correct behaviour of learning-enabled CPS is still an open question which must be addressed in order to deploy these systems in real-world safety-critical settings.

Within this research question, we welcome contributions about the formal analysis of learning-enabled CPSs. Examples include but are not limited to:

- Verifying systems with machine-learning components in the loop (including model checking and theorem proving).
- Formal reasoning about the model's (epistemic) uncertainty.
- Learning safe data-driven models for CPS monitoring and control.
- Dealing with distribution shifts induced by (possibly adversarial) runtime changes of the CPS (e.g., detecting when existing system properties cease to hold and adapting the verification methods to account for such runtime changes).

How to contribute to this Question

If you believe you can contribute to answering this Question with your research outputs, find out how to submit in the Instructions for authors (https://www.cambridge.org/core/journals/research-directions-cyber-physical-systems/information/author-instructions/preparing-your-materials). This journal publishes Results, Analyses, Impact papers and additional content such as preprints and "grey literature". Questions will be closed when the editors agree that enough has been published to answer the Question so before submitting, check if this is still an active Question. If it is closed, another relevant Question may be currently open, so do review all the open Questions in your field. For any further queries, check the information pages (https://www.cambridge.org/core/journals/research-directions-cyber-physical-systems/information/about-this-journal) or contact this email (cps@cambridge.org).

Competing interests. The authors declare none.

References

Motivation and challenges

Seshia SA, Sadigh D and Sastry SS (2022) Toward verified artificial intelligence. *Communications of the ACM* **65**, 46–55.

State-of-the-art in machine learning and neural network certification, including CPS with neural controllers

Huang X, *et al.* (2020) A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability. *Computer Science Review* **37**, 100270.

Tambon F, *et al.* (2022) How to certify machine learning based safety-critical systems? A systematic literature review. *Automated Software Engineering* **29**, 1–74.

Tran H-D, Xiang W and Johnson TT (2020) Verification approaches for learning-enabled autonomous cyber-physical systems. *IEEE Design & Test* **39**, 24–34.

Liu C, *et al.* (2021) Algorithms for verifying deep neural networks. *Foundations and Trends® in Optimization* **4**, 244–404.

Li L, *et al.* (2020). SoK: Certified robustness for deep neural networks. ArXiv preprint arXiv:2009.04131.

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Verification approaches dealing with uncertainty and probabilistic guarantees

Lauri M, Hsu D and Pajarinen J (2022) Partially observable Markov decision processes in robotics: A survey. *IEEE Transactions on Robotics*.

Jansen N, *et al.* (2020) Safe reinforcement learning using probabilistic shields. In *31st International Conference on Concurrency Theory (CONCUR 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Cleaveland M, *et al.* (2022) Risk verification of stochastic systems with neural network controllers. *Artificial Intelligence* **313**, 103782.

Dixit A, *et al.* (2022) Adaptive Conformal Prediction for Motion Planning among Dynamic Agents. ArXiv preprint arXiv:2212.00278.

Cairoli F, Paoletti N and Bortolussi L (2022) Conformal Quantitative Predictive Monitoring of STL Requirements for Stochastic Processes. ArXiv preprint arXiv:2211.02375.

Bortolussi L, *et al.* (2021) Neural predictive monitoring and a comparison of frequentist and Bayesian approaches. *International Journal on Software Tools for Technology Transfer* **23**, 615–640.

Akintunde ME, *et al.* (2022) Formal verification of neural agents in non-deterministic environments. *Autonomous Agents and Multi-Agent Systems* **36**, 1–36.

Yan R, *et al.* (2022) Strategy Synthesis for Zero-Sum Neuro-Symbolic Concurrent Stochastic Games. ArXiv preprint arXiv:2202.06255.

Bacci E and Parker D (2020) Probabilistic guarantees for safe deep reinforcement learning. In *International Conference on Formal Modeling and Analysis of Timed Systems*. Springer, Cham.

Wicker M, *et al.* (2021) Certification of iterative predictions in Bayesian neural networks. In *Uncertainty in Artificial Intelligence*. PMLR.

https://doi.org/10.1017/cbp.2023.2 Published online by Cambridge University Press