MODPROD

Generative AI in Simulation-Based Test Environments for Large-Scale Cyber-Physical Systems: An Industrial Study

Masoud Sadrnezhaad, José Antonio Hernández López, Torvald Mårtensson, Dániel Varró

20th MODPROD Workshop on Model-based Cyber-physical Product Development
February 4, 2026

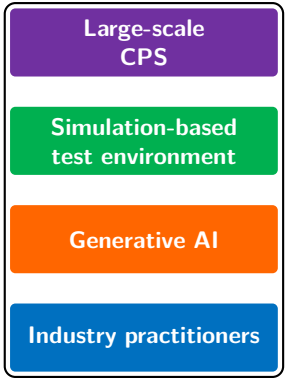
Table of Contents

MODPROD

Introduction

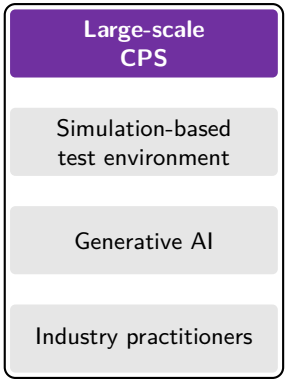
- **Research Question**

  What is the potential of generative AI in simulation-based test environments for large-scale cyber-physical systems, according to industry practitioners?



Large-scale Cyber-Physical Systems

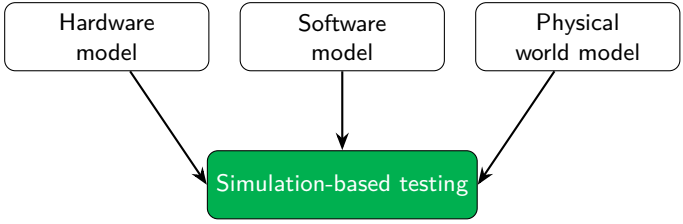
- Modern CPSs such as aircraft, autonomous trucks, and telecom networks combine:
  - hardware
  - software
  - physical processes
- They require:
  - a wide variety of heterogeneous tests
  - numerous simulators



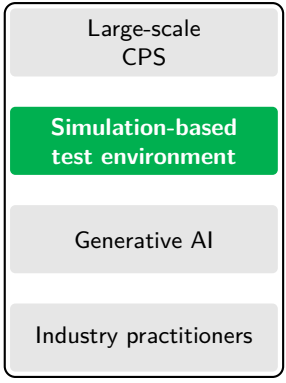
Simulation-based Test Environments

MODPROD

- Engineers rely on detailed models of hardware, software, and the physical world
  - to test safely and cheaply



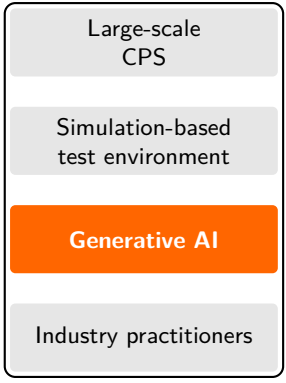
- Extensive resources are required
  - to develop and maintain simulation models
  - cost rises as systems grow



Generative AI Opportunities

MODPROD

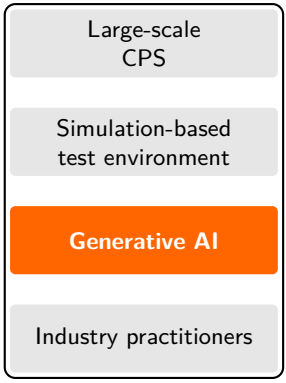
- Recent advances (e.g. ChatGPT, Copilot)
  - Generate source *code*
  - Produce *test cases*
  - Analyse *test results*
- Promised benefits in software projects
  - Reducing manual effort
  - Increasing test coverage
- Raise question:
  - Will these gains carry over to CPS testing?
    → *"Can generative AI cut the cost of simulation-based testing for large-scale CPSs?"*



Generative AI Opportunities

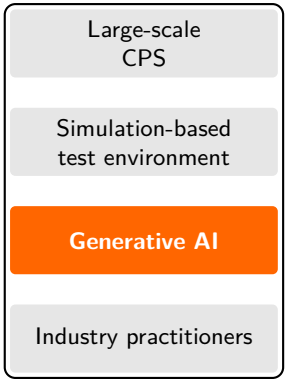
MODPROD

- Recent advances (e.g. ChatGPT, Copilot)
  - Generate source *code*
  - Produce *test cases*
  - Analyse *test results*
- Promised benefits in software projects
  - Reducing manual effort
  - Increasing test coverage
- Raise question:
  - Will these gains carry over to CPS testing?
    → *"Can generative AI cut the cost of simulation-based testing for large-scale CPSs?"*



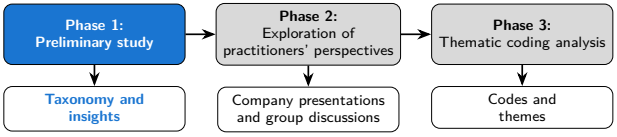
Generative AI Opportunities

- Recent advances (e.g. ChatGPT, Copilot)
  - Generate source *code*
  - Produce *test cases*
  - Analyse *test results*
- Promised benefits in software projects
  - Reducing manual effort
  - Increasing test coverage
- Raise question:
  - Will these gains carry over to CPS testing?
    → *"Can generative AI cut the cost of simulation-based testing for large-scale CPSs?"*

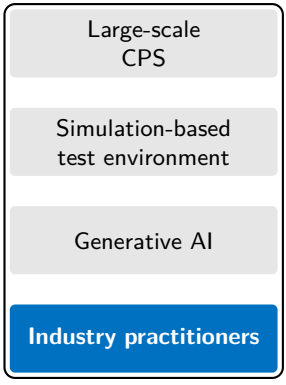


Research Method

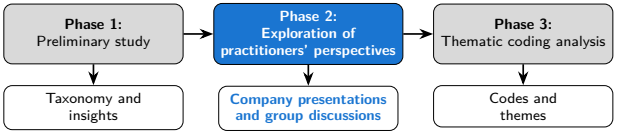
Research Method



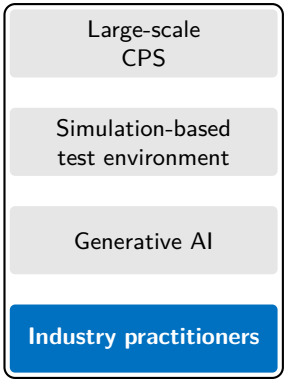
- Literature snowballing
- Taxonomy and classification
- Key insights extracted
- Material shared with participants



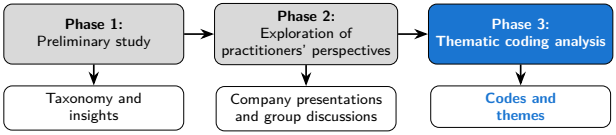
Research Method



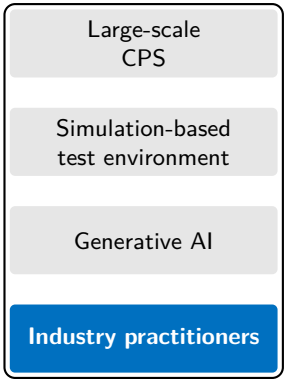
- One-day, in-person workshop
- 6 companies | 33 participants
- 8 company presentations
- 5 cross-company discussion groups



Research Method



- Thematic coding analysis of workshop presentations and group discussions
- Two independent coders
- Two reviewers for validation
- Three main themes identified



Key Contributions

Key Contributions

Industry-grounded insights

into the **current challenges** of simulation-based testing

Actionable research agenda

structured around the **three high-impact directions** shown on the right to support informed adoption decisions

AI-generated scenarios & environment models

- Novel scenario generation
- Environment model generation
- Scenario evaluation & selection

Simulators & AI in CI/CD pipelines

- Large-scale data handling
- Continuous integration
- Simulators interoperability

Trustworthiness of generative AI for simulation

- Model/Simulation fidelity
- Over-reliance
- Standardisation & regulation
- Traceability

Novel scenario generation

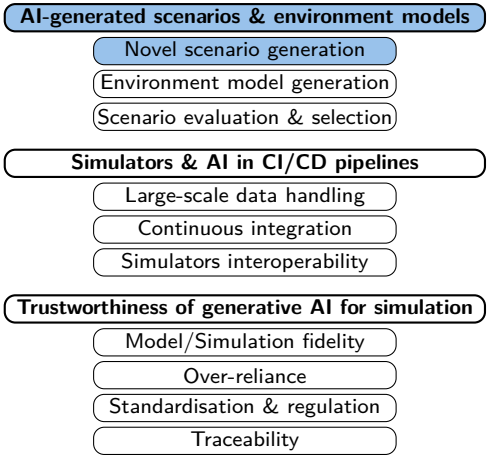
MODPROD

Hopes and challenges

- Produce rare or near-accident scenarios
- Move beyond rigid rule-based frameworks
- Increase coverage without extra logging

Example(s) from industry

- "Create more extreme and challenging test scenarios from critical incidents in our real-world data"



Environment model generation

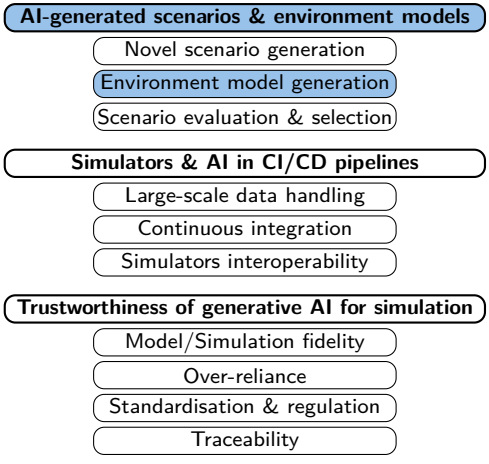
MODPROD

Hopes and challenges

- Generate high-fidelity environment models
- Build accurate digital twins of real environments
- Reduce manual effort in physical modelling
- Support sensor realism for training and validation
- Ensure realism and avoid violating physics constraints

Example(s) from industry

- AI-built customer-network twins for secure tests



Scenario evaluation & selection

MODPROD

Hopes and challenges

- Rank generated scenarios by novelty and realism
- Prioritise runs when compute resources are limited
- Close feedback loop between generation and execution
- Define metrics for novelty and realism
- Demands having strong domain-specific insights and a deep understanding of the operational context

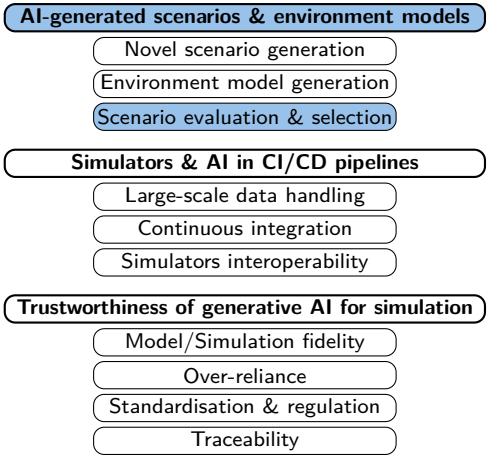
Example(s) from industry

- Demand for automatic scenario-quality metrics

Example(s) from literature

- RAG-based LLM prioritises faulty scenarios

(10.1109/AITEST62860.2024.00014)



Large-scale data handling

MODPROD

Hopes and challenges

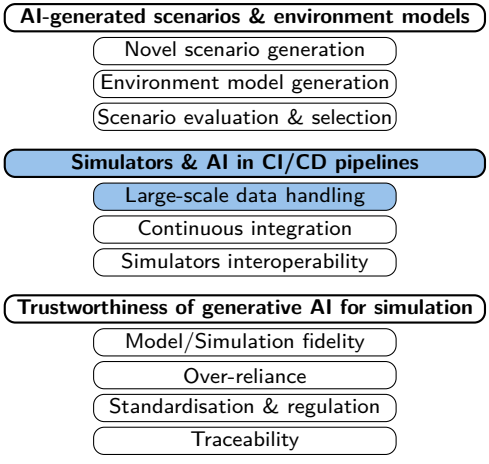
- Process large-scale operational logs
- Clean, label, and store signals efficiently
- Feed scalable AI training and validation loops

Example(s) from industry

- Managing tens of thousands of vehicle signals

Example(s) from literature

- RAG to detect anomalies and faults in CPSs
  (10.1109/AITEST62860.2024.00014)



Continuous integration & simulators interoperability

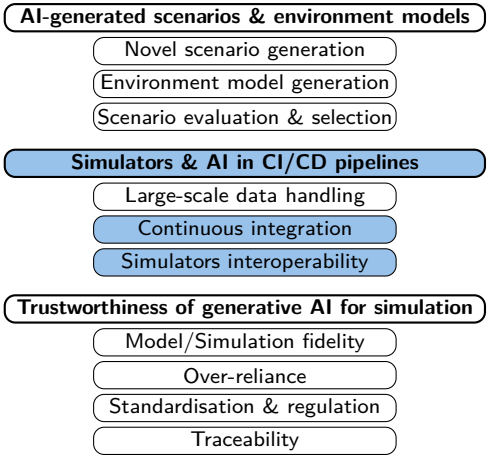
MODPROD

Hopes and challenges

- Embed AI models and simulators in CI/CD pipelines
- Automate retraining and validation on every commit
- Provide continuous feedback for CPS development
- Connect and orchestrate standalone simulators for end-to-end tests

Example(s) from industry

- Challenges in adapting AI tools to established CI processes
- Capturing subtle, high-level phenomena and modeling complex system-of-systems interactions remains difficult



Model/Simulation fidelity

MODPROD

Hopes and challenges

- Guarantee that AI-generated simulations can be used in safety-critical contexts
- Align simulation outputs with real-world behaviour
- Detect divergence between simulation and real-world data
- Maintain accuracy as system complexity grows

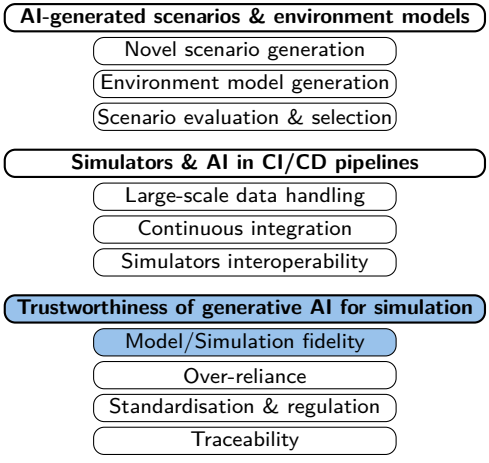
Example(s) from industry

- Cross-checking flight traces against simulations

Example(s) from literature

- Digital-twin loop corrects simulation via feedback

(10.1109/ETFA61755.2024.10710900)



Over-reliance, Standardisation & Traceability

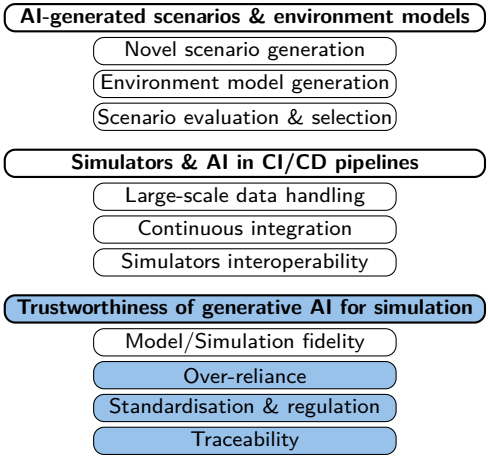
MODPROD

Hopes and challenges

- Provide systematic oversight and mitigate hallucination
- Enforce internal standards and external regulations
- Maintain traceability between generated artefacts and source requirements

Example(s) from industry

- Warned about over-reliance on AI-generated code in production and risks of potentially poor-quality output



Conclusion

Three take-aways

Focus on the three priority directions

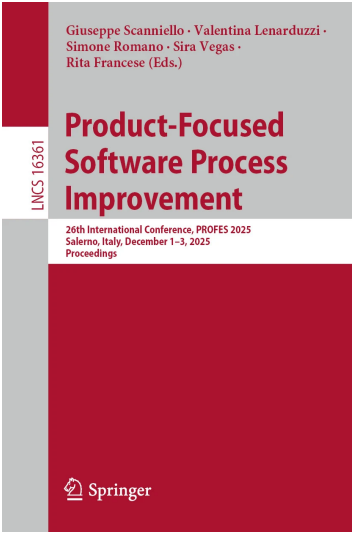
AI-generated scenarios & environment models, simulators & AI inside CI/CD pipelines, and trustworthiness of generative AI for simulation

Move from exploratory use to continuous practice

Industry needs evaluation mechanisms and seamless pipeline integration before generative-AI solutions can leave the "one-off demo" stage

Deepen industry–academia collaboration

Practitioners lack a concrete roadmap; producing visible proofs-of-concept and robust metrics is essential to unlock responsible, large-scale adoption

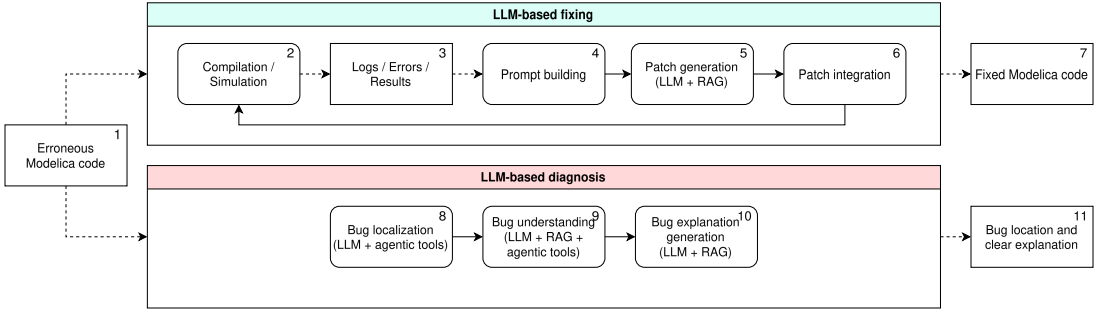


Full paper

Poster

[1] M. Sadrnezhaad, J. A. H. López, T. Mårtensson, and D. Varró, "Generative AI in simulation-based test environments for large-scale cyber-physical systems: An industrial study," in *Product-Focused Software Process Improvement - 26th International Conference, PROFES 2025, Salerno, Italy, December 1-3, 2025, Proceedings*, ser. Lecture Notes in Computer Science, vol. 16361, Springer, 2025, pp. 203–219.

- **Problem:** Debugging CPS simulation models is difficult
- **Motivation:** Existing diagnostics lack explanations aligned with the modeller's conceptual view of the system
- **Approach:** Agentic LLM-based diagnosis and automated repair



Thank you for your attention. ☺

MODPROD

- **Research Question**

  What is the potential of generative AI in simulation-based test environments for large-scale cyber-physical systems, according to industry practitioners?

- **Contributions**

Industry-grounded insights
into the **current challenges** of simulation-based testing

Actionable research agenda
structured around the **three high-impact directions**

Full paper

My LinkedIn

AI-generated scenarios & environment models
- Novel scenario generation
- Environment model generation
- Scenario evaluation & selection

Simulators & AI in CI/CD pipelines
- Large-scale data handling
- Continuous integration
- Simulators interoperability

Trustworthiness of generative AI for simulation
- Model/Simulation fidelity
- Over-reliance
- Standardisation & regulation
- Traceability