Developing an AI-enabled Industry 4.0 platform – Performance experiences on deploying AI onto an industrial edge device

Holger Eichelberger, Gregory Palmer*, Claudia Niederée*

Software Systems Engineering
University of Hildesheim

L3S Research Center
University of Hannover

eichelberger@sse.uni-hildesheim.de

Gefördert durch:

Bundesministerium
für Wirtschaft
und Energie

Motivation (1)

Ecosystem on **easy-to-use AI** in the **industrial production** focusing on interoperability, vendor-neutrality, openness

Talk by Ahmad

Talk by Christian

Motivation (2)

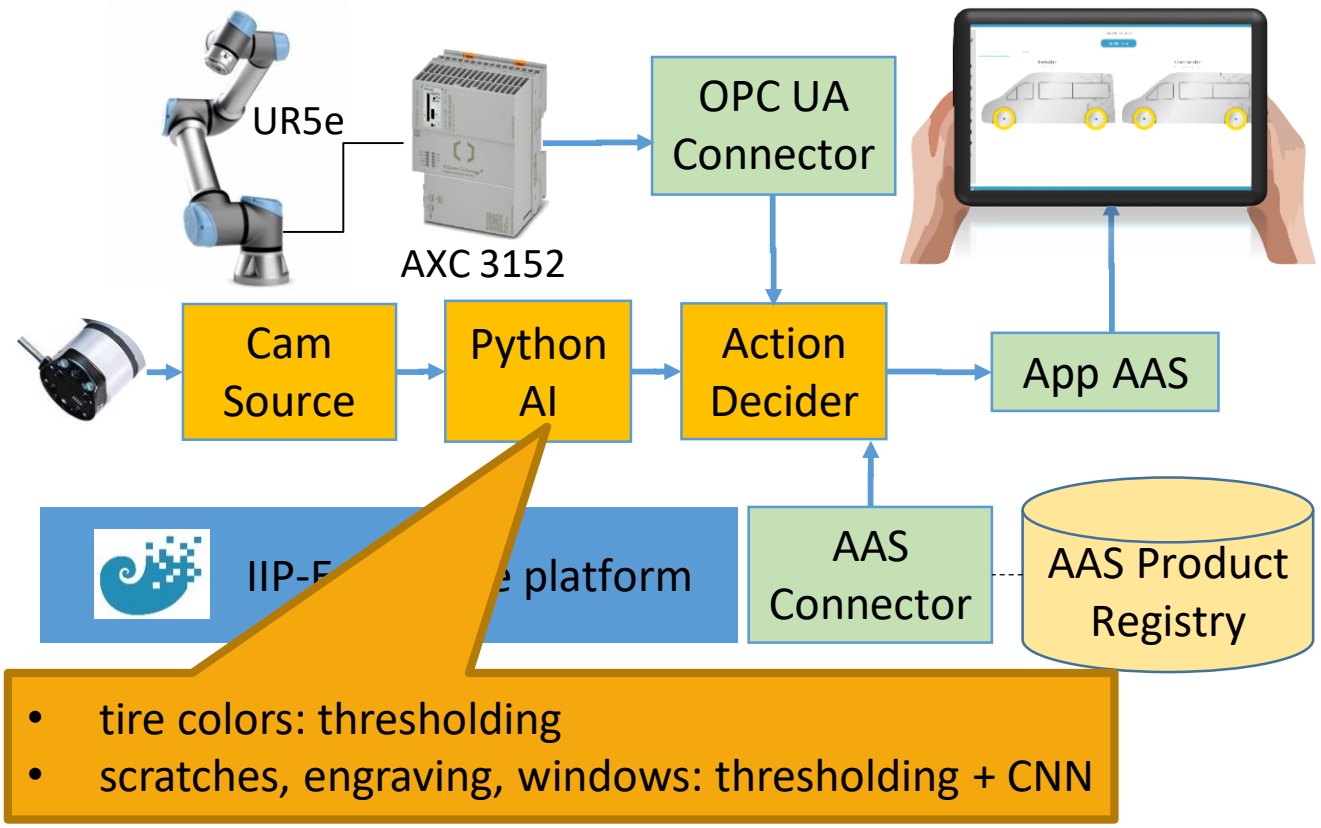
Focus: AI-enabled IIoT platform and it‘s early evaluation

- Industrial hardware, e.g., edge devices
- AI based on usual frameworks
- Public demonstrator for visual quality inspection
- Dynamic deployment and resulting performance effects

Performance characterization, not a formal experiment

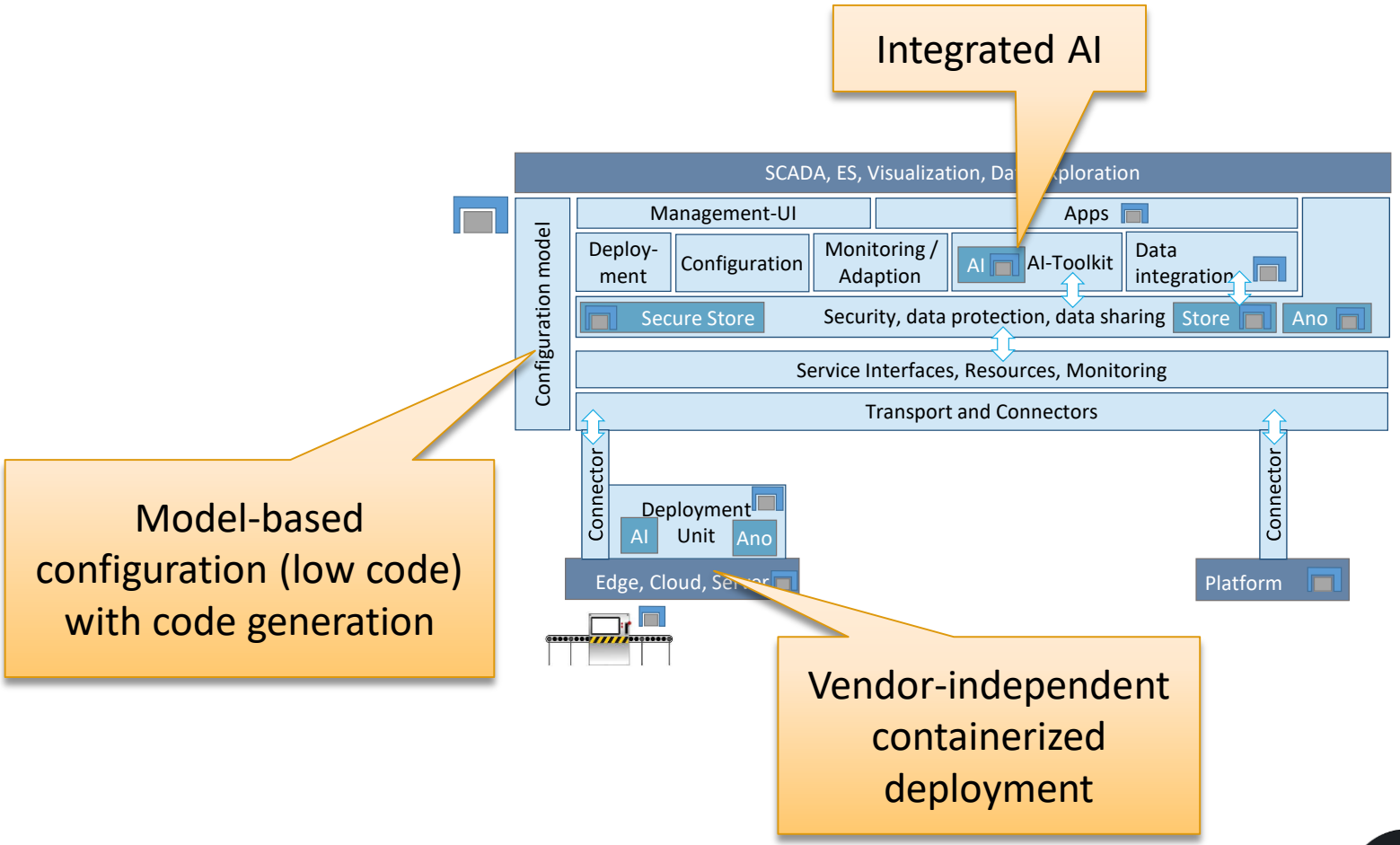
Demonstrator / Subject (1)

Demonstrator / Subject (2)



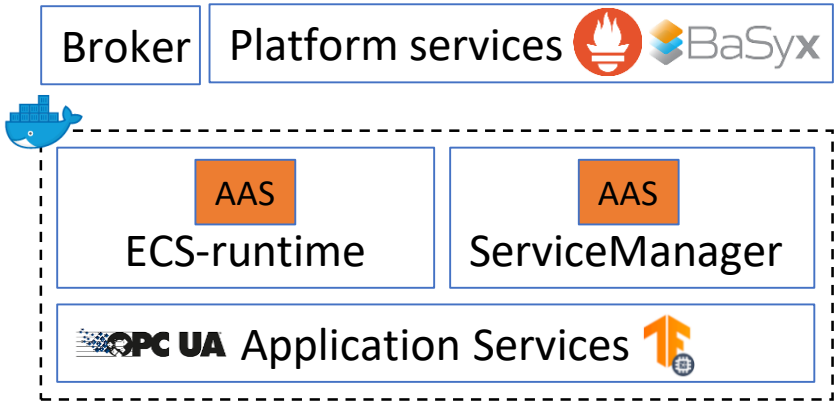
H. Eichelberger, G. Palmer, S. Reimer, T. T. Vu, H. Do, S. Laridi, A. Weber, C. Niederée, and T. Hildebrandt, Developing an AI-enabled IIoT platform - Lessons learned from early use case validation, SASI4, 2022

IIP-Ecosphere IIoT Platform

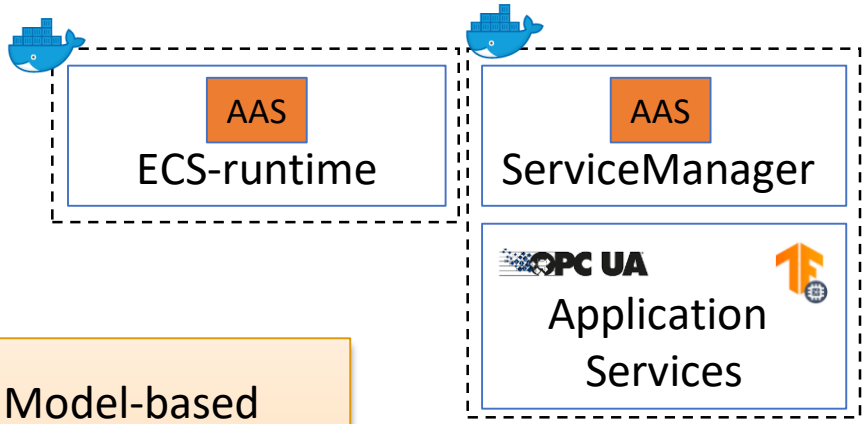


Distributed deployment

Central installation (Server)



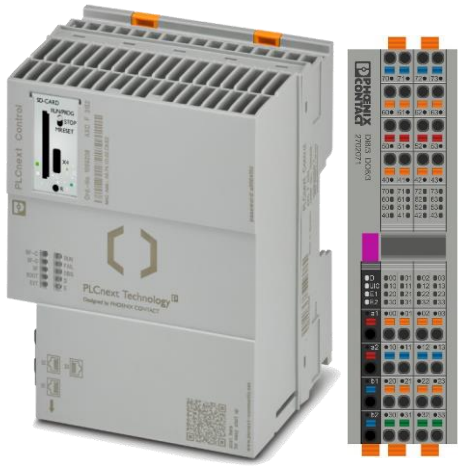
Distributed installation (Device, Edge, …)



Model-based generation of containers

Experiment setup

- Hardware
  - **Edge:** Phoenix Contact AXC 3152, 2 core Intel Atom, 2 GB RAM, 32 GB SDD HDD, Linux
  - **„Server“:** Intel Core i5-7400 3.00 GHz, 32 GByte RAM, 2 TByte SDD, Debian 11 Linux
- AI
  - Python 3.8.10
  - Tensorflow lite 2.8.0
- Experiment variants
  - AI on the edge with broker
  - AI on the edge without broker
  - AI on the „Server“, camera / OPC source on the edge



More devices may lead to better separation of results.

[Pictures: PHOENIX CONTACT]

Experiment result

Variant	AI @ Edge, no broker	AI @ Edge, broker	AI @ „Server“
first	2.13 s	2.20 s	0.61 s
min	1.79 s	1.80 s	0.38 s
avg	1.83 s	1.84 s	0.51 s
stddev	0.03 s	0.02 s	0.14 s
max	1.89 s	1.91 s	0.83 s

- Measurements by manual Python instrumentation, Prometheus
- 10 repetitions of full inspection process
- Baseline on edge: 1 s

- Edge runs „hot“: 95% memory usage

DOI 10.5281/zenodo.7157607

Discussion

- Initial measurements: 3-5 s on the edge
- Within the demonstrator: 50 s per quality inspection
- In a real setting: 8 ms machine pace
- Potential influences / optimizations
  - Different AI library
  - Smaller container, modularized dependencies
  - Code and AI model optimization
  - Hardware acceleration, e.g., TPU, GPU, FPGA

[Pictures: PHOENIX CONTACT]

Conclusion and Future Work

- I4.0 / CPPS is integration of software, hardware, standards
- AI @ edge is a vision, … , a requirement
- Successfully deployed ordinary AI to industrial edge
- Performance …

- Future
  - Dig into optimization alternatives
  - Easy-to-use AI, federated learning

IIP-Ecosphere
Next Level Ecosphere for Intelligent Industrial Production