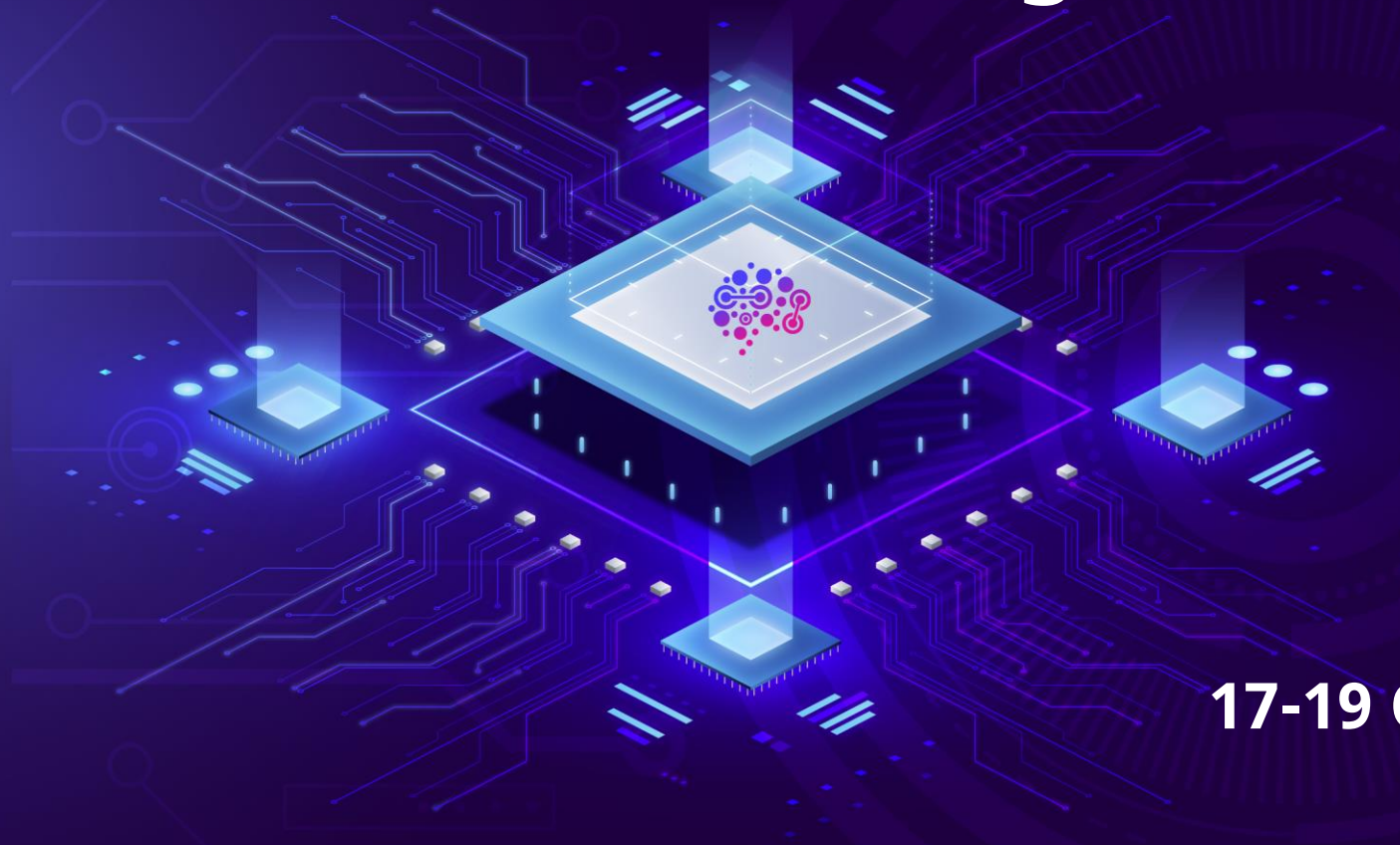
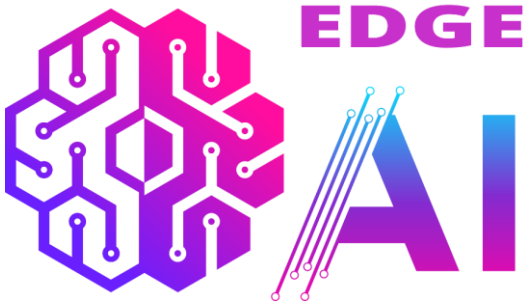
European Conference on EDGE AI Technologies and Applications - EEAI



Advancing emerging edge AI technologies and driving next generation intelligent applications.

17-19 October 2023, Athens, Greece

European Conference on EDGE AI Technologies and Applications - EEAI



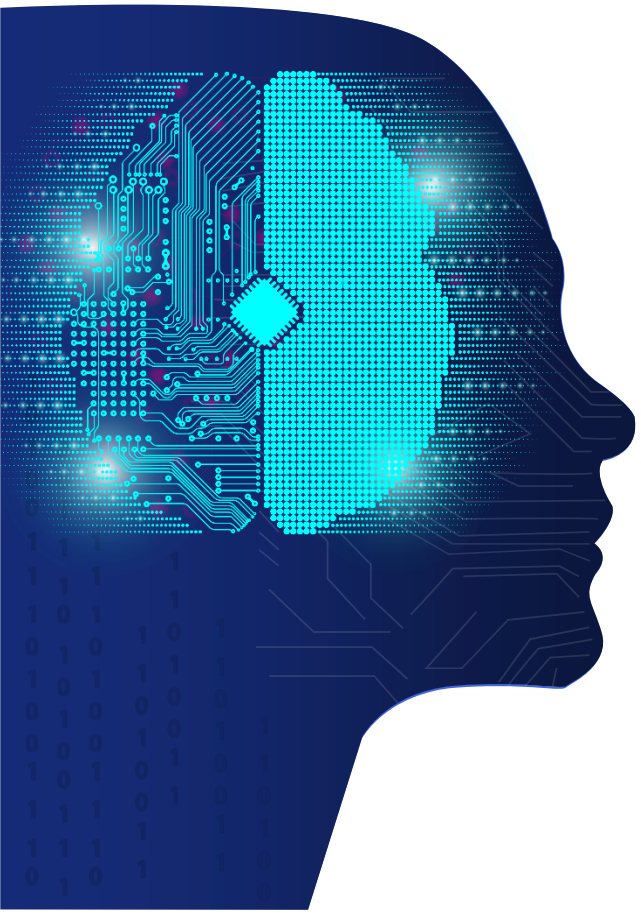
Outlook on the Deployment and Execution of Mixed-Criticality Edge-AI Applications

Hans Dermot Doran, **Suzana Veljanovska**, ZHAW, Switzerland

17-19 October 2023 Athens, Greece

Copyright © 2023

Presentation Outline



- Introduction
- Research Goals
- State-of-the-art
- Proposals
- Findings

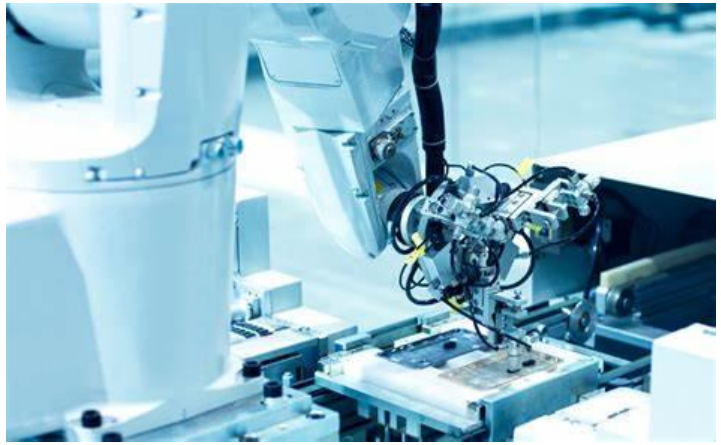
Copyright © 2023

Introduction

- Application domain includes functional safety and security
  - Aeronautics (DO178C and DO254)
  - Factory automation (IEC61508)

➢ **Mixed criticality** system

- Safety-critical and non-critical sections implemented on a common hardware platform



Copyright © 2023

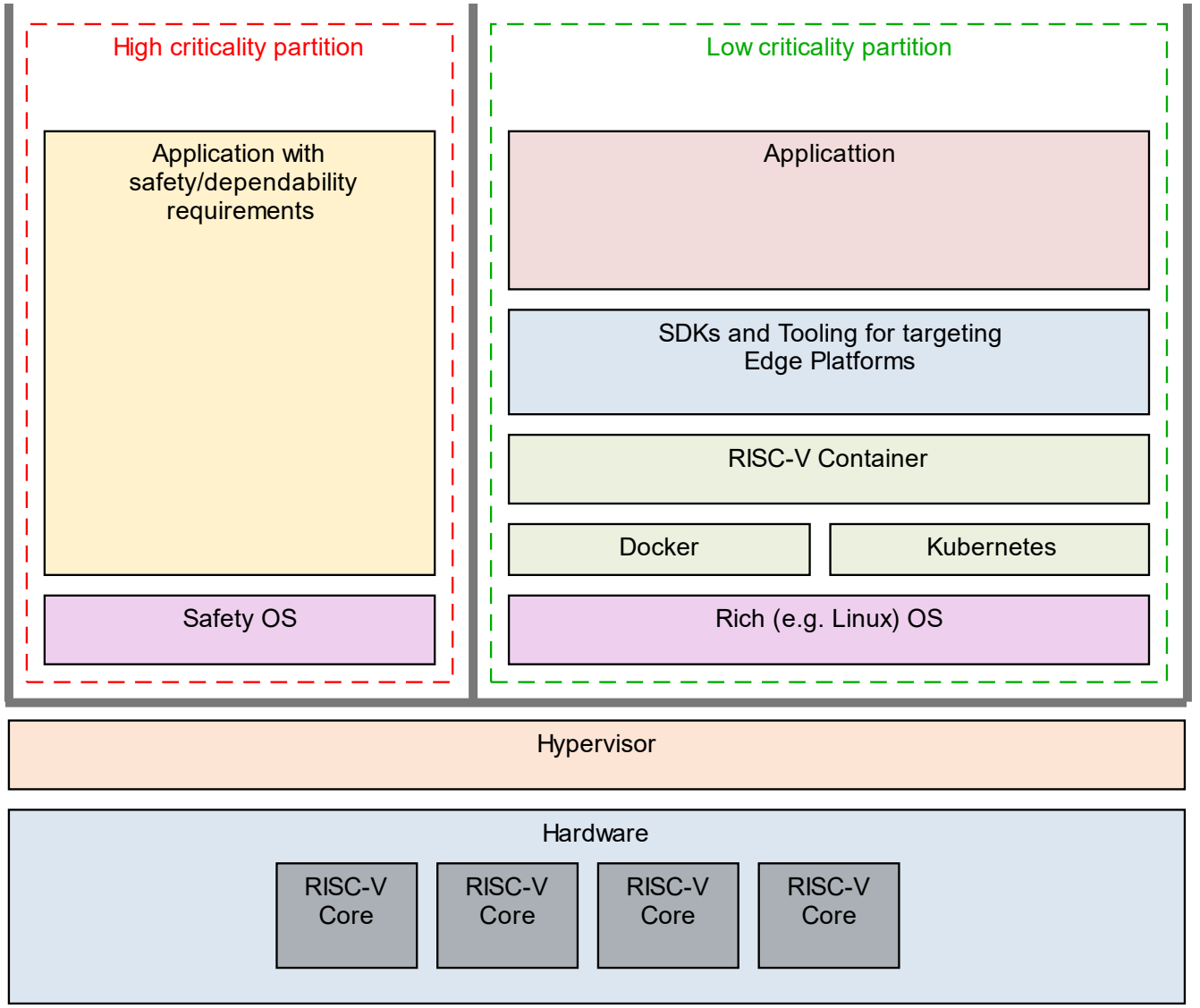
Research Goals

- Overarching Research Goal – in the context of REBECCA project
  - Application-domain high-criticality execution
- Current Research Goal
  - Design of a multi-core mixed criticality system architecture for eventual safe and dependable AI algorithm execution
- Current Problem
  - Deployment of mixed criticality AI system

Copyright © 2023

State of the Art

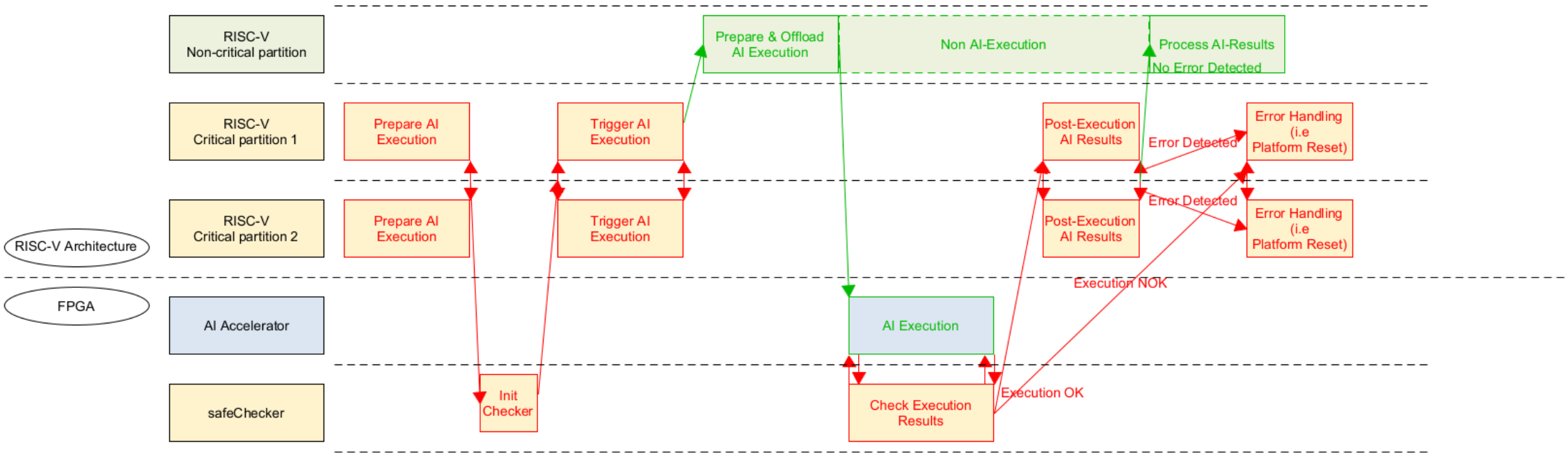
- Typical mixed-criticality system
- Consists of two or more partitions:
  - Critical (safe/secure) partition
  - Non-critical partition
- Hypervisor implements a separation kernel
  - Ensures memory and execution integrity
- REBECCA will integrate AI accelerators, both off-chip and hosted in an FPGA.
- Require AI-aware mixed criticality system



Copyright © 2023

High Criticality Execution of AI-Model: Schedule

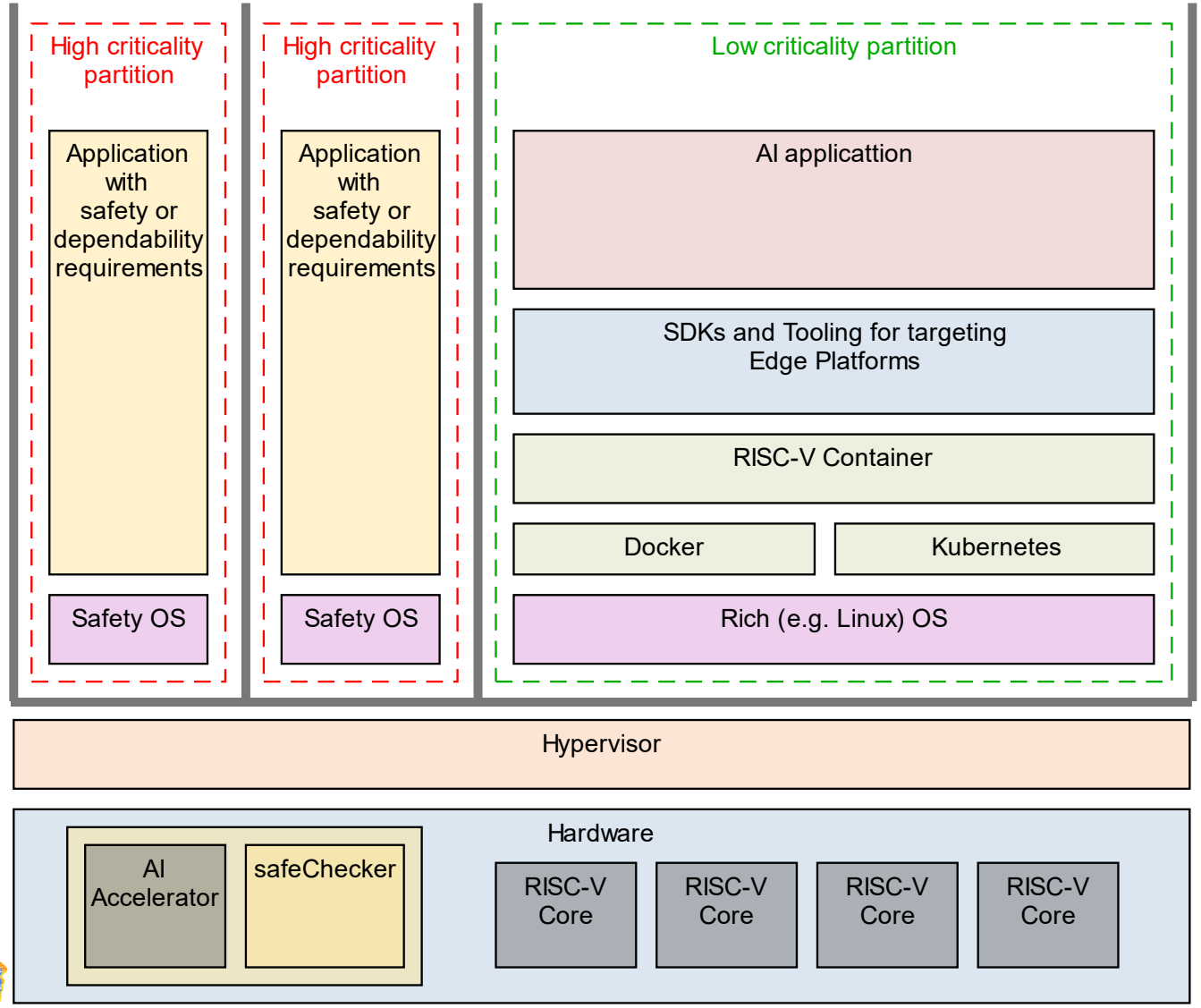
- “Mixed criticality” on a non-intrinsically dependable platform implies the use of redundancy which implies two high criticality hypervisor partitions
- Redundancy implies twofold execution which implies communication between hypervisor partitions
- Requires an FPGA unit to check local dependable operation of AI accelerator -> safeChecker
- These implications suggest the schedule of operations below (simplified - w/o extensive exception handling)



Copyright © 2023

Proposed Architecture

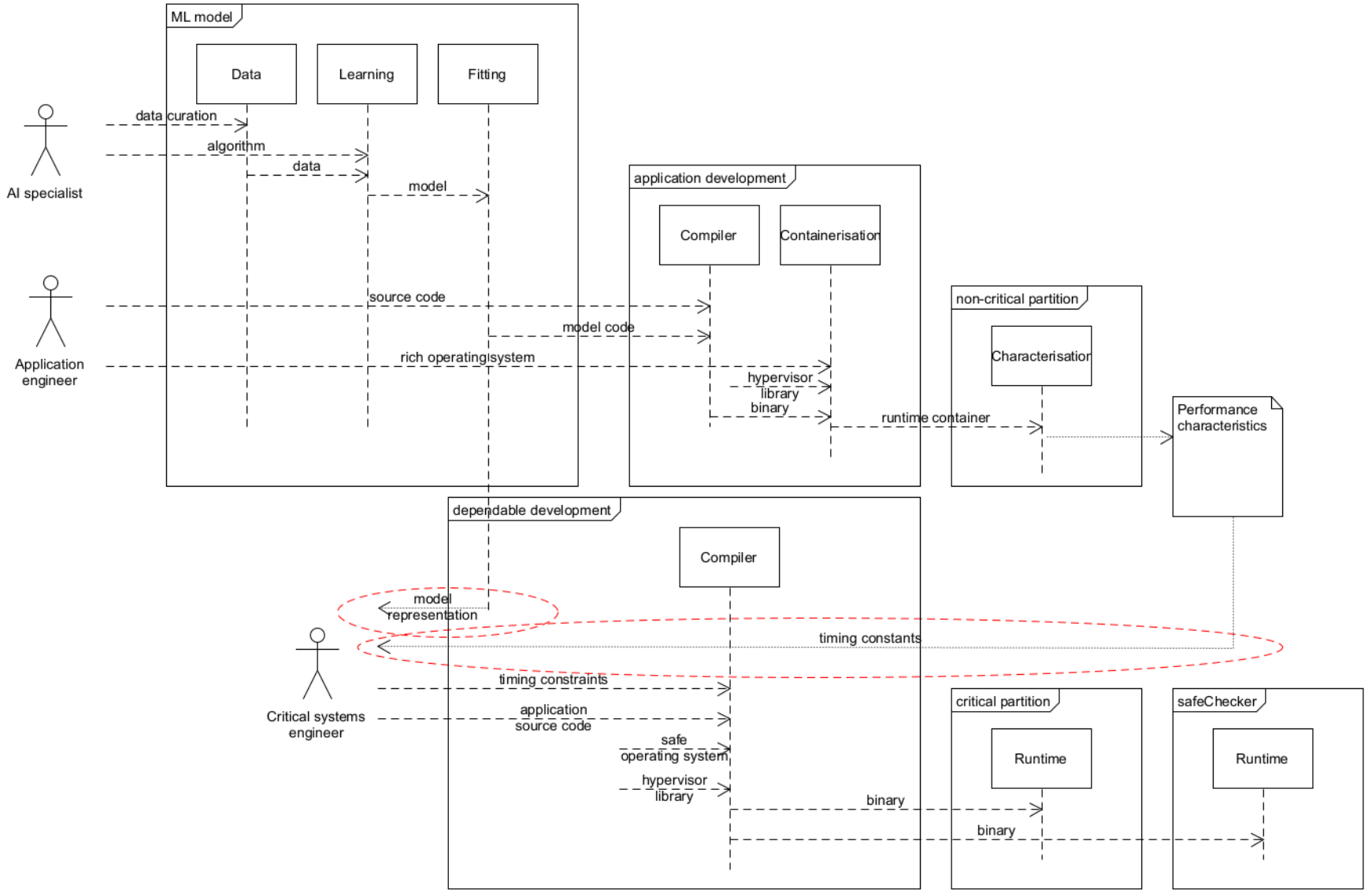
- Two high criticality hypervisor partitions (redundancy)
- Dependability features for FPGA-based AI-Accelerators
- Ramifications for deployment of the AI-application
  - Requires execution triggers received from high-criticality checkers
  - Triggers based on common understanding of AI task
  - Timeouts observed by dependable system need to be determined and configured
- If each high-criticality partition gets dedicated RISC-V core then significant curtailing of computational power



Copyright © 2023

State of the Art: AI focused design and implementation

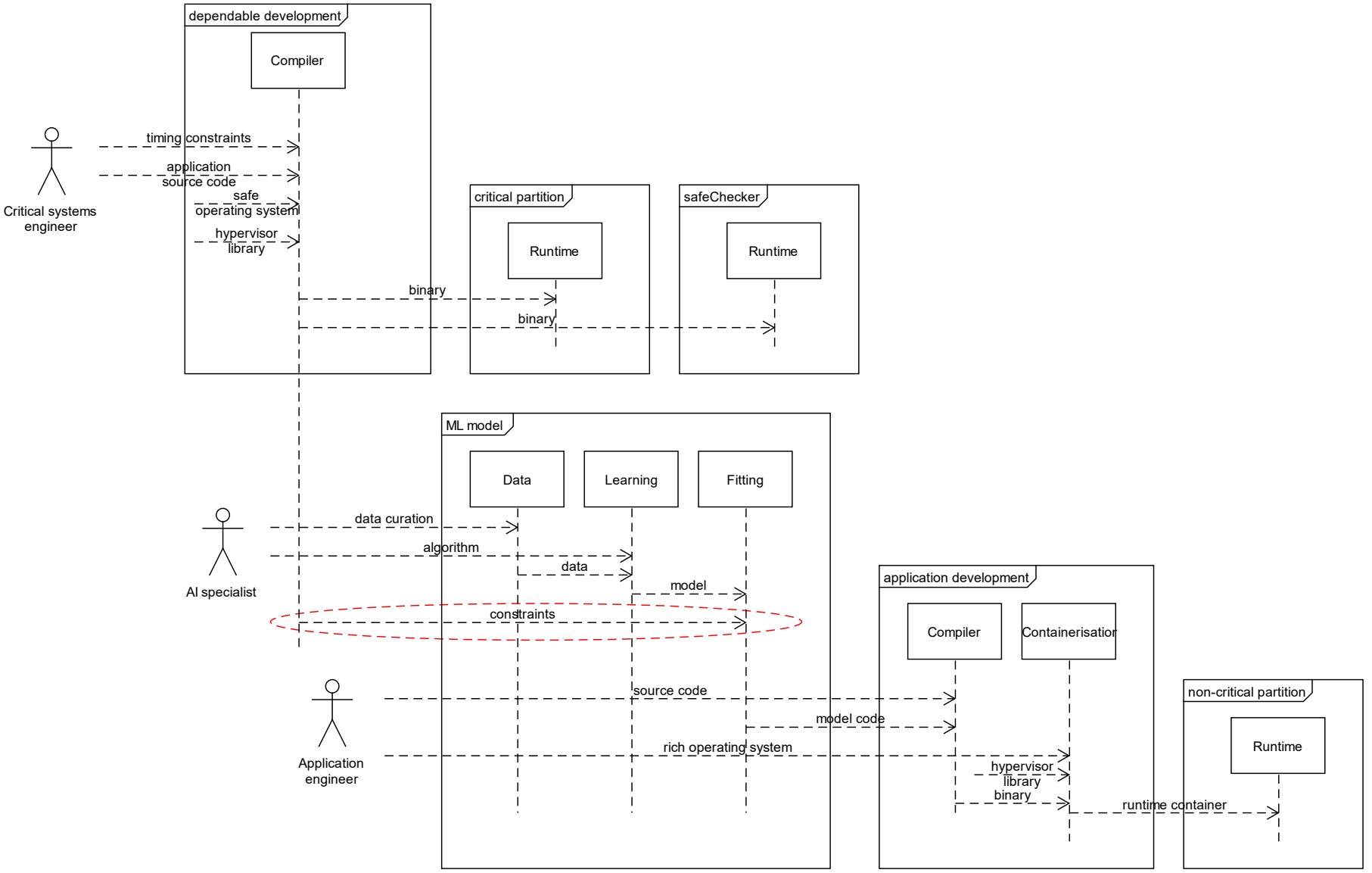
- Non-critical and critical partition runtimes have some input dependencies
- Electronic outputs of non-critical development process may not be directly used as inputs for high criticality development process
  - Dependencies need to be resolved by hand
  - High criticality workflow restricts flexibility of AI application development and deployment



Copyright © 2023

Proposal design and implementation workflow

- By determining dependability constraints a priori – high criticality dependencies become an input for the model generation stage
- Maximum flexibility is assured for AI model generation
- Requires standard definition of constraints for *at an absolute minimum* the model fitting stage
  - Will most likely demand non-functional algorithm modifications



Copyright © 2023

Findings

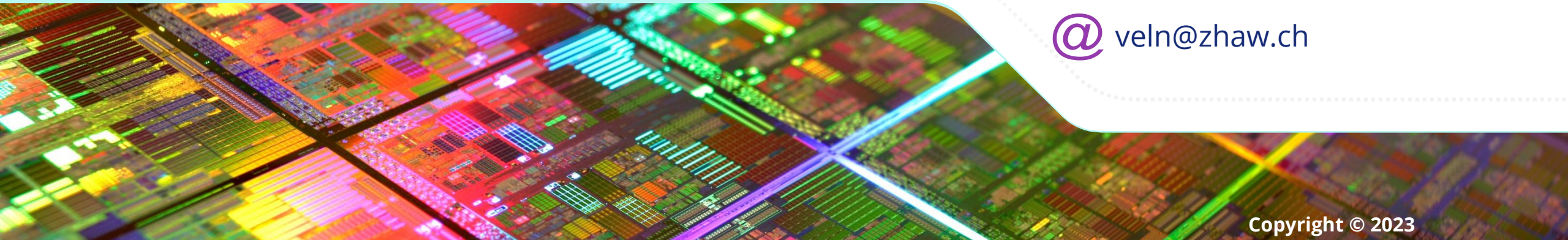
- Critical Systems engineer must define the architecture resources that AI engineers may utilize for any specific level of criticality
  - De-facto an a priori determination of non-functional algorithm architecture
  - Require a format to electronically define the resources made available and any other constraints
- Digital constraints inputs must be integrated into the standard AI model creation process
- Open questions with regards deployment of high criticality SW on statically defined cores and statically defined schedule or on dynamically determined cores and schedules

Copyright © 2023

EDGE AI

Thank You

For your attention



@ veln@zhaw.ch

Copyright © 2023