SAFEXPLAIN
Safe and Explainable
Critical Embedded Systems based on AI

HiPEAC

SAFEXPLAIN Results in Action: *the integrated SW Platform*

Carlo Donzella

exida DEVELOPMENT

Enrico Mezzetti

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

23rd Workshop on Automotive Software & System
29/05/2025 Bergamo (Italy)

Funded by European Union

This project has received funding from the European Union's Horizon Europe programme under grant agreement number 10106959

Some context: where are we coming from?

Until ten years ago, the **automotive domain** was considered one the most mature industrial sectors, with very **consolidated, regulated and high-quality production standards and methodologies**… *but* with a **slow pace of technological innovation** (at least, if compared if other "digitally driven" areas like *consumer electronics, telecom, aerospace, etc*...)

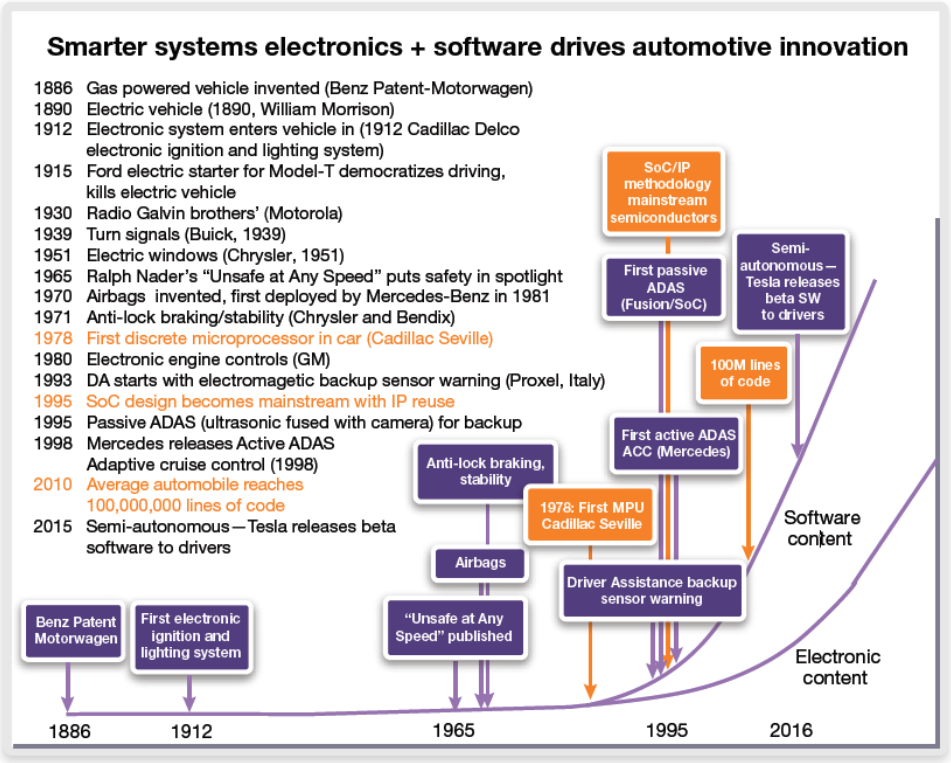
SLOWLY, THEN SUDDENLY, EVERYTHING CHANGED…

- During his Welcome Note at the **Automotive SYS Conference** in Berlin, in **July 2014**, Dr. Ulrich Eichhorn, **VDA Managing Director**, announced that the car of the future would be:

electric, connected, autonomous

- Today this makes us smile but ten years ago it was a ***revolutionary*** statement!

Actually, technology in car has a long - but not linear… - tradition



Source: Synopsys

Remarkable, sometimes stunning, facts:

- First EV: 1**8**90!
- First electronics part in car: 1**9**12!
- First airbag: 1970
- First micro in car: 1978
- First ADAS: 1995

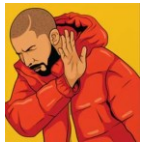
*(extra-automotive bonus from little-known facts - first plane autopilot: 1**9**12!)*

…but… how far are we from the Holy Grail of AI in cars? …that is to say, full autonomy (A.K.A. SAE LEVEL 5)?

There is hardly a more controversial issue: even among most qualified experts, we have two extreme, highly conflictual, irreconcilable positions…

#1 Fully Autonomous Vehicles are obviously already here! [endless list of fancy videos with spectacular driverless trips and of fake/useless technical reports] Let's trash the dumb outdated standards and regulations that are crippling them, let's free the animal spirits of AI-powered turbo capitalism!

#2 Full Autonomy will be obviously never ever achieved! [endless list of methodological/technical unresolved issues and of videos of dreadful driverless accidents] Let's overregulate to create a new AI winter, let's save the world from killer cars before carmaggedon happens!



Enter SAFEXPLAIN… how can we qualify and certify AI-based solutions in EV? Can we provide a basis for a reasonable, balanced, solid position #3?

R&D goals and results

- **AI challenges common practice for FUSA-related software**
  - Failure rates, data used for software design, *[many more issues…]*
- SAFEXPLAIN goals and results:
  - Made **ML/DL components explainable and traceable** by design
    - Already available ML/DL components built with FUSA in mind
  - **Adapted FUSA standards** to allow qualifying/certifying ML/DL software
    - Extensions of processes and methods amenable to manage intrinsic ML/DL characteristics are available
  - **Preserving sufficiently high levels of performance** to meet safety goals
    - Early Core Demo running on developed platform ***on show now at the SAFEXPLAIN booth!***

SAFEXPLAIN results are the basis of a reasonable ***"third"*** position:

- plenty of low-, mid- and high-safety ADAS functions ***can*** (in same cases ***have to***) be implemented with relatively **consolidated ML/DL solutions**;
- (new and updated) **standards and regulations on Quality, Safety and Security of AI** are **needed and welcome** to qualify/certify new solutions before they hit the road.

SAFEXPLAIN
Safe and Explainable
Critical Embedded Systems based on AI

HiPEAC

SAFEXPLAIN Results in Action:
the integrated SW Platform

The SAFEXPLAIN Landscape

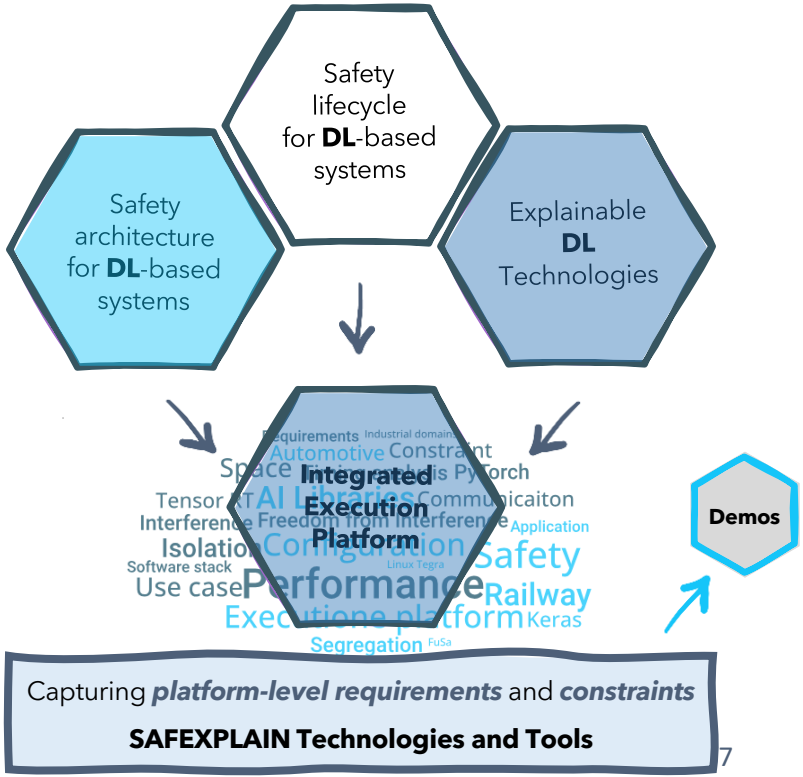
23rd Workshop on Automotive Software & System
29/05/2025 Bergamo (Italy)

Funded by European Union

This project has received funding from the European Union's Horizon Europe programme under grant agreement number 101069595.

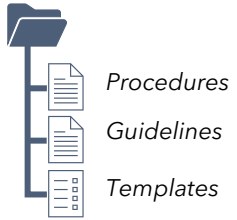
SAFEXPLAIN

- **SAFEXPLAIN distinguishing trait**
- **Holistic approach for Safe & Explainable AI**
  - **Process and procedures**
  - **Design**
  - **Analysis**
  - **Run time**
- **Integrated SW Platform overarching role**
  - All *requirements* stemming from SAFEXPLAIN methodologies must be met
  - All *technological solutions* can be deployed



SAFEXPLAIN AI-FSM

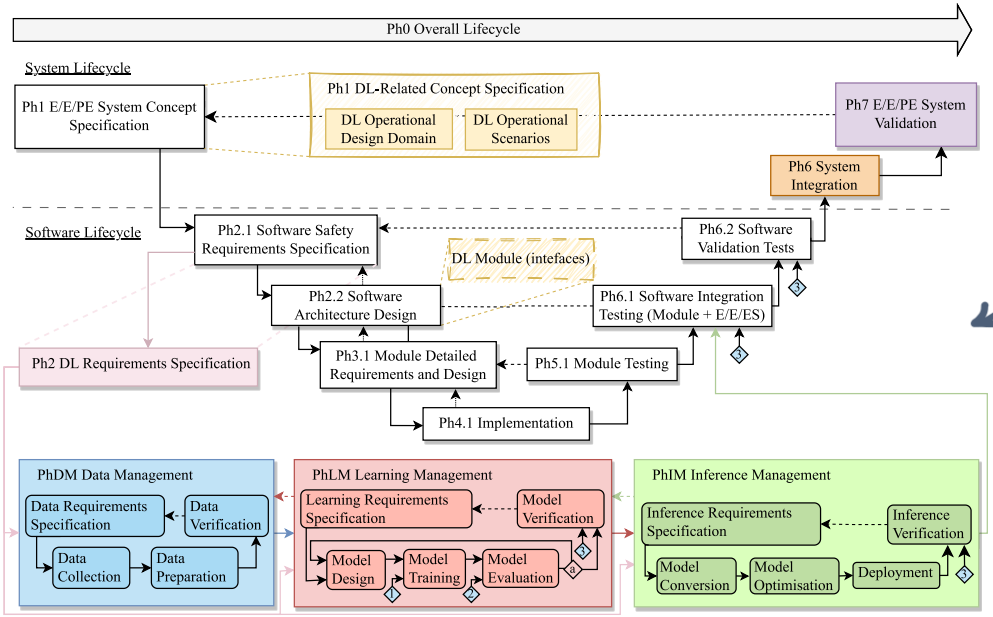
- **Functional Safety Management (FSM)**
  - All essential activities in Functional Safety lifecycle phases (IEC 61508-1)
  - Prevent errors in specification, design, development, manufacturing, and commissioning

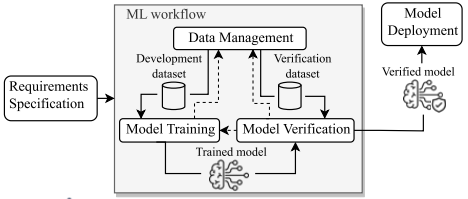


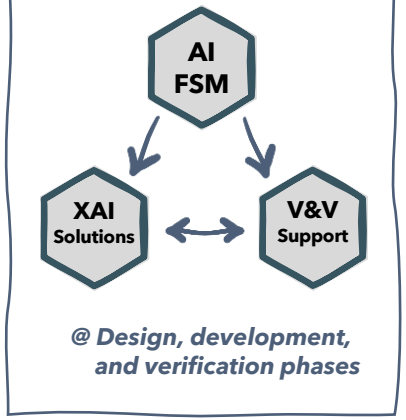
Safety lifecycle for **DL**-based systems

SIL3 Phases

- ✓ Ph0 Overall Life Cycle
- ✓ Ph1 System Concept Spec.
- ✓ Ph2 System Architecture Spec.
- ✓ Ph3 Module Detailed Design
- ✓ Ph4 Implementation
- ✓ Ph5 Module Testing
- ✓ Ph6 Integration Testing
- ✓ Ph7 Validation Testing







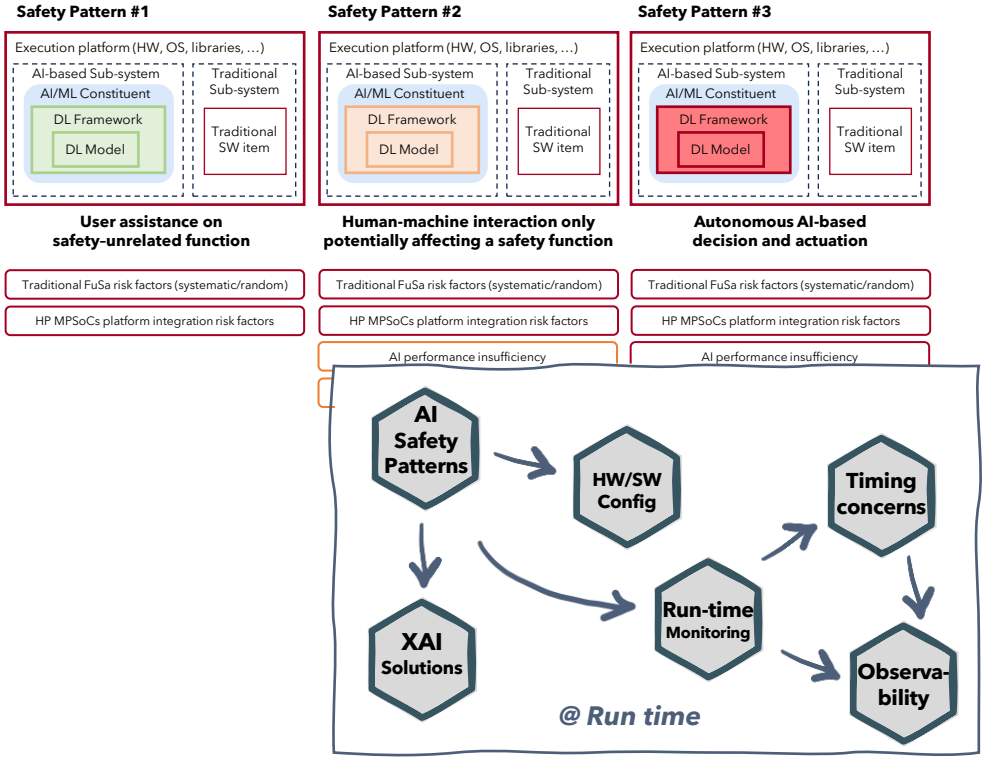
SAFEXPLAIN

SAFEXPLAIN FuSa Architecture and Patterns

Safety architecture for **DL**-based systems

- **Traditional Functional Safety (FuSa)**
  - *Capture random and systematic faults*
  - *Master HW / SW platform complexity*
    - *Segregation, interference, mixed-criticality approaches, use of resources, etc.*
- **AI/DL specific traits**
  - *Capture DL model insufficiencies*
  - *Support DL explainability solutions*
- **Architectural archetypes**
  - *Diverse redundancy*
  - *Supervision block*
  - *Layered diagnostic and monitoring*

- **Safety patterns**
  - Incremental addressing of risk factors on AI-based Sub-system



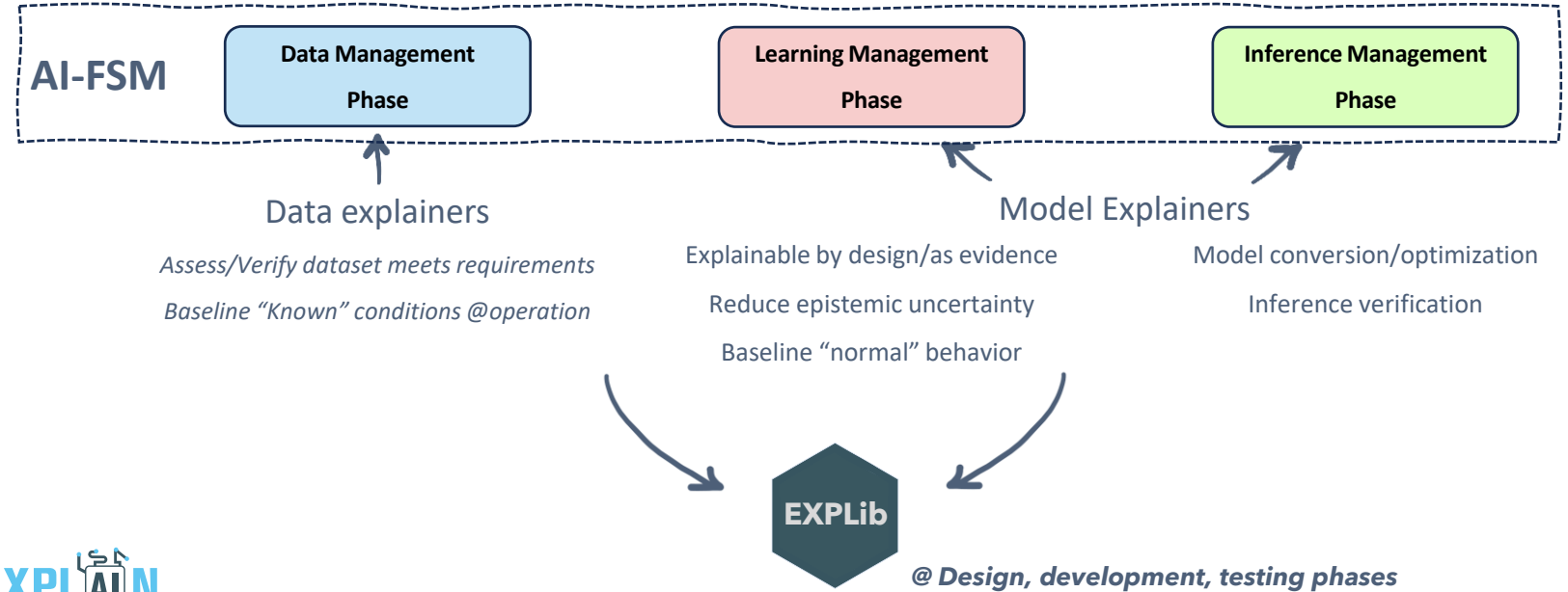
SAFEXPLAIN

SAFEXPLAIN XAI

Explainable **DL** Technologies

- **Explainable Artificial Intelligence (XAI)**

*Processes, algorithms, and methods that allows humans to **understand** and **trust** the results and output generated by machine learning algorithms*



SAFEXPLAIN XAI

Explainable **DL** Technologies

- **Explainable Artificial Intelligence (XAI)**

*Processes, algorithms, and methods that allows humans to **understand** and **trust** the results and output generated by machine learning algorithms*

FuSa Architecture

Model supervisor

Decision function

EXPLib

- *Anomaly detection (input, model, output)*
- *Surrogate model for consistency check*
- *Uncertainty estimation*

IN: Model prediction + supervisor score

OUT: Ensemble and trustworthiness score

DLLib

@ Run time

SAFEXPLAIN

SAFEXPLAIN
Safe and Explainable
Critical Embedded Systems based on AI

HiPEAC

SAFEXPLAIN Results in Action:
the integrated SW Platform

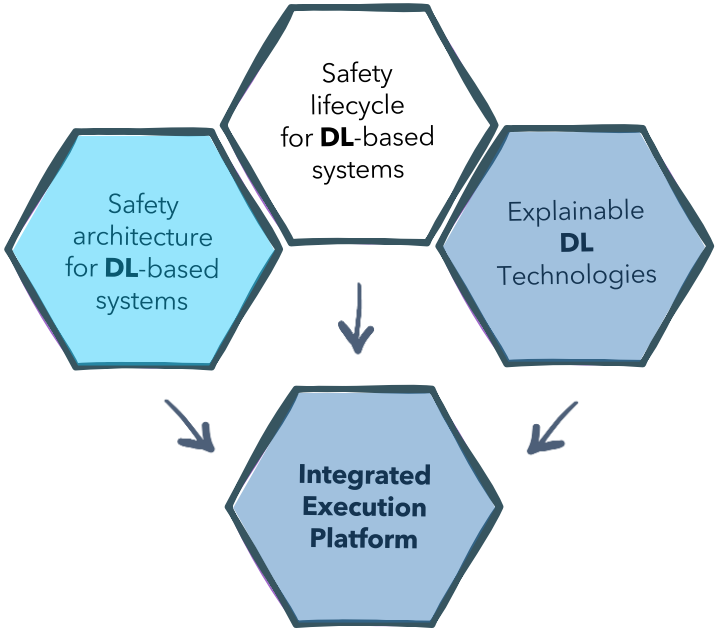
SAFEXPLAIN Integrated Platform

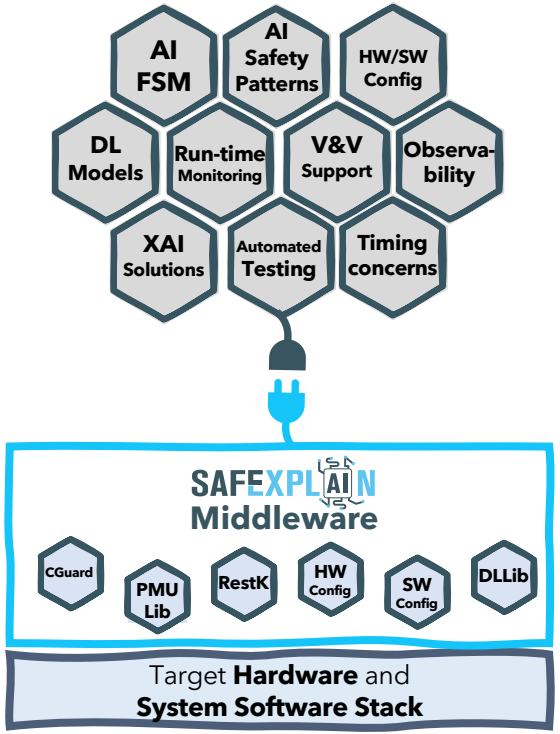
23rd Workshop on Automotive Software & System
29/05/2025 Bergamo (Italy)

Funded by European Union

This project has received funding from the European Union's Horizon Europe programme under grant agreement number 101069595.

Where things are supposed to come together…





Integrated Execution Platform

SAFEXPLAIN HW/System SW Platform

Target **Hardware** and **System Software Stack**

Ubuntu (Linux Tegra)

Hardware

NVIDIA Jetson AGX Orin (32GB Dev Kit)

Computing requirements of complex AI-based systems

200 TOPS of AI performance for autonomous systems

NVIDIA Ampere GPU

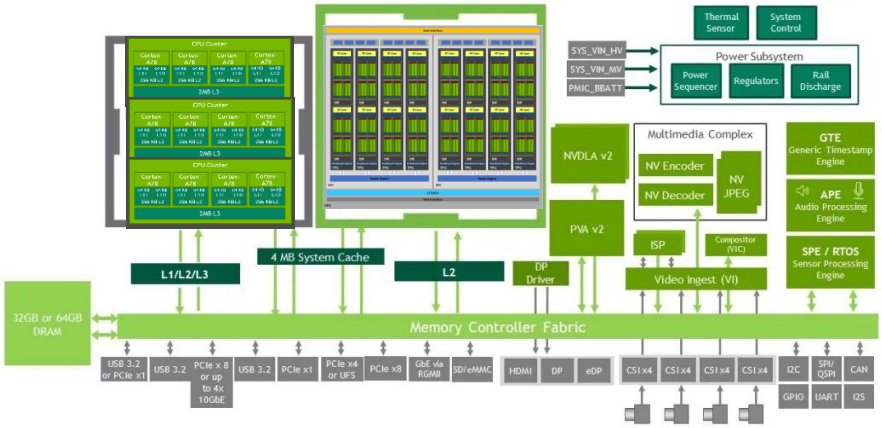
Arm® Cortex®-A78AE CPU

Next-gen accelerators for DL and Vision (NVDLA, PVA)

Video encoder and decoder

High speed I/O, 204 GB/s of memory bandwidth

32GB of DRAM + 2TB NVMe



NVIDIA reference SW stack

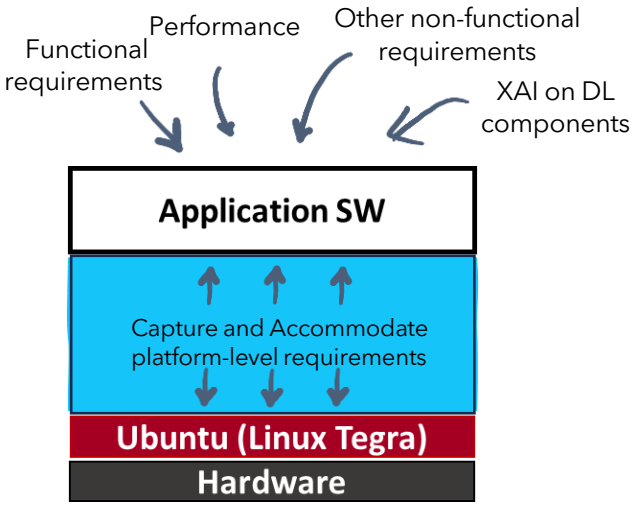
Jetson Linux 36.3 (Ubuntu 22.04)

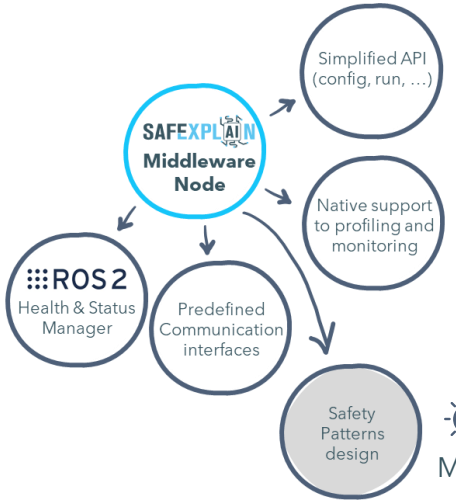
Linux Tegra 5.15

JetPack 6.0.1 SDK

SAFEXPLAIN Middleware concept

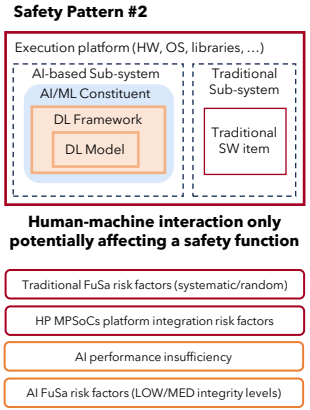
- **Building on ROS2 concepts and support**
  - Widespread adoption
  - Modular design and scalability (pub-sub semantics)
  - Native support for FuSa and V&V
- **Pivotal concept** in SAFEXPLAIN platform
  - Set of node archetypes as specializations (wrapper)
    - Communication patterns – data flows
    - Bespoke monitoring mechanism
    - Simplified semantics specification
  - Compliance to Safety Patterns *by design*
    - Mandatory software nodes/components
    - User-defined nodes for application semantics
    - Multi-layered monitoring and diagnostic
  - Support for deployment configurations
    - Statically pre-defined setups and configurable options
  - Test integration and V&V support
  - Seamless integration of all technologies/tools

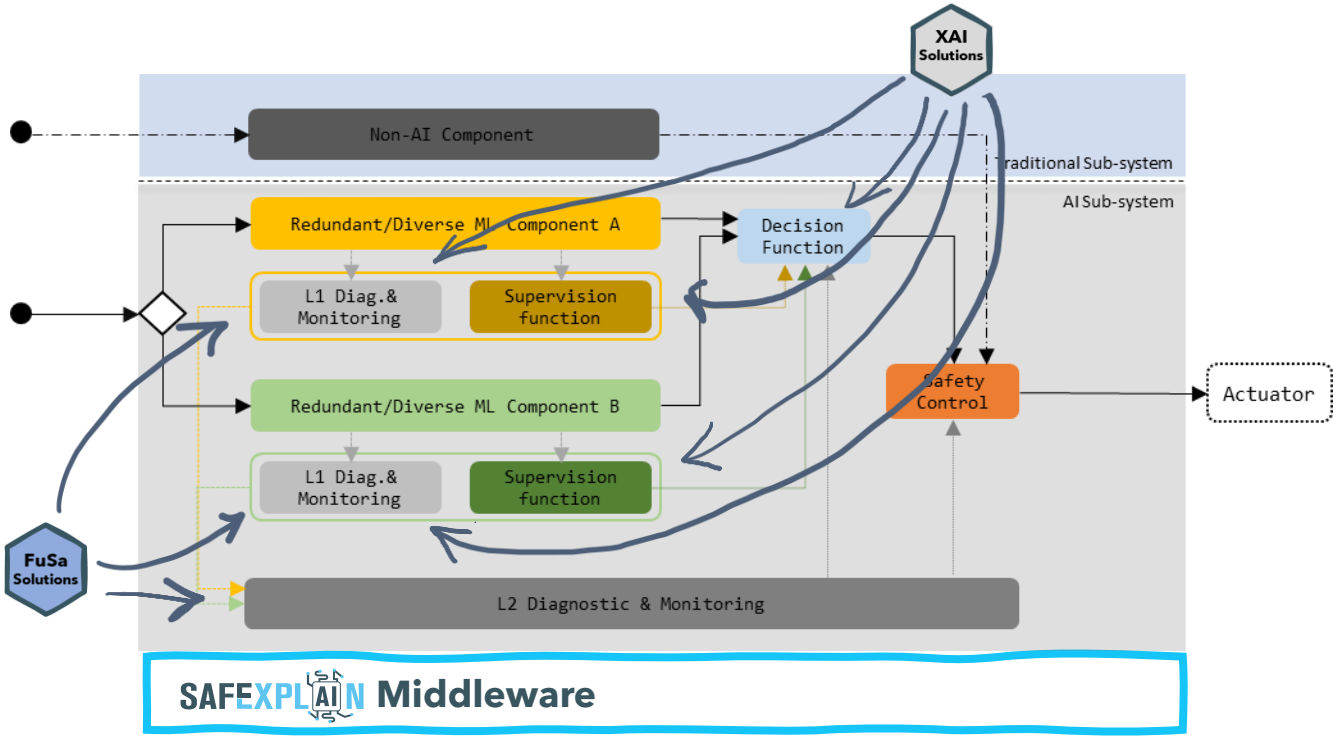




Safety pattern example

- SP2 - AI/ML constituent may partially affect the decision process

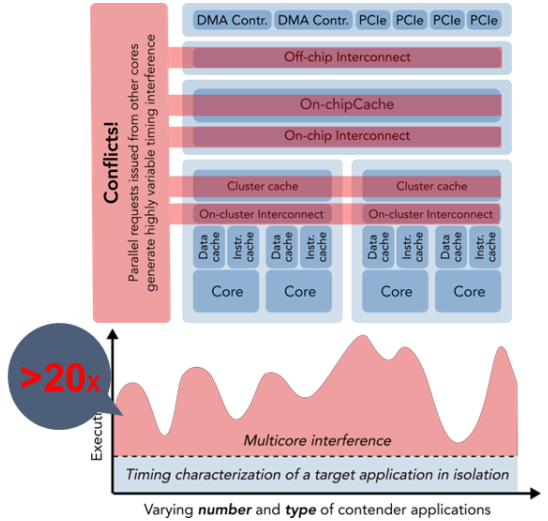


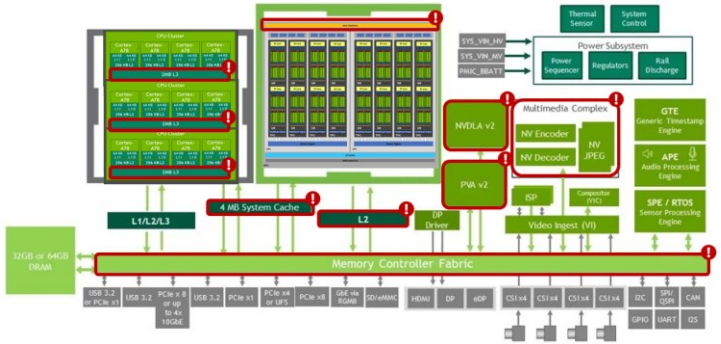


Plug-in 'functional' code
- ✓ *ML nodes*
- ✓ *Supervisor*
- ✓ *Decision function*
- ✓ *Control logic*

Platform support – HW and SW configuration

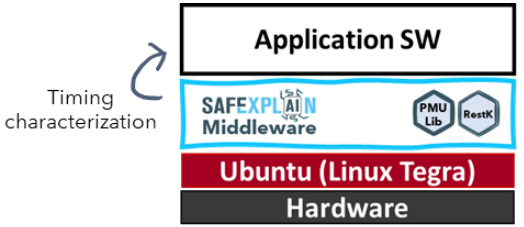
- **COTS MPSoCs** build on **massive HW resource sharing**
  - **Unprecedented performance** meeting AI requirements
  - **Contention** on multiple simultaneous accesses
    - A concern for timing V&V practice
      - ***Freedom from interference*** – ISO 26262
      - ***Interference channels*** – CAST 32A and A(M)C 20-193
- **Focus on configuration and deployment options**
  - Promote **segregation** **AND** meet **performance** requirements
- **Strategical objectives**
  - Analysis and classification of **sources of timing interference**
  - Identification of **HW/SW mechanisms** for interference control
  - Identification of **deployment configurations**
  - Support the **instantiation of the identified setups**

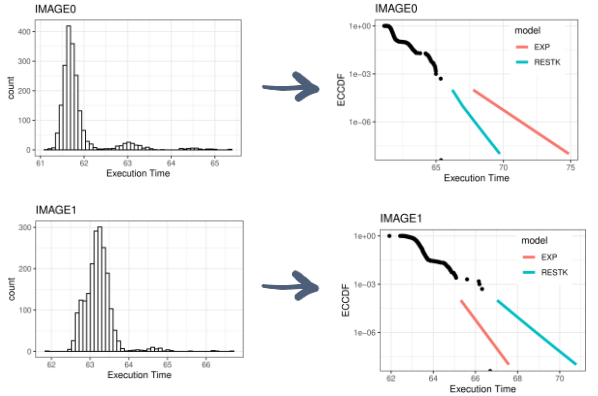


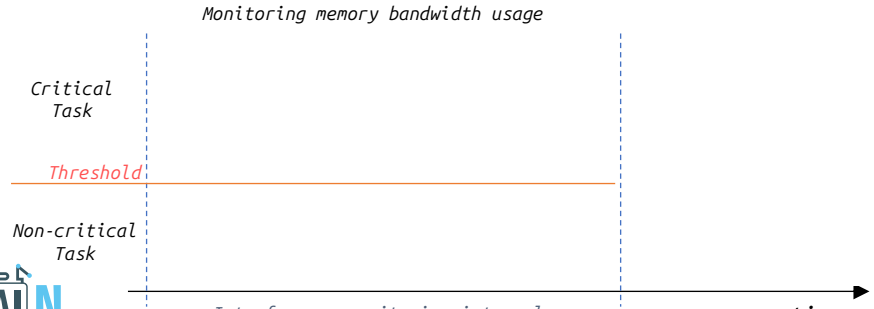


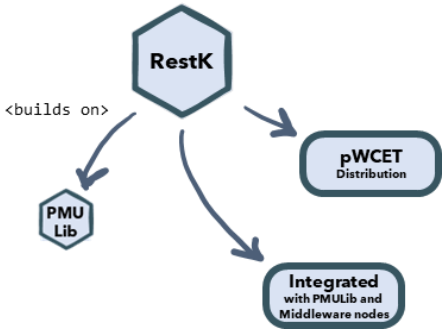
Timing characterization method for AI-based solutions

- **Complex AI-based SW on advanced heterogeneous MPSoCs**
- **Probabilistic timing analysis techniques**
  - Use of *statistical analysis tools* to produce WCET estimates.
  - Restricted k (**RestK**)
    - Based on *Markov's inequalities* adapted to mixture distributions
- **Capture and control *residual* interference**
  - After solutions deployed as part of Safety Patterns
  - Full *freedom from interference* is hard to achieve
  - Non-critical tasks are prevented from generating unwanted timing interference on critical tasks (inferred from correlated HW events)
  - Tasks exceeding the *interference thresholds* are throttled
    - Until next period or critical tasks terminate



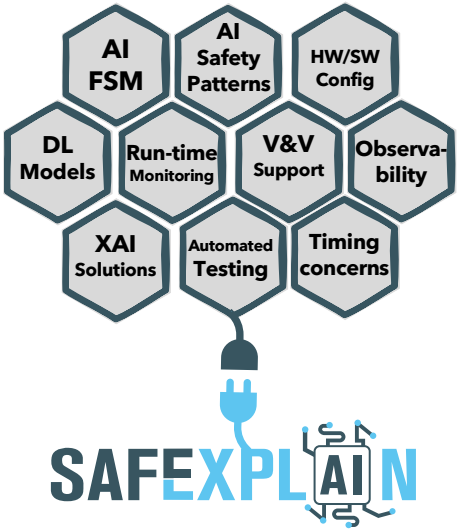


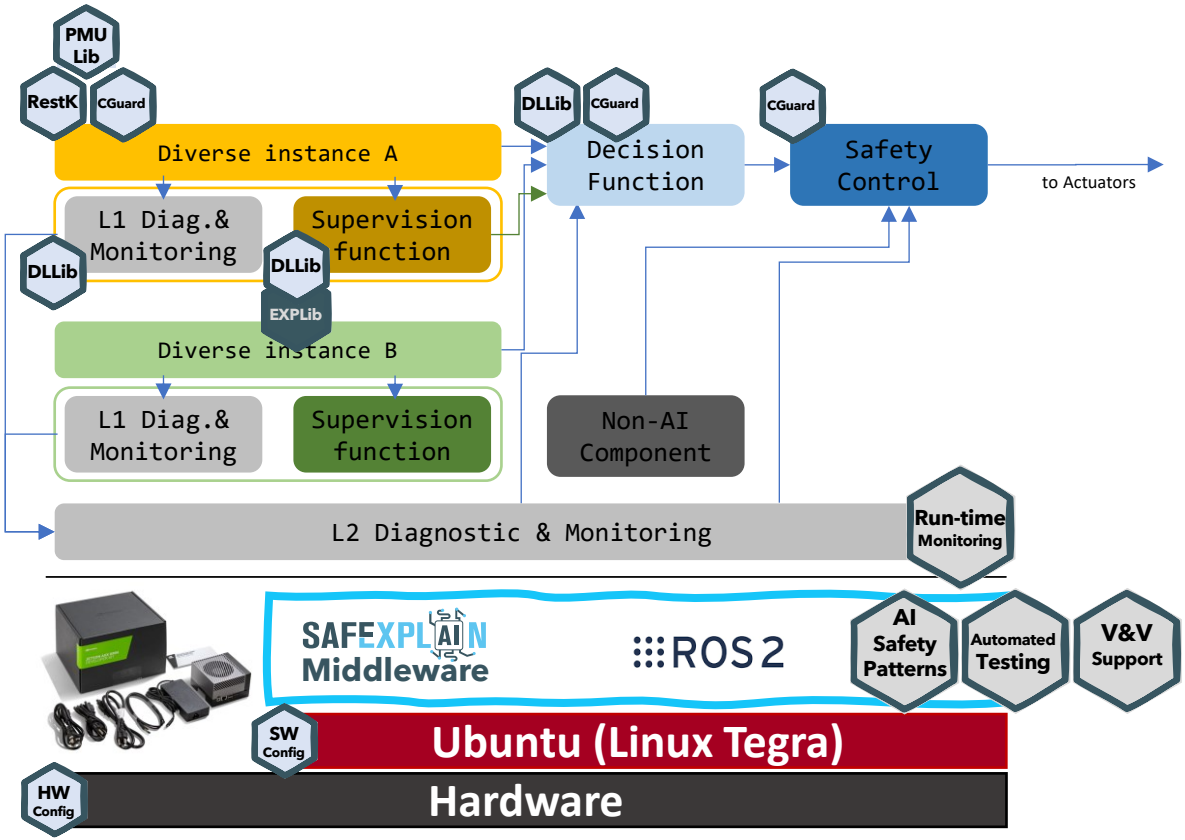




SAFEXPLAIN

SAFEXPLAIN Platform Recap

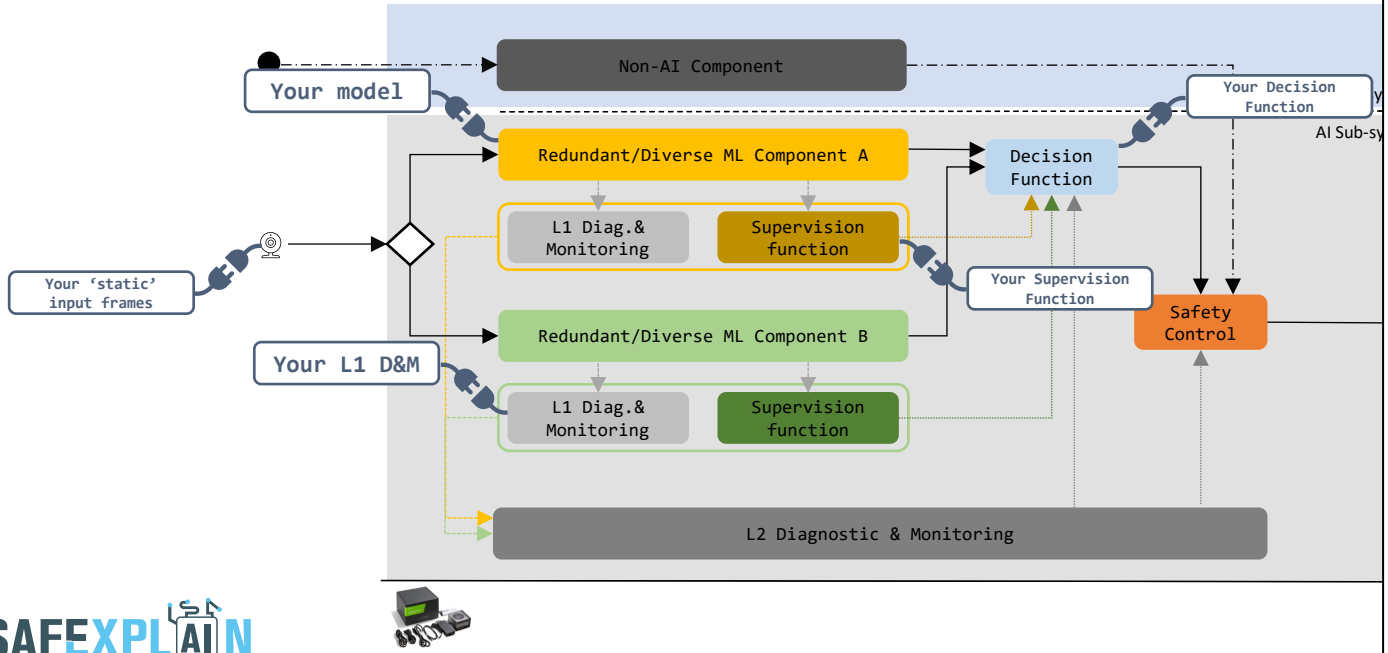


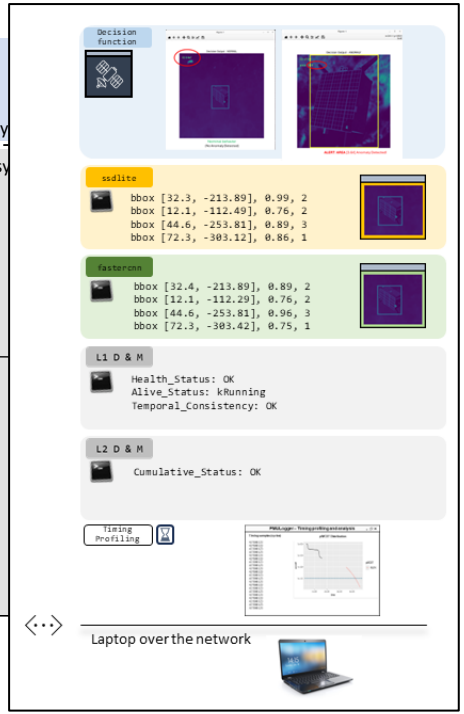


SAFEXPLAIN CORE Demo

- **CORE Demo as Adaptable Cross-domain demonstrator**
  - Small-scale, simplified, yet representative open demonstrator
    - SAFEXPLAIN technologies and tools and platform's key features
  - Replaceable building blocks to showcase specific features and scenarios

Technology demonstrators
SAFEXPLAIN Core Demo(s)
Open-source elements or fully documented





SAFEXPLAIN

OCT 2022 – SEP 2025

SAFEXPLAIN

Safe and Explainable
Critical Embedded Systems based on AI

BARCELONA SUPERCOMPUTING CENTER (BSC)
https://www.bsc.es/

IKERLAN, S. Coop (IKR)
https://www.ikerlan.es/

AIKO SRL (AIKO)
https://www.aikospace.com/

RISE RESEARCH INSTITUTES OF SWEDEN AB (RISE)
https://www.ri.se/

NAVINFO EUROPE BV (NAV)
https://www.navinfo.eu/

EXIDA DEVELOPMENT SRL (EXI)
https://www.exida-eu.com/

- It is instrumental in closing the gap between **AI solutions** and **safety culture**
- It provides **guidelines, processes, libraries and software tools**
- It enables **AI adoption in Critical Embedded Systems**
- Now it is close to completion and has just released the SW **Core Demo**, teasing for the **Full Demos** in Sept 2025
- It is looking for **early users/adopters** in **industries** and **agencies**

SAFEXPLAIN
Safe and Explainable
Critical Embedded Systems based on AI

Want to know more about SAFEXPLAIN?

Learn about the upcoming activities

Making certifiable AI a reality for critical systems: CORE DEMO

Discover how SAFEXPLAIN technology can accommodate scenarios with critical functionalities in three selected toy examples from the automotive, rail and space domains.

TRUSTWORTHY AI IN SAFETY-CRITICAL SYSTEMS

Overcoming adoption barriers

Join us to explore safe, reliable AI in fields like automotive, aerospace, rail, and robotics. Learn through demos, tech sessions, and discussions on making AI robust, explainable, and standards-compliant.

Thanks

SAFEXPLAIN

Safe and Explainable
Critical Embedded Systems based on AI

Follow us on social media:

www.safexplain.eu

Funded by the European Union

This project has received funding from the European Union's Horizon Europe programme under grant agreement number 101069595.