SAFEXPLAIN Unveils Core Demo: A Step Further Toward Safe and Explainable AI in Critical Systems

Barcelona, 03 July 2025

The SAFEXPLAIN project has just publicly unveiled its Core Demo, offering a concrete look at how its open software platform can bring safe, explainable and certifiable AI to critical domains like space, automotive and rail.

Showcased in a recent webinar led by partners from the Barcelona Supercomputing Center and exida development, the Core Demo walkthrough and corresponding slides are now publicly available, giving the broader community a first look at the project's holistic, end-to-end approach to AI safety and explainability.
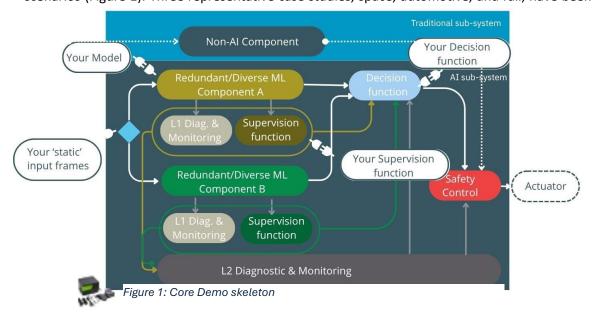
Addressing AI's Clash with Functional Safety

As AI takes on increasingly complex roles in safety-critical applications, from autonomous vehicles to satellite systems, the need for trustworthy AI is quickly growing. However, today's AI mostly remains in a "black box" that is difficult to verify, lacks traceability and is at odds with the rigorous certification process demanded by sectors like transportation and healthcare.

SAFEXPLAIN has contributed to bridging this gap. The project's Core Demo is a small-scale, modular demonstrator that highlights the platform's key technologies and showcases how AI/ML components can be safely integrated into critical systems. The SAFEXPLAIN approach complements emerging standards like SOTIF by enabling runtime monitoring and explainability-by-design. The Core Demo focuses on an illustrative *safety pattern* scenario- where AI influences, but does not solely determine, decision-making- to illustrate how SAFEXPLAIN technology makes AI-based safety-critical systems safe by construction, without compromising performance.

A Modular and Representative Demo for Space, Automotive and Rail domains

The Core Demo is built on a flexible skeleton of replaceable building blocks for Inference, Supervision, or Diagnostic components, among others, that allow it to be adapted to different scenarios (Figure 1). Three representative case studies, space, automotive, and rail, have been

Figure 1: Core Demo skeleton

instantiated within the framework to highlight its applicability across domains. The technical specification describing this skeleton and it applicability can be found [here](#).

The webinar showed the Core Demo instantiated to the space case, where the AI component is responsible for the docking maneuvers between two spacecrafts, including target (satellite) identification and pose estimation. In this case, AI supports, but does not dominate system behavior. It partially affects the decision process and, ultimately, the operation of the system. The SAFEXPLAIN execution platform actively monitors the system for potential drifts and sensor faults. In the automotive case, the Core Demo is envisioned to support assisted braking systems, while in the rail case it aids in anomaly detection in signaling or onboard diagnostics.

Figure 2: Excerpts of the Core Demo in Action in space and automotive cases, presented by Enrico Mezzetti

What's Next: From Core Demo to Full-Scale Use Cases

- The Core Demo webinar is now available online. SAFEXPLAIN partners invite early users and interested stakeholders to explore its capabilities in their own environments by contacting safexplainproject@bsc.es.
- The Core Demo also offers a preview of what is ahead: **fully developed end-to-end demonstrators in space, automotive and rail cases**, complete with Operational Design Domain (ODD)- based scenarios and test suites.
- The final demonstrators will be officially released in September 2025, during the "[Trustworthy AI In Safety-Critical Systems: Overcoming adoption barriers](#)" event. Registration is open until 8 September 2025, or until the venue capacity has been reached.

For more information and to access the webinar recording with the Core Demo walkthrough and presentation slides, visit the [website](#) or contact safexplainproject@bsc.es.

About SAFEXPLAIN

The [SAFEXPLAIN](#) (Safe and Explainable Critical Embedded Systems based on AI) is a HORIZON Research and Innovation Action financed under grant agreement 101069595. The project began on 1 October 2022 and will end in September 2025. The project is formed by an inter-disciplinary consortium of six partners coordinated by the Barcelona Supercomputing Center (BSC). The consortium is composed of three research centres, RISE (Sweden; AI expertise), IKERLAN (Spain; FUSA and railway expertise) and BSC (Spain; platform expertise) and three CAIS industries, NAVINFO (Netherlands; automotive), AIKO (Italy; space), and EXIDA DEV (Italy; FUSA and automotive).