ARTIFICIAL INTELLIGENCE

PNRR/SERICS - Space-Aware Safety and Security of AI

Funded By	Dipartimento DAUIN MINISTERO DELL'UNIVERSITA' E DELLA RICERCA [P.iva/CF:97429780584] Politecnico di TORINO [P.iva/CF:00518460019]
Supervisor	DI CARLO STEFANO - stefano.dicarlo@polito.it
Contact	SAVINO ALESSANDRO - alessandro.savino@polito.it DI CARLO STEFANO - stefano.dicarlo@polito.it
Context of the research activity	The increasing deployment of Artificial Intelligence (AI) in space applications introduces both significant opportunities and critical challenges. This PhD proposal aims to address the safety and security concerns associated with AI systems operating in the space environment. By leveraging advanced machine learning techniques, we seek to enhance the robustness and reliability of AI systems, ensuring their resilience against both environmental hazards and cyber threats with focus on RISC-V based systems. The approach includes developing comprehensive security frameworks and fail-safe mechanisms to safeguard these AI systems, facilitating their safe operation and enhancing mission success rates. Use-case applications from aerospace industry will be considered. "Progetto finanziato nell'ambito del PNRR - M4C2, Investimento 1.3 - Avviso n. 341 del 15/03/2022 - PE0000014 Security and Rights in the CyberSpace (SERICS) - CUP E13C22001850001"
	The proposed research aims to address the critical challenges of safety and security for Artificial Intelligence (AI) systems deployed in space applications. The research will encompass several key areas, including the development of robust AI algorithms, the creation of comprehensive security frameworks, and the attestation of the quality of solutions in realistic use case applications. The overarching goal is to enhance the reliability, resilience, and security of AI systems operating in the unique and harsh conditions of space. The initial phase of the research will involve a thorough review of existing literature on AI safety and security, specifically focusing on applications in space. This will include an analysis of current AI algorithms used in space missions, their vulnerabilities, and the existing security measures in place. The goal is to identify the gaps and limitations in current technologies and methodologies, providing a foundation for the subsequent research phases. One of the core activities will be the development of robust AI algorithms that

Objectives	can operate reliably in the challenging space environment. This will involve: • Error Detection and Correction: Designing AI systems capable of detecting and correcting their errors autonomously. This is crucial for maintaining operational integrity, given the difficulty of human intervention in space. • Fault Tolerance: Developing algorithms and architectures with built-in fault tolerance to handle unexpected conditions and anomalies. • Adaptive Learning: Creating AI systems that can adapt to new and unforeseen circumstances without human intervention, enhancing their long-term reliability. Such fault-tolerance do not include cyber-attacks protections, for which developing robust security frameworks is essential. This research will focus on: • Threat Modeling: Identifying potential threats and attack vectors specific to space-deployed AI systems. • Secure Communication Protocols: Developing secure communication protocols to protect data transmitted between space AI systems and ground control. • Intrusion Detection Systems: Implementing advanced intrusion detection systems that can identify and respond to security breaches in real-time. • Cryptographic Techniques: Utilizing advanced cryptographic techniques to secure data at rest and in transit. As the final goal is to support with design exploration optimization techniques that will allow the definition of safety and security constraints, the further step will aim at designing specific training strategies to support the application of fault-tolerance and security protection during the model generation. Such strategies might complement architectural solutions with data-aware ones. Simulations and modeling will be used extensively to test the developed AI algorithms and security measures under various conditions, including fault injection and cyber-attacks simulations. To ensure the practical applicability of the developed solutions, experimental validation will be conducted through collaborations with space agencies and industry partners. The findings from this research will be disseminated through various channels to maximize impact: • Journal Publications: Publishing research papers in high-impact journals, such as IEEE Transactions on Computers, IEEE Transactions on Neural Networks and Learning Systems • Workshops and Conferences: targeting high-level conferences such as DATE, DAC, DFT, ESWEEK, MICRO, etc., and Space dedicated events.

Skills and competencies for the development of the activity	Hardware accelerator design, FPGA prototyping, with an emphasis on reconfigurable architectures and programming skills in Python, C.