ARTIFICIAL INTELLIGENCE

DM 629 PNRR - AI-ready institutional platform for the collaborative development and valorization of FAIR-by-design software

Funded By	MINISTERO DELL'UNIVERSITA' E DELLA RICERCA [P.iva/CF:97429780584] UNIVERSITA' DI PISA [P.iva/CF:00286820501] Politecnico di TORINO [P.iva/CF:00518460019]
Supervisor	CAPPELLUTI FEDERICA - federica.cappelluti@polito.it
Contact	TORCHIANO MARCO - marco.torchiano@polito.it CAPPELLUTI FEDERICA - federica.cappelluti@polito.it VETRO' ANTONIO - antonio.vetro@polito.it
Context of the research activity	This project aims at experimenting and validating advanced Artificial Intelligence services and tools for the development, release and sharing of software in line with Open Science best practices and the FAIR principles (findability, accessibility, interoperability, reuse). The AI-based tools will be built on top of the GitLab@PoliTo platform, in collaboration with the Center Study on Open Science, to gather faculty and researcher needs. Progetto finanziato dal PNRR a valere sul DM 629/2024 "Generica di ricerca PNRR" - CUP: E14D24002310006
	Software is fundamental to scientific research, from STEM disciplines to the humanities and social sciences. and, along with publications and data, is recognized as one of the enabling pillars for Open Science. The peculiarities of software compared to other digital objects, such as its inherent evolutionary nature, composite architecture, and machine language coding for the purpose of executability require appropriate management tools to meet specific FAIR principles (findability, accessibility, interoperability, reuse) and the FAIR-by-design paradigm. Reproducibility of software is a key element of high-quality research and fosters collaboration; versioning facilitates technology transfer. It is therefore important for research institutions to support, monitor and enhance the production of software both in the context of funded and open research as well as to ensure its long-term preservation. In this context, the Politecnico di Torino has launched a pilot project to develop the FAIR software platform GitLab@PoliTo, which rests on an institutional instance of GitLab, an environment for collaborative

Objectives	development and version control. The goal of the research is to test and integrate into the GitLab@PoliTo platform advanced artificial intelligence (AI) services with the dual purpose of valorizing the software developed by the academic community and providing researchers with advanced tools for its development and release according to Open Science principles and best practices. The research will leverage GitLab's AI architecture, which integrates multiple components that facilitate the development of AI services on its platform. Advanced AI services may cover: automatic production of documentation, testing and synthetic data for testing with ready-to-use deep learning models; intelligent assistance in linking an artifact to Open Science platforms (e.g., EOSC, Zenodo); support for exploitation through patents (e.g., automatic classification); and the creation of pipelines for building, validating and containerizing AI models (e.g., MLops), to make GitLab@PoliTo an "AI-ready" platform and help address the so-called "reproducibility crisis." Additional services may be defined by means of surveys and focus groups in the University and within international networks on Open Science in which the Politecnico di Torino participates (e.g., EOSC-A), in collaboration with the Center Study on Open Science, to gather the needs and ideas of faculty and researchers in a timely and inclusive manner. The services developed as part of the PhD thesis may also contribute to the GitLab community's ongoing development of the AI Gateway, a new standalone service that will provide access to AI capabilities to all GitLab users, facilitating the integration of external AI services, model deployment, and real-time inference in GitLab.Finally, the research activity will include an analysis of the non-functional requirements that need to be met for the platform to achieve adequate levels of quality to provide services to all researchers at Politecnico. The following points identify the track of activities of the work. It will be subject to constant revision. 1. Literature Review and Background Research on the following aspects: • Open Science and the FAIR principles. • Software Engineering for AI (SE4AI) latest advancements, and vice versa (AI4SE). • Reproducibility crisis in AI research and existing technical solutions. 2. Analysis of current GitLab's AI Architecture and future developments • Analyze GitLab's existing architecture and future developments (AI Gateway), focusing on CI/CD pipelines, Kubernetes integration, and data/package management. • Identify the opportunities for integrating AI services into GitLab@PoliTo. 3. Requirement Gathering • Produce requirements for the identified services and matching with the components of the GitLab AI architecture. • Collaborate with the Center Study on Open Science to gather requirements from faculty and researchers. 4. Development of AI Services • Implement selected advanced AI services into the GitLab@PoliTo platform. 5. Evaluation and Optimization • Create and run test cases to validate the functionality, usability and performance of the AI services, also in relation to the state of the art. • Optimize/improve the services based on the evaluation results. 6. Guidelines for long-term Preservation and Reproducibility • Develop strategies for the long-term preservation of software and reproducibility mechanisms at Polito. • Conduct a thorough analysis of non-functional requirements for providing the services to all researchers at Politecnico.

	Dissemination of results and contribution to the GitLab community will be cross-cutting activities.

Skills and competencies for the development of the activity	The candidate should have: • Very good knowledge of AI techniques. • Good knowledge of statistical methods and tools for analyzing experimental data. • Basic knowledge on software testing concepts, techniques, and methodologies. • Strong programming skills. • Basic knowledge of Open Science and FAIR principle, also in relation to software. • Research aptitude and curiosity to cross disciplinary boundaries. The candidate should also possess good communication and presentation skills.