ARTIFICIAL INTELLIGENCE

MUR DM 118 - Automated Assessment and Improvement of the Quality of Training Data for Machine Learning Models

Funded By	Università degli Studi del MOLISE [P.iva/CF:00745150706] MINISTERO DELL'UNIVERSITA' E DELLA RICERCA [P.iva/CF:97429780584]
Supervisor	DI CARLO STEFANO - stefano.dicarlo@polito.it
Contact	Prof. Rocco Oliveto, Università degli Studi del Molise, rocco.oliveto@unimol.it
Context of the research activity	Quality of Training Data for Machine Learning Models. Progetto finanziato nell'ambito del PNRR – DM 118/2023 - CUP E14D23001820006
Objectives	This research aims at defining novel approaches for automatically assessing the quality of data used to train Machine Learning (ML) models. First, a comprehensive catalogue of data-related problems will be defined for several properties (e.g., effectiveness and fairness). Then, automated approaches for removing such problems and, thus, improving data quality will be devised. Finally, it will be empirically verified to what extent removing such problems allows for improving the related property.
Skills and competencies for the development of the activity	The candidate should have excellent programming skills (possibly in Python). He/she should know how to train and test a Machine Learning model, including how to compute commonly used metrics for evaluating a model. In particular, the candidate should know how to train and test Deep Neural Networks and, specifically, Large Language Models.