Politecnico di Torino

COMPUTER AND CONTROL ENGINEERING

DAUIN - A Data Redistributive and Cooperative Model for Fostering Circular Social Impact of Trustworthy AI

Funded By	Dipartimento DAUIN
Supervisor	CERQUITELLI TANIA - tania.cerquitelli@polito.it
Contact	
Context of the research activity	This research activity develops a data policy engine to give individuals control over their data while enabling cooperative AI services for broader social benefit (e.g., health insurance for people with illnesses). Using LLM-based synthetic data generation and Retrieval Augmented Generation (RAG), it ensures privacy-preserving data use. Applications in banking and health insurance will show how ethical AI promotes transparency, inclusivity, and a sustainable data economy.
	As AI systems become increasingly embedded in society, the need to realign their development with ethical, transparent, and socially equitable principles grows urgent. Today's data ecosystem often centralizes value extraction among large entities, marginalizing the individuals and communities that contribute the data powering these technologies. This research seeks to reverse that dynamic by designing AI solutions that redistribute data value through cooperative, citizen-centered mechanisms. The study will explore how privacy-preserving technologies, such as LLM-based synthetic data generation and Retrieval Augmented Generation (RAG), can enable ethical, decentralized data use for AI training and service delivery. To achieve this ambitious goal, a number of interdependent research objectives (ROs) will be pursued: RO1. Design of a framework for data cooperatives as trustworthy custodians of user-contributed data. ++ RO1a. Investigate governance models that ensure citizen control, consent, and transparent decision-making over data usage. ++ RO1b. Define incentive structures that encourage voluntary data contribution and foster engagement within cooperative models. ++ RO1c. Develop ethical principles and technical requirements to guide cooperative data reuse for AI services. RO2. Develop privacy-preserving data pipelines based on LLM-generated synthetic data. ++ RO2a. Evaluate the use of generative AI to produce tabular, textual, and image-based synthetic data while preserving statistical utility.

Objectives	++ RO2b. Assess model fidelity, bias transfer, and representational fairness in synthetic datasets derived from cooperative data pools. RO3. Enable knowledge extraction from decentralized data sources via Retrieval Augmented Generation (RAG). RO4. Validate the cooperative AI framework through domain-specific use cases (e.g., applications in banking and health insurance). These objectives span AI, data ethics, digital policy, and human-computer interaction, fostering a multidisciplinary research trajectory that prioritizes equity and transparency in AI system design. Year 1: The candidate will study existing models of data governance, synthetic data generation, and privacy-enhancing technologies. They will survey frameworks for data cooperatives, focusing on how to formalize citizen engagement, ownership, and consent. A first iteration of the cooperative model will be defined, integrating early synthetic data generation pipelines based on LLMs. Year 2: The focus will shift toward technical development and experimentation. The candidate will implement and evaluate LLM-based synthetic data generators, testing their fidelity, privacy, and applicability across formats. Parallel efforts will build RAG pipelines capable of securely querying distributed data. Pilot studies in banking and insurance domains will assess system performance, ethical compliance, and user perception. Year 3: In the final phase, the candidate will develop a user-facing platform enabling citizens and institutions to contribute data, view impact metrics, and manage consent dynamically. The full cooperative AI system will be deployed in at least two sectors (banking and insurance), with technical benchmarking and participatory evaluations involving stakeholders. The candidate will also initiate partnerships to explore the biodiversity use case and contribute to policy dialogues around data sovereignty. Throughout Years 2 and 3, the Phd student will promote cross-sector collaboration and open dissemination of results, through academic publications, public workshops, and joint initiatives with civil society and industry actors. The study aims not only to advance foundational AI research but also to operationalize ethical, inclusive data economies with long-term social value.

Skills and competencies for the development of the activity	Strong background in data science and deep learning and natural language processing.