ARTIFICIAL INTELLIGENCE

DM 630/Univ. Ca' Foscari/LARUS - Towards Trustworthy AI with Graphs

Funded By	UNIVERSITA' DI VENEZIA - CA' FOSCARI [P.iva/CF:00816350276] LARUS BUSINESS AUTOMATION S.R.L. [P.iva/CF:03540680273] MINISTERO DELL'UNIVERSITA' E DELLA RICERCA [P.iva/CF:97429780584]
Supervisor	DI CARLO STEFANO - stefano.dicarlo@polito.it
Contact	Dr. Sebastiano Vascon (UniVE) Prof. Marcello Pelillo (UniVE)
Context of the research activity	AI technologies have significant potential to transform society, however they also carries out risks that can negatively impact individuals, groups and organizations. AI risk management is therefore mandatory for a responsible development and use of AI systems. This research aims to demonstrate how effective graph-based methods are to mitigate risks and improve trustworthiness of AI. Research topics include improving explainability, resistance to malicious attacks and constraining model biases. This scholarship refers to an industrial collaboration with the company LARUS Business Automation Srl. Progetto finanziato dal PNRR a valere sul DM 630/2024 - CUP: E14D24002330004
	AI technologies have significant potential to transform society and people's lives, from commerce to health, from transportation to cybersecurity, to the environment and our planet. They can therefore foster inclusive economic growth and support scientific advances that improve the conditions of our world. However, AI technologies also carry risks that can negatively impact individuals, groups, organizations, communities, societies, the environment, and the planet. Without adequate controls, AI systems can amplify, perpetuate, or exacerbate inequitable or undesirable outcomes for individuals and communities, taking unclear or even unexplainable decisions. Conversely, with adequate controls,

Objectives	AI systems can mitigate and manage inequitable outcomes. These risks make AI a particularly challenging technology for both organizations and society to implement and use. AI risk management is a key component of the development and responsible use of AI systems. This research aims to exploit graph-based methods and models for mitigating risks on artificial intelligence, such as understanding and eliminating/reducing models' bias, improving explainability, which is fundamental for making informative decisions, and resilience to malicious attacks. The approach involves using graph-based models (mainly graph neural networks) and graph data to exploit contextual relationships. Exploiting these relationships should, in principle, provide collective evidence of the right decision to make, making these models more robust and mitigating the effect of potential biases. During the research activities, the selected candidate will explore the state of the art on explainable AI in the area of deep neural networks for graphs and develop new mechanisms for that, which will later be industrialized. The study will focus initially on static graphs, moving forward to dynamic settings where the underlying graph evolves, to conclude by considering higher-order relationships (hypergraphs). Finally, the Ph.D. student will delve into the topic of robustness as a key aspect for the adoption of responsible AI - especially in the application of such models in regulated contexts (healthcare, finance, etc.) and in compliance with new regulations (AI Act, ISO 42001, etc.). Indeed, both aspects (explainability and robustness) are affected by model and dataset biases; hence techniques to mitigate this effect (like regularization, diverse and representative training data sampling, human in the loop, etc...) will be part of the exploration.

Skills and competencies for the development of the activity	The candidate should have a solid background on the topics for this research, in particular: - AI and Machine Learning - Linear algebra and statistics - Python for Data Science - Deep Learning Libraries (Pytorch, TensorFlow, etc) - Proven experience in building graph-based models and (graph)-neural networks is an important plus.