DIPARTIMENTO DI
INFORMATICA, SISTEMISTICA E
COMUNICAZIONE
Viale Sarca, 336
20126 Milano

ReGAInS

ONLINE SEMINAR ANNOUNCEMENT

The next meeting of the Neuro-Symbolic Seminar (NeSS) will be held on Tuesday November 19th at 4.30 pm

The seminar will be online. Registration is mandatory at the following link:
https://unimib.webex.com/weblink/register/r048232b17492ee070afbf6c7e3c93cc6

Interpretable Concept-Based Memory Reasoners

Speaker
Giuseppe Marra
KU Leuven

Abstract: The lack of transparency in the decision-making processes of deep learning systems presents a significant challenge in modern artificial intelligence (AI), as it impairs users' ability to rely on and verify these systems. To address this challenge, Concept-Based Models (CBMs) have made significant progress by incorporating human-interpretable concepts into deep learning architectures. This approach allows predictions to be traced back to specific concept patterns that users can understand and potentially intervene on.
However, existing CBMs' task predictors are not fully interpretable, preventing a thorough analysis and any form of formal verification of their decision-making process prior to deployment, thereby raising significant reliability concerns. To bridge this gap, we introduce Concept Memory Reasoner (CMR), a novel CBM designed to provide a human-understandable and provably-verifiable task prediction process. Our approach is to model each task prediction as a neural selection mechanism over a memory of learnable logic rules, followed by a symbolic evaluation of the selected rule. The presence of an explicit memory and the symbolic evaluation allow domain experts to inspect and formally verify the validity of certain global properties of interest for the task prediction process.
Experimental results demonstrate that DCR achieves better accuracy-interpretability trade-offs to state-of-the-art CBMs, discovers logic rules consistent with ground truths, allows for rule interventions, and allows pre-deployment verification.

Bio: Giuseppe Marra is Assistant Professor at the Department of Computer Science, KU Leuven, where he is part of the Declarative Languages and Artificial Intelligence (DTAI) research group. He was an FWO Post-Doc Fellow on the project "Deep Statistical Relational Learning" with Prof. Luc De Raedt. He obtained his PhD at the University of Florence, Italy, in 2020 under the supervision of Prof. Marco Gori. His research interests cover the integration of neural computation and relational reasoning, with a particular focus on logical and probabilistic reasoning. This endeavour intersects with the fields of explainable AI, safe reinforcement learning, generative AI and geometric deep learning.

Contact person for the seminar: rafael.penaloza@unimib.it