From: "Martens, J.J.M. (Jan)" <j.j.m.martens@liacs.leidenuniv.nl>
Subject: LIACS research seminar with Inês Gomes
Date: 24 October 2024 at 13:04:14 CEST
To: LIACS - Staff <Staff@liacs.leidenuniv.nl>, "ines.gomes@fe.up.pt" <ines.gomes@fe.up.pt>

Dear all,

please come attend the next

LIACS Research Seminar

the monthly seminar series where LIACS researchers present their research to each other. The speaker this time will be

Inês Gomes

Inês Gomes
(Porto University)

on Mon 4 Nov at noon
(room DM1.09)

(Free lunch included: please register a week in advance by accepting this calendar invite)

Also if the topic is not directly related to your research, please do come! Because such a seminar is an opportunity to learn from each other, about the research highlights within LIACS.

Title: Pushing the Limits: Stress-Testing Machine Learning Models to Understand Decisions

Abstract: The widespread adoption of machine learning and deep learning models has revolutionized various fields. However, it has also introduced significant challenges related to interpretability and responsible use, particularly in critical applications. While advances have improved model interpretability, a substantial gap exists in understanding how they handle low-confidence

predictions. This talk introduces stress testing, a data-driven method to identify model weaknesses by analyzing their decision boundaries. We propose a framework that leverages generative models to create data near a classifier's decision boundary. We can better understand how models handle low-confidence images by selecting representative boundary samples (prototypes) and applying post hoc explanation techniques.

This framework can also improve transparency by enhancing model cards - a report summarising a model's performance, limitations, and ethical considerations. By incorporating insights from stress testing into these reports, stakeholders can make more informed decisions when deploying machine learning systems.

When: Monday 4 November 2024, 12.00-13.00

Where: DM1.09

Speaker: Inês Gomes