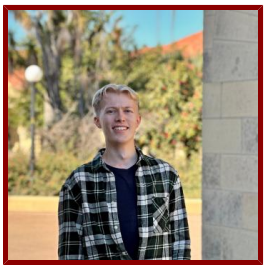
Stanford



Anders Gjølbye Madsen

Graduate Visiting Researcher Student, Computer Science

Bio

BIO

Anders Gjølbye Madsen is a PhD fellow at the Technical University of Denmark. His research focuses on trustworthy machine learning for healthcare, with an emphasis on explainability, interpretability, and reliable evaluation of models in high-stakes settings. He works broadly with modern deep learning methods, including self-supervised learning, and is interested in questions of robustness and alignment. He is the author of PatternLocal, a NeurIPS 2025 paper on reducing false-positive attributions in explanations of non-linear models by refining local explanation approaches. He earned a BSc in Artificial Intelligence and Data from DTU and completed an MSc in Engineering in Applied Mathematics at DTU, including a study exchange in Computational Science and Engineering at ETH Zürich. Anders will spend 2026 as a visiting researcher at Stanford University's Trustworthy AI Research (STAIR) Lab, working with Professor Sanmi Koyejo.