

Théo Sourget

tsou@itu.dk | <https://www.tsourget.fr> | <https://github.com/TheoSourget>

EDUCATION

Université de Rouen

Master in Data Science and Engineering, highest honors, Valedictorian

Rouen, France
2021-2023

Université de Rouen

Bachelor in Computer Science and Data Science, honors, Valedictorian

Rouen, France
2020-2021

IUT de Vannes

2 Years University Diploma in Computer Science (DUT Informatique)

Vannes, France
2018-2020

CURRENT & RECENT EMPLOYMENT

IT University of Copenhagen

PhD Student, Pattern Recognition Revisited lab

Copenhagen, Denmark
September 2025 – Current

- I study data quality and the evaluation of machine learning algorithms
- Supervised by Veronika Cheplygina

CentraleSupélec, University of Paris-Saclay

Research Engineer, MICS laboratory

Gif-sur-Yvette, France
April 2025 – September 2025

- I studied CLIP-based models on the classification of chest X-rays
- I assessed their fairness across subgroups of patients and their robustness to shortcut learning

IT University of Copenhagen

Research Assistant, Pattern Recognition Revisited lab

Copenhagen, Denmark
October 2023 – January 2025

- I assessed the usage of non-relevant information in medical images (chest x-rays, eye fundus) by CNN models
- I studied the presence of public medical datasets in scientific papers and created open-access tools based on the research
- I defined and wrote a PhD funding application on the evaluation of multi-modal datasets and models
- I assisted in the data collection and data processing of other lab projects

IT University of Copenhagen

Assistant Lecturer, Department of Computer Science

Copenhagen, Denmark
October 2023 – July 2024

- I assisted students during practical sessions of the Data In the Wild Course of Fall 2023 (data processing, visualisation, web scraping)
- I prepared and gave lectures to bachelor's students of the Project in Data Science course of Spring 2024 (Git, decision trees, ensemble of classifiers, common problems in machine learning projects)
- I designed the exercises and coordinated the teaching assistants for the practical sessions

Capgemini Engineering

Data Scientist Intern, Medic@

Illkirch-Graffenstaden, France
April 2023 – September 2023

- I performed teeth detection, segmentation and numbering in X-ray dental panoramic to help diagnose rare diseases for both adults and children.
- I compared the performances of Mask-RCNN and Detection Transformer (DETR)
- I defined an algorithm to generate new dental panoramics using existing images and inpainting models
- I used different generative models (GAN, diffusion models) to generate new teeth

PUBLICATIONS

- Fairness and Robustness of CLIP-Based Models for Chest X-rays,
T. Sourget, D. Restrepo, C. Hudelot, E. Ferrante, S. Christodoulidis, M. Vakalopoulou, FAIMI MICCAI Workshop, 2025
- Mask of Truth: Model Sensitivity to Unexpected Regions of Medical Images,
T. Sourget, M. Hestbek-Møller, A. Jiménez-Sánchez, J.J. Xu, V. Cheplygina. Journal of Imaging Informatics in Medicine, 2025
- In the Picture: Medical Imaging Datasets, Artifacts, and their Living Review,
Amelia Jiménez-Sánchez, [...], **Théo Sourget**, [...], Veronika Cheplygina, ACM FAccT, 2025
- Copycats: the many lives of a publicly available medical imaging dataset,
A. Jiménez-Sánchez, N.R. Avlona, D. Juodelyte, **T. Sourget**, C. Vang-Larsen, A. Rogers, H.D. Dariusz Zajac, V. Cheplygina. NeurIPS Datasets and Benchmarks Track, 2024.
- [Citation Needed] Data usage and citation practices in medical imaging conferences.
T. Sourget, A. Akkoç, S. Winther, C. L. Galsgaard, A. Jiménez-Sánchez, D. Juodelyte, C. Petitjean, and V. Cheplygina. Medical Imaging with Deep Learning (MIDL), 2024.
- Detection transformer for teeth detection, segmentation, and numbering in oral rare diseases: Focus on data augmentation and inpainting techniques,
H. Kadi, **T. Sourget**, M. Kawczynski, S. Bendjama, B. Grollemund, and A. Bloch-Zupan. International Conference on Computational Science and Computational Intelligence (CSCI), 2023.
- Can Segformer be a true competitor to U-net for medical image segmentation?
T. Sourget, S. N. Hasany, F. Mériaudeau, and C. Petitjean. Medical Image Understanding and Analysis (MIUA), 2023.

OTHER ACHIEVEMENTS

Presentations

- **MICS Seminar (Talk)**: Practices and problems with publicly available medical datasets
- **D3A Conference 2.0 2024 (Poster)**: Mask of truth: model sensitivity to unexpected regions of medical images
- **”In the Picture: Medical Imaging Datasets” workshop (Oral)**: [Citation Needed] Data usage and citation practices in medical imaging conferences
- **MIDL 2024 (Oral and Poster)**: [Citation Needed] Data usage and citation practices in medical imaging conferences
- **D3A Conference 1.0 2024 (Poster)**: [Citation Needed] Data usage and citation practices in medical imaging conferences
- **GRETSI 2023 (Poster)**: Can Segformer be a true competitor to U-net for medical image segmentation?

Fundings

- **DDSA visit grant** of EUR 1645 in 2023 to work with Dr Veronika Cheplygina at the IT University of Copenhagen
- **Graduate school MinMacs scholarship** of EUR 15 000 (funded by the Normandy region) during both the first and second year of my Master’s degree, giving access to funding, a supervisor and an immersion in laboratories

Other roles

- Reviewer for the FAIMI MICCAI workshop 2025
- Reviewer for the MELBA 2025 Special issue on Fairness of AI in Medical Imaging (FAIMI)
- Reviewer for the Medical Imaging with Deep Learning (MIDL) 2025 conference
- Member of the organisation committee of the “Datasets through the Looking-Glass” webinar
- Assisted in the technical setup of the “In the Picture: Medical Imaging Datasets” workshop