T H O M A S   V A K I L I

Department of Computer and Systems Sciences
Stockholm University
+46 8 16 16 59
thomas.vakili@dsv.su.se

Körsbärsvägen 9 lgh 2102
114 23 Stockholm, Sweden
+46 704 38 34 26
thomas@vakili.net

Education

PhD Student - Stockholm University   2021 -

Currently studying towards a PhD degree in computer science at the Department of Computer and Systems Sciences. The position is focused on privacy preserving machine learning for natural language processing and my supervisor is Professor Hercules Dalianis.

MSc in Computer Science - KTH Royal Institute of Technology   2013 - 2020

Studied the five-year engineering program which grants an engineering diploma and an MSc in computer science. My specialization was natural language processing and machine learning.

Persian Philology - Uppsala Universitet   2016 - 2017

Between the bachelor and master parts of the engineering program I studied Persian for a year at the Department of Linguistics and Philology. This improved my skills in my second mother tongue and also deepened my knowledge of linguistics.

Industry Experience

Consultant - Netlight   Q3 2019 -

Netlight is a prestigious IT consultancy firm based in Stockholm and with offices in multiple European cities. I am currently on a leave of absence during my PhD studies.

Data Engineer - Spotify (via Netlight)   Q4 2020 - Q1 2021

Worked as a data engineer in a multinational team at Spotify. The focus of this assignment was helping Spotify develop their data pipeline to improve their curation of core datasets.

Data Scientist - PostNord (via Netlight)   Q3 2019 - Q4 2020

PostNord is a major logistics company based in the Nordics. I worked as a data scientist and backend developer to build machine learning-based route planning and time series prediction.

Backend Developer - Truecaller   Q2 2018 - Q4 2018

Truecaller is a Swedish software company that provides dialer identification and payments in many different countries. I worked as a backend developer in their search team.

Publications

Vakili, T., & Dalianis, H. (2023). Using Membership Inference Attacks to Evaluate Privacy-Preserving Language Modeling Fails for Pseudonymizing Data. *Proceedings of the 24th Nordic Conference on Computational Linguistics.*

Dolk, A., Davidsen, H., Dalianis, H., & Vakili, T. (2022). Evaluation of LIME and SHAP in Explaining Automatic ICD-10 Classifications of Swedish Gastrointestinal Discharge Summaries. *Proceedings of the 18th Scandinavian Conference on Health Informatics*, 166–173.

Vakili, T., Lamproudis, A., Henriksson, A., & Dalianis, H. (2022). Downstream Task Performance of BERT Models Pre-Trained Using Automatically De-Identified Clinical Data. *Proceedings of the 13th Language Resources and Evaluation Conference (LREC 2022)*, 4245–4252.

Bridal, O., Vakili, T., & Santini, M. (2022). Cross-Clinic De-Identification of Swedish Electronic Health Records: Nuances and Caveats. *Proceedings of the Workshop on Ethical and Legal Issues in Human Language Technologies and Multilingual De-Identification of Sensitive Data In Language Resources at LREC 2022*, 49–52.

Jerdhaf, O., Santini, M., Lundberg, P., Bjerner, T., Al-Abasse, Y., Jonsson, A., & Vakili, T. (2022). Evaluating Pre-Trained Language Models for Focused Terminology Extraction from Swedish Medical Records. *Proceedings of the Workshop on Terminology in the 21st Century: Many Faces, Many Places at LREC 2022*, 30–32.

Vakili, T., & Dalianis, H. (2022). Utility Preservation of Clinical Text After De-Identification. *Proceedings of the 21st Workshop on Biomedical Language Processing at ACL 2022*, 383–388.

Vakili, T., & Dalianis, H. (2021). Are Clinical BERT Models Privacy Preserving? The Difficulty of Extracting Patient-Condition Associations. In *Proceedings of the AAAI 2021 Fall Symposium on Human Partnership with Medical AI: Design, Operationalization, and Ethics.*

Vakili, T. (2020). A Method for the Assisted Translation of QA Datasets Using Multilingual Sentence Embeddings. Presented at the workshop on *RESOURCEs and representations For Under-resourced Languages and domains at SLTC 2020.*

Teaching Experience

Internet Search Techniques and Business Intelligence   Fall 2022 -

I am a lecturer and lab examiner in this master-level course. My lecture is about opinion mining and sentiment analysis, and I also examine and develop the labs in the course.

Language Technology   Fall 2022 -

I teach a bachelor level course about language technology. My duties include examining lab assignments as well as giving a lecture about lexical resources.

Natural Language Processing   Spring 2022 -

I am a teacher in the natural language processing course which is given to master students in my department. My responsibilities include holding a lecture on deep learning in NLP as well as designing and examining lab assignments and project reports.

Digital Business Strategies and Change Management   Fall 2021

I was a teaching assistant in a course on digital management in which I was responsible for grading essays and presentations.

Information Retrieval   Spring 2019

During my studies at KTH, I was a teaching assistant in a course on information retrieval. I was responsible for grading lab assignments and helping students understand the material.

Thesis Supervision

Master's thesis of Alexander Dolk and Hjalmar Davidsen   Spring 2022

Main supervisor for a thesis project about explainable AI applied to clinical NLP. The thesis investigated whether clinicians preferred LIME or SHAP for explaining the predictions of ICD-10 codes. Resulted in a conference paper at SHI 2022.

Master's thesis of Xiaoman Tang   Spring 2022

Main supervisor for a thesis project about bias in NER. The thesis evaluated to what extent NER models based on Spacy and models based on BERT misclassified names associated with marginalized demographic categories.

Language Skills

Swedish – native
English – fluent
Persian – advanced
Spanish – intermediate