D Research qualifications portfolio

1 Reflection on research activities

Research Goal and Vision

My overarching research goal is to understand humans through the lens of computational, probabilistic, and complexity-oriented science, to contribute to a more equitable, sustainable, and knowledge-driven society. Across my work, I studied human cognition, behaviour, interaction, and social organization as complex systems shaped by uncertainty, bounded rationality, and structural constraints. While my methodological toolkit is rooted in mathematics, computer science, and machine learning, my scientific motivation is fundamentally social and humanistic. I want to provide the methods to understand how humans learn, interact, innovate, make decisions and are affected by their social context.

My long-term ambition is helping to develop methodological frameworks to understand human systems from micro-level (cognitive) processes to macro-level social dynamics. These methods should not only be scientifically rigorous but also interpretable, communicable, and socially actionable.

A principle that has always guided me, and ultimately the reason why I am in science, is: “commit to an idea that is bigger than yourself.” My educational and academic development has always been driven by curiosity, the desire to learn new things, and the wish to contribute to this elusive thing we call science, which seems greater than the individual researchers contributing to it.

In the following, I reflect on my personal journey, including the experiences, challenges, and moments of growth that ultimately led me here.

Academic Journey and Achievements

Despite still being in the early phase of my academic career, my research profile already includes publications across top-tier venues in cognitive science, education research, human–computer interaction (HCI) and data science. To date, I have published 15 peer-reviewed articles in journals such as *Scientific Reports*, *Learning and Instruction*, *Computers in Human Behavior*, *MethodsX*, and *British Journal of Educational Psychology*, as well as flagship conferences in HCI and data science (ETRA, ISMAR, ICMI, CHI, PAKDD). These publications combine traditional RCT[1]-based research designs with information-theoretic and network-based analyses, interpretable machine learning, and multimodal data analysis.

I have always been strongly interested in methods. During my PhD, I learned how to design and conduct RCT-based experiments, develop questionnaires, build virtual reality environments, and analyse eye-tracking data. This training provided the foundation for my later transition toward more computational and data-heavy research. In parallel, I continued building on my master-level training, which was strongly oriented toward data science and machine learning. In my dissertation papers, I used machine learning approaches to investigate individual behaviour and better understand cognitive processes. Through this work, I became familiar with using machine learning to investigate heterogeneous treatment effects. In parallel, I gained experience with network analysis and multimodal behavioural modelling using complex observational data. My PhD period was also marked by COVID-19, which limited opportunities for conference participation and international research exchange. Nevertheless, the period allowed me to develop a broad methodological foundation spanning experimental research, computational modelling, and multimodal data analysis.

After completing my Dr. rer. nat., I deliberately transitioned into a position with a stronger methodological and computational focus. This brought me into the current position in human geography, where I have now been formally a postdoctoral fellow for less than two years in order to extend my research toward computational and data-intensive analyses of social and spatial processes. Within this relatively short period, I established machine learning and computational social science workflows for large-scale register and online data, including high-performance computing, data harmonization, data imputation, sampling, and the design and training of models on multi-million-record datasets. This work already resulted in an accepted publication at a top-tier data mining conference (PAKDD 2026), alongside additional manuscripts in late-stage preparation on latent representations of human capital and industrial opportunity spaces. In parallel, I secured initial competitive funding from AI Lund, contributed to national and EU-level grant applications, and built interdisciplinary connections spanning economic geography, innovation research, computational social science, GeoAI, and applied machine learning. Further details are provided in the following section on my current interdisciplinary research profile.

[1] Randomized control trial

At the Department of Human Geography, I was the first computer scientist to be hired. I built computational infrastructures and helped establish an understanding that data-driven research processes are often highly non-linear and require substantial work on data curation, harmonization, and model development before evaluation itself becomes possible. Part of my work therefore involved not only establishing computational structures and practices within the department, but also communicating differences in workflows, expectations, and research timelines across disciplines. I believe this effort is one reason why true interdisciplinary research often produces more fruitful scientific results in the long run. In this sense, I strongly identify with the idea that “if you want to go fast, go alone; if you want to go far, go together.”

At the same time, this transition gave me the opportunity to introduce data science and computational approaches into the department more broadly. Because I had already worked across disciplines, I was able to quickly establish connections across methodological traditions, and together with colleagues we founded the Hägerstrand Lab as a shared computational research environment. Beyond research itself, I greatly enjoy developing research environments that bring people together around shared methodological and scientific questions. I had already experienced this during my PhD through the collaborative VR lab environment we created, and I hope to continue building similar interdisciplinary spaces in the future.

This interdisciplinary and computational research environment forms the basis of my current research program, which I outline in the following sections.

Current Interdisciplinary Profile

Human Capital, Mobility, and Economic Dynamics

A central strand of my research focuses on register-based analyses of human capital, career trajectories, and labor market dynamics. Within the NEXUS project - *Machine Learning on Register Data* - I develop machine learning approaches for large-scale Swedish microdata. A key idea in this work is to treat human capital as a latent and dynamic construct that can be inferred from individual career trajectories, rather than as a static variable measured only through education, occupation, or income.

We use the idea that life trajectories can be represented as sequences. In our published paper on text-based life trajectories from Swedish register data, we used Transformer models to predict individual residential mobility from longitudinal register data. The next goal is to link these individual trajectories more directly to questions of economic development, innovation, and regional transformation.

A second part of this research concerns the temporal harmonization of Swedish industry classifications. Together with Andreas Erlström, Alvaro von Borries, Markus Grillitsch, Ola Hall, and Alexandros Sopasakis, we work on methods that allow longitudinal analyses of industrial transitions despite changes in classification systems over time. This work has also led to the development of a latent representation of Swedish industry classifications, which we are currently preparing for publication.

Building on this, we aim to integrate graph-based approaches using employer-employee relations and firm transitions in the register data. This allows career mobility, firm connections, and labour market structures to be analysed as relational processes. Together with Yuan Liao and Markus Grillitsch, I also recently started a project combining mobile phone and register data to study inter-firm knowledge connectivity and labour mobility as mechanisms of regional learning and economic transformation.

Opportunity Structures and Spatial Inequality

A second strand of my research focuses on regional opportunity structures, industrial change, and spatial inequality. In particular, I am interested in how regional economic development emerges from the interaction between industries, labour mobility, and locally available capabilities.

A central project within this strand is the development of latent industry representations for studying regional path development. The goal is to construct a shared representation space that combines different forms of industry relatedness, including worker mobility, semantic similarity, hierarchical industry structures, and input–output relations. Building on this, we developed a concept for opportunity space that links latent industry structures with region-specific economic compositions. This allows regional development, industry emergence, and industrial decline to be analysed as geometric and relational processes within a shared latent space.

Methodologically, this work combines representation learning, labour mobility data, and language embeddings to model industrial relatedness and regional transformation across time.

Further, I recently started a project on racial spatial inequality and urban transformation in Brazil together with Lorena Melgaço, Luana X. P. Coelho, and Fabio Schreiber. Using three decades of grid-based census data, the project investigates racial concentration, segregation patterns, and urban transformation processes surrounding large-scale urban development projects in Brazil. More specifically, the project examines how racialized spatial inequalities persist beyond socioeconomic differences and how urban planning and economic development processes contribute to the reproduction of unequal spatial structures.

Institutions, Discourse, and Multimodal AI

A third strand of my research focuses on informal institutions, discourse, and symbolic systems using large-scale textual and multimodal data. In particular, I am interested in the use of large language models (LLMs) and multimodal large language models (MLLMs) as computational tools for scaling semantic analysis within the social sciences. This includes both substantive applications and methodological questions related to uncertainty, interpretability, reliability, and the role of these models within scientific workflows.

Together with Jonathan Friedrich, Melissa Cardona, and Linda Stihl, I currently work on AI-supported newspaper analysis to evaluate institutional adaptation and policy failure in EU textile waste governance. This work combines large-scale newspaper corpora with semantic and temporal analysis to examine discursive shifts, institutional coordination failures, and policy dynamics surrounding the green transition. At the same time, the project also serves as a methodological testing ground for understanding how LLM-based approaches can be integrated into social scientific text analysis in reliable and theoretically grounded ways.

In a complementary project together with Ola Hall, Eva-Charlotte Ekström, and Magnus Jirström, in cooperation with researchers in Addis Ababa, we apply vision-language models to analyse advertising practices in Ethiopian souk shops. The project investigates how visual communication strategies relate to local economic conditions, market structures, and informal institutional contexts using multimodal AI approaches. More broadly, this work explores how multimodal models can be used to analyse symbolic and visual communication processes across different social and spatial settings.

Across these projects, my broader interest lies in understanding how symbolic systems, discourse, and institutional structures emerge, stabilize, and transform through human interaction and communication processes. At the same time, I increasingly use these projects to explore the methodological potential and limitations of LLMs and MLLMs for computational social science, particularly regarding semantic analysis at scale, interpretability, trustworthiness, and observational inference.

Causal Machine Learning and Computational Social Science

An increasingly important direction of my research concerns causal machine learning, such as causal representation learning. Much of contemporary machine learning is highly successful at prediction but often remains limited in its ability to identify underlying mechanisms, intervention effects, and transferable causal structures. I am therefore increasingly interested in how causal approaches can improve the interpretability and robustness of machine-learning-based research.

This perspective is becoming increasingly relevant within several of my ongoing projects. Together with Ola Hall, Alexandros Sopasakis, and Marina Toger, I am currently involved in a project application to the Swedish National Space Agency (Rymdstyrelsen) that combines machine learning and satellite imagery to develop longitudinal uncertainty-aware imputation methods for studying the relationship between temperature and health. Within this work, causal and interventional approaches play an important role for addressing uncertainty, transferability, and observational bias in large-scale spatiotemporal data.

Conceptually, this work is strongly influenced by the causal framework introduced by Judea Pearl, which became particularly prominent within epidemiology and computer science through the use of structural causal models and directed acyclic graphs (DAGs) to formalize causal structures and interventions. For me, this also reconnects to earlier interests in probabilistic inference that I encountered during my master studies, particularly through the lecture on probabilistic inference by Philip Hennig in Tübingen. Conceptually, this research is guided by ideas from causal representation learning and interventional machine learning developed by researchers such as Bernhard Schölkopfand collaborators.

Research Environment and Scientific Exchange

Over time, I realized that I work best in close collaboration with others. Computational research often allows for a large number of possible modelling decisions, configurations, and evaluation strategies. This complexity made me increasingly appreciate the importance of continuous methodological

exchange and short feedback loops within collaborative environments. I therefore consciously worked on exposing ideas earlier, discussing methodological decisions openly, and integrating feedback throughout the research process. At the same time, I realized how important it is to work in environments where colleagues share at least parts of the same scientific language and where open methodological discussion is encouraged.

Collaboration has therefore become one of the central motivations behind my work. The projects I found most scientifically productive were usually those that brought together researchers with different competencies and methodological perspectives around a shared problem. This also required me to develop communication skills that allow me to bridge between disciplines and research traditions. My experience is that successful interdisciplinary collaboration emerges when strong domain knowledge is combined with clear research questions, deep understanding of the data, and methods that appropriately fit the problem at hand.

During my relatively short time at Lund University, I became involved in a broad range of interdisciplinary collaborations and research environments. Beyond the Hägerstrand Lab, I am connected to CIRCLE through collaborations in innovation research. I am also involved in activities connected to AI Lund, the Social Science Methods Centre, and the Regional Dynamics Group at the Department of Human Geography. In addition, I participate in collaborations connected to SESAC, WASP Lighthouse, and eSSENCE through projects involving machine learning. Together with colleagues, I am currently involved in national and European grant applications and in the organization of training activities related to AI and machine learning methods for early-career researchers by being part of the organizing team for an autumn school on AI, co-organized with EU-SPRI.

My work is guided by principles of open science and reproducibility. Whenever possible, I publish open access, share code through GitHub, and work toward transparent and reusable research workflows following FAIR principles.

Beyond my own research, I contribute to different scientific communities. I am regularly involved in reviewing for journals and conferences across data science, human-computer interaction, machine learning, and cognitive science, including *Acta Psychologica*, *PLOS Computational Biology*, *International Journal of Human–Computer Interaction*, ISMAR, ETRA, and ICMI. In addition, I contributed to the organization of the ACM Symposium on Eye Tracking Research and Applications (ETRA) in 2023 and 2024, including technical support, student volunteer coordination, and on-site conference management for a large international conference community. These activities further strengthened my interest in collaborative and interdisciplinary research environments.

Future Research Plan

Over the coming years, I plan to continue and expand my ongoing research with register data, with a stronger focus on combining register data with additional data sources such as web data, mobile phone data, and earth observation data. The goal is to develop multimodal machine learning models that can serve as foundational models for simulating human behaviour, social dynamics, and spatial processes.

Another direction of my future research concerns large language models as statistical models and scientific tools. I am particularly interested in the growing intersection between statistics and LLM research, especially regarding uncertainty, interpretability, distribution shift, and the role of LLMs within scientific workflows. In this context, I want to further explore the use of LLMs for tasks such as data collection, data cleaning, and data analysis. Their contextual understanding and scalability can improve both efficiency and analytical depth, while also enabling textual data to be analysed in semantically meaningful ways at an unprecedented scale. At the same time, I am interested in the question of how LLMs internalize statistical structure and observational distributions and how these models can be integrated into computational sciences in theoretically grounded ways.

This development also goes hand in hand with a stronger engagement with causal machine learning. I therefore want to further expand the use of causal and interventional approaches to address out-of-distribution problems, transferability of research findings, and uncertainty-aware modelling within observational settings. This also includes the development of complex probabilistic models for data imputation and simulation. As a first step, I started planning a workshop on causal machine learning with observational data.

A further strand of my future research addresses interpretability, visualization, and uncertainty communication in AI- and machine-learning-based research. I am particularly interested in developing visualization techniques for complex probabilistic models that support scientific interpretation, uncertainty communication, and policy relevance. I want to codevelop these approaches in research-based teaching and interdisciplinary computational lab environments.

In terms of scientific dissemination, I aim to continue balancing publications between domain-specific journals and machine learning conferences. I consider this important both for maintaining strong connections to substantive research questions and for staying engaged with rapidly evolving methodological developments. To support my plans, I will pursue targeted funding to secure research assistance, computational infrastructure, and academic staff through national funding bodies. In the longer term, I aim to lead a Horizon Europe consortium and apply for ERC Starting or Consolidator Grants in order to establish a sustainable long-term research group.