Niclas Griesshaber

niclasgriesshaber@outlook.com   niclasgriesshaber.com   github.com/niclasgriesshaber

+49 176 4260 4186   linkedin.com/in/niclasgriesshaber   huggingface.co/niclasgriesshaber

Education

2024	**University College London**, *United Kingdom* M.Sc. Computational Statistics and Machine Learning, Distinction
2023	**University of Oxford**, *United Kingdom* M.Sc. Economic and Social History, Distinction, Dissertation Prize, Top of Class
2022	**National University of Singapore**, *Singapore* Exchange Semester
2022	**University of Tübingen**, *Germany* B.Sc. Cognitive Science, First Class Honours, Top 10%
2021	**University of Tübingen**, *Germany* B.Sc. International Economics, First Class Honours, 3rd Best Bachelor's Degree
2017	**Johannes-Kepler-Gymnasium**, *Germany* Abitur, 4th Best High School Graduate out of 116 Students
2015	**Holy Trinity Catholic Secondary School**, *Canada* Year Abroad during High School

Fellowships & Awards

2023	Joan Thirsk Prize for the Best Dissertation in the M.Sc. Economic and Social History
2021	RWT GmbH Prize for the 3rd Best Bachelor's Degree in International Economics
2019–	Scholarship Fellow of the Konrad-Adenauer-Foundation
2017	Winner Baden-Württemberg in the Federal President's History Competition
2017	Special Prize in Economics by Reutlingen's District Savings Bank
2017	History Prize of Reutlingen by the Historical Society of Reutlingen
2017	E-Fellows Online Scholarship for 4th Best High School Graduate

Teaching

University of Tübingen

Chair of Computer Vision, Andreas Schilling

2022 Summer	Teaching Assistant in Introduction to Machine Learning

Chair of Statistics, Econometrics and Quantitative Methods, Martin Biewen

2020 Winter	Teaching Assistant in Quantitative Methods
2019 Winter	Teaching Assistant in Explorative Data Analysis

Research Assistantships

2023 (5 mos.)	Dr. Giulia Caprini, Fellow at Nuffield College, University of Oxford.
2019 (3 mos.)	Professor Zhaoping Li, Max Planck Institute for Biological Cybernetics, Tübingen.
2017 (1 mos.)	Research Internship at the Berks History Center in Reading, Pennsylvania, USA.

Research Grants

2023	College Contribution Fund of Harris Manchester College, University of Oxford.

PhD Market Paper

Transplanting Craft Guilds to Colonial Latin America: A Large Language Model Analysis
(with Sheilagh Ogilvie; currently under review at The Journal of Economic History)

What can we learn about institutional transplantation by analyzing craft guilds in colonial Latin America? We use large language models (LLMs) to investigate colonial guild ordinances, addressing two major bottlenecks in assessing institutions: digitizing qualitative sources efficiently and analyzing them quantitatively. Our newly designed methodology reveals both long-term continuities and striking differences between craft guilds in colonial Mexico and Peru, particularly with regard to human capital and product quality. The LLM-based approach identifies patterns that were previously not discernible using standard methods in economic history, its results are reproducible, and it can easily be extended to other historical settings.

Papers in Progress

A New Historical Dataset for Machine Translation between Early German and English: Cross-Lingual Transfer, Many-Shot In-Context Learning, and Fine-Tuning
(with Sheilagh Ogilvie, Jiayi Wang, and Yao Lu; in preparation for submission to ACL)

We present a new historical text dataset for machine translation consisting of 3,873 paragraph-level pairs between Early Modern Bohemian German and English. The dataset contains more than 750,000 words and is based on handwritten legal court records dated from 1582 to 1787. The primary sources were manually transcribed and translated across several years by the Chichele Professor of Economic History at the University of Oxford. We evaluate the translation accuracy of open-source large language models (LLMs) on this low-resource historical language pair. These models struggle to generalise their abilities through cross-lingual transfer, as evident from their low translation scores (e.g., for Early Modern Bohemian German to English: BLEU 11.50, COMET 0.57; for the reverse direction: BLEU 2.19, COMET 0.48). To address these challenges, we successfully replicated a recent many-shot in-context learning paradigm for low-resource languages, demonstrating continuous improvements from 0 to 128 shots. Using LLaMA-3.1-8B-Instruct, BLEU scores increased to 24.32 (COMET 0.66) for Early Modern Bohemian German to English and to 7.23 (COMET 0.54) for the reverse direction. Fine-tuning remains the performance ceiling across all experiments, with BLEU scores of 36.70 (COMET 0.69) and 9.55 (COMET 0.57) respectively, highlighting that in-context learning may not reach the same level of effectiveness in handling out-of-distribution data. These findings reinforce the need for more interdisciplinary research, as adapting LLMs to historical languages not only advances our understanding of transformer architectures but also sharpens new analytical tools to advance the field of history.

Multimodal Large Language Models for Layout Detection and Post-OCR Correction: German Business Directories from the 18th and 19th Centuries
(with Gavin Greif)

Historical microdata is scarce prior to the emergence of modern census records. Despite their limitations, city and trade directories may be the most regular and widely available source on urban economic activity at a granular level for the pre-census era. Thus far, the systematic exploitation of historical directories by economic historians has been prevented by the tremendous costs associated with their transcription. We solve this bottleneck by developing the first fully automated data extraction workflow for eighteenth- and nineteenth-century directories. For this, we develop a novel multimodal LLM-based post-OCR correction approach, which accurately corrects the mistakes made by OCR engines. We demonstrate that our pipeline works for difficult historical typefaces (e.g., Fraktur) and different layouts (e.g., multi-column and sub-heading-based classification) without the need for custom pipeline adaptation, model training, or any scan pre-processing. We apply this pipeline to a corpus of German city and trade directories published between 1770 and 1870 to build a database of millions of individual directory entries across 35 German cities. Finally, we develop solutions for classifying individual entries and linking entries across subsequent directories.

Seminars & Conferences

2024	Microdata in Economic History: Beyond the Full-Count Census
2024	Joan Thirsk Dissertation Prize Lecture at the Annual Graduate Workshop, All Souls College, University of Oxford. I was the pre-speaker to the Sir John Hicks Memorial Lecture in Economic History, delivered by Professor Joel Mokyr.
2023	Invitation to participate in the All Souls College Fellowship Examination

Languages & Software Skills

Languages	English (TOEFL: 111/120), German (Native), Spanish (Native), French (DELF B1)
Skills	Python (5 years), GitHub, LaTeX, Hugging Face, R, Matlab, Java, MS Office

Community Service

2022–2023	Middle Common Room Treasurer at Harris Manchester College, University of Oxford
2016–	Certified DSV Ski Instructor (Trainer C) at the Skiclub Offenburg
2018–2020	Co-founder and HR Director of Enactus Tübingen e.V.

Industry Experience

2019 (2 mos.)	**KPMG, Intern in Digital Compliance**, *Stuttgart, Germany*
	• Developed automated workflows for SAP audit processes using BluePrism and VBA, enhancing efficiency in audit report generation.
	• Collaborated with the IT compliance team to identify automation opportunities, reducing manual input and error rates.
2018 (3 mos.)	**Ernst & Young, Intern in Corporate Law**, *San José, Costa Rica*
	• Advised clients on selecting legal structures for business incorporation, ensuring compliance with local regulatory requirements.
	• Created client-facing presentations and internal reports in PowerPoint, clarifying complex legal concepts for non-legal stakeholders.

Personal Information

Citizenships:	Ecuadorean and German

References

Sheilagh Ogilvie, FBA
Chichele Professor of Economic History
University of Oxford

Pontus Stenetorp
Professor of Natural Language Processing
University College London

Andreas Schilling
Professor of Computer Vision
University of Tübingen

Jörg Baten
Professor of Economic History
University of Tübingen

Giulia Caprini
Fellow at Nuffield College
University of Oxford

Martin Biewen
Professor of Statistics
University of Tübingen