

Education

Free University of Amsterdam

2021 - 2024

PhD thesis: "Language Models as Measurement Tools: Using Instruction-Based Models to Increase Validity, Robustness and Data Efficiency" ([PDF](#))

Sciences Po Paris

Free University of Berlin

2016 - 2018

Double Master: political economy, tech policy, quantitative & qualitative methods
Average Grade: 1,2 (top 10%)

University of Augsburg, Germany

2011 - 2015

BA in quantitative social sciences

Average Grade: 1,34 (top 10%)

American University, Washington DC

2013, fall semester

Average Grade: A

Notable Distinctions

German National Academic Foundation (Studienstiftung des deutschen Volkes)

Fellow of Germany's biggest scholarship for excellent and voluntarily active students

Heinrich Böll Stiftung

PhD fully funded by foundation for excellent and engaged students

Languages, Computer Skills

Languages: *German* (mother tongue), *English* (fluent), *French* (fluent)

Computer skills:

- Programming languages:
 - Python: advanced
 - R: intermediate
 - JavaScript: beginner-intermediate
- PyTorch, Hugging Face libraries, sklearn, LLM orchestration libraries etc.
- Git, GitHub, AWS, Docker etc.
- VSCode, Cursor, Jupyter
- StackOverflow [reputation](#): 3200+



Work Experience

🤗 Hugging Face

Dec. 2023 - present

Machine Learning Engineer

- Helping customers build useful things with AI (LLMs, VLMs, text/image classifiers, text/image embeddings) in different industries (finance, health, manufacturing, visual media, retail, public sector) for different use-cases (RAG, code agents, complex LLM systems, text/image search) along the full ML life cycle (use-case definition, data annotation, fine-tuning, evaluation, inference optimization, deployment)
- Training +100 models that have been downloaded +100 million times ([HF profile](#))
- Servicing multiple contracts with a total value of \$1.7 million
- Open-source software: [prompt-templates](#); [zeroshot-classifier](#); [LitServe-HF-containers](#)
- Communicating on social media ([LinkedIn](#), [X](#)); [blog post](#) on synthetic data

Freelance NLP Consultant

Sept. 2022 - Dec. 2023

Part-time, alongside PhD

- Consulted clients on building Retrieval Augmented Generation systems (RAG) (clients were for example [arcee.ai](#), [software-brauerei.ch](#), or [CEPS](#))
- Organised the [European Policy Data Science Network](#)
- Acquisition of clients and public communication of new developments in ML

Free University of Amsterdam

2021 - 2024

PhD Candidate

- PhD Thesis: "Language Models as Measurement Tools: Using Instruction-Based Models to Increase Validity, Robustness and Data Efficiency" ([manuscript PDF](#))
- Showed how instruction-based NLI text classifiers perform better with less data and produce more valid and less biased results than other models in a variety of settings
- Released 0-shot classifiers that were downloaded millions of times per month ([collection](#))
- Two of my models are among the top 10 on the HF hub [ranked by IBM](#)
- Implemented multi-GPU training using the Dutch supercomputer Snellius
- Manuscript finished Dec. 2023, successful defence Oct. 2024

Data Science Lab, Hertie School Berlin

Jan. 2023 - Dec. 2023

Visiting PhD Candidate

- Teaching: e.g. ["Hands-on Transformers"](#) workshop

Centre for European Policy Studies (CEPS), Brussels

June 2019 – Sept. 2022

Data Science Consultant, Researcher (June 2020 – Sept. 2022)

Data Science Consultant, Research Assistant (June 2019 – June 2020)

- Built solutions for and consulted EU institutions with machine learning tools. E.g. analysed citizen consultations using SBERT for EU impact assessments
- Implemented and coordinated large grants and won tenders using NLP methods, e.g. NER, natural language inference, part-of-speech tagging, semantic clustering
- Deployed an interactive dashboard on AWS using Elastic Beanstalk
- Research on socio-economic impacts of new technologies. E.g. contributed to [cost assessment of EU AI Act](#) for European Commission; lead for [FlexiGroBots](#) project

CARSA International, Brussels

Oct. 2018 – June 2019

Consultant, Technology & Innovation

- Projects included: impact assessment of Industry 4.0 applications through surveys & interviews; evaluation of Digital Innovation Hubs; impact assessment of privacy law

Open-Source Contributions (Selection)

Deep Learning Models

- Open-sourced +100 models, especially for English and Multilingual 0-shot classification. The models have been downloaded +100 million times as of January 2025 ([Hugging Face Profile](#), [TCTF leaderboard](#))
- [Benchmark from IBM](#): two of my models rank among the top 10 of models on the HF hub for downstream classification fine-tuning.

Datasets

- [Multilingual-NLI](#): +2.7 million NLI text pairs in 26 languages spoken by +4 billion people. For multilingual 0-shot classification.
- [synthetic zeroshot mixtral](#): A synthetically generated dataset for training more general 0-shot classifiers.
- [CEPS EurLex Dataset](#): Collected, cleaned and published dataset of 142.036 EU legal acts.

Open-source software

- [prompt-templates](#): A library for modularizing prompt templates in standardized .yaml files and for sharing templates on the HF Hub
- [litserve-huggingface-ie](#): A guide for creating LitServe containers for serving any ML model on HF Inference Endpoints.
- [GPT-google-sheets](#): Custom functions for using ChatGPT in Google Sheets without programming knowledge.
- [zeroshot-classifier](#): An easy-to-reproduce pipeline for reproducing my popular 0-shot classifiers.
- Contributions to various repositories

Other Activities (Selection)

DataMine Europe ASBL (formerly ‘European Elections Stats’)

June 2018 - 2021

Co-Founder

- Developed an automatic seat projection for the European Parliament elections 2019 based on national polls from all 28 member states. Used by leading [French](#), [German](#), [Spanish](#), [U.S.](#) media.
- Created a Twitter Bot, using the BART model to summarise international news on Covid-19 (@NewsMiner_Covid).
- Developed a prototype text mining software analysing 2000+ daily news articles from 25 newspapers in 6 EU member states.

Publications, Conference Papers (Selection)

Laurer, Moritz, van Atteveldt, W., Casas, A., & Welbers, K. (2024). On Measurement Validity and Language Models: Increasing Validity and Decreasing Bias with Instructions. *Communication Methods and Measures*, 1–17. <https://doi.org/10.1080/19312458.2024.2378690>

Laurer, Moritz, van Atteveldt, W., Casas, A., & Welbers, K. (2023). Building Efficient Universal Classifiers with Natural Language Inference (arXiv:2312.17543). arXiv. <https://doi.org/10.48550/arXiv.2312.17543>

Laurer, Moritz, Van Atteveldt, W., Casas, A. S., & Welbers, K. (2023). Less Annotating, More Classifying – Addressing the Data Scarcity Issue of Supervised Machine Learning with Deep Transfer Learning and BERT-NLI. *Political Analysis, Cambridge University Press*. 1-33. doi:10.1017/pan.2023.20

Laurer, Moritz, Van Atteveldt, W., Casas, A., & Welbers, K. (2023). Lowering the Language Barrier: Investigating Deep Transfer Learning and Machine Translation for Multilingual Analyses of Political Texts. *Computational Communication Research*, 5(2), 1. <https://doi.org/10.5117/CCR2023.2.7.LAUR>

Borrett, Camille; Laurer, Moritz 2020: “The new ‘EurLex dataset’ of 142.036 EU legal acts and an illustrative use case”. PolMeth 2021 poster presentation. Received “PolMeth Europe Best Poster Award for Open Science”.

Laurer, Moritz 2020: “Measuring Agency with Natural Language Processing”. COMPTTEXT 2020 conference paper.

Borrett, Camille; Laurer, Moritz 2020: “A Computational Analysis of Emerging Actors in 92.215 EU Regulations”. COMPTTEXT 2020 conference paper.

Borrett, Camille; Laurer, Moritz 2020: “The CEPS EurLex dataset: 142.036 EU laws from 1952-2019 with full text and 22 variables”. Harvard Dataverse, V2. Available online: <https://doi.org/10.7910/DVN/0EGYWY>

Laurer, Moritz; Seidl, Timo 2020: “Regulating the European Data-Driven Economy: A Case Study on the General Data Protection Regulation”. In: Policy & Internet. <https://doi.org/10.1002/poi3.246>

Renda, Andrea; Laurer, Moritz 2020: “IoT 4 SDGs – What can the Digital Transformation and IoT achieve for Agenda 2030?”. CEPS study, Brussels. Available online: <https://www.ceps.eu/ceps-publications/iot-4-sdgs/>

Renda, Andrea; Reynolds, Nicole; Laurer, Moritz; Cohen, Gal 2020: “Digitising Agrifood – Pathways and Challenges.” CEPS study, Brussels. Available online: <https://www.ceps.eu/ceps-publications/digitising-agrifood/>