

David Pohl

(+81) 70-8428-5072 | de.david.pohl@gmail.com | linkedin.com/in/pohl-david | github.com/david-pohl

TECHNICAL SKILLS

Languages	Python, JavaScript, Java, SQL
Libraries	Keras, Tensorflow, NumPy, Pandas, Hugging Face, FastAPI, React.js, Selenium, DSPy
Tools	Docker, Git, Ansible, Concourse, OpenAI/Cohere API

EDUCATION

Institute of Science Tokyo - M.Sc. Computer Science - GPA: 97/100 • Research on LLM tokenization under Prof. Naoaki Okazaki	Apr 2024 – Mar 2026 <i>Tokyo, JP</i>
Heidelberg University - B.Sc. Computer Science - GPA: 1.0/1.0 (with distinction) • Advanced coursework in NLP, Information Retrieval, and Symbolic AI	Oct 2020 – Sep 2023 <i>Heidelberg, DE</i>
North Carolina State University - Visiting Student - GPA: 4.0/4.0 Schiller-Gymnasium Cologne - Abitur (A-level) - GPA: 1.1/1.0 (valedictorian)	<i>Raleigh, US / 2022</i> <i>Cologne, DE / 2020</i>

EXPERIENCE

Software Engineering Intern - Woven by Toyota, Inc. • Developing automated checks for software requirements using LLMs • Building a browser extension for assisted technical writing based on internal standards	Jun 2025 – Aug 2025 <i>Tokyo, JP</i>
Strategy Consulting Intern - Boston Consulting Group • Conceptualized strategies for establishing a new risk division for a German bank • Initiated a corporate-wide document hierarchy for credit risk and compliance	Aug 2024 – Oct 2024 <i>Cologne, DE</i>
Research Intern - University of Tokyo • Developed a novel, competition-aware news recommendation technique with Nikkei newspaper	Oct 2023 – Mar 2024 <i>Tokyo, JP</i>
Software Engineering Intern - Deutsche Telekom • Automatized DevOps workflows by building pipelines with Ansible, Concourse, and Docker	Aug 2021 – Feb 2022 <i>Dresden, DE</i>

PUBLICATIONS

Decoding-Free Sampling Strategies for LLM Marginalization • Proposed a novel automata-based sampling method for efficient marginalization • Achieved orders-of-magnitude speedups over model-based sampling at an improved accuracy • <i>David Pohl</i> , Marco Cognetta, Junyoung Lee, Naoaki Okazaki. (under review)
Pitfalls, Subtleties, and Techniques in Automata-Based Subword-Level Constrained Generation • Addressed key challenges in automata-based constrained generation for subword-level LLMs • Proposed a unified pipeline ensuring reliability, extensibility, and testability • Marco Cognetta*, <i>David Pohl</i> *, Junyoung Lee, Naoaki Okazaki. TokShop at ICML 2025 (shared first authorship)
Online DATEing: A Web Interface for Temporal Annotations • Developed and deployed a web interface and API for easy access to temporal tagging models • Dennis Aumiller, Satya Almasian, <i>David Pohl</i> , Michael Gertz. SIGIR 2022

PROJECTS

(Ongoing) Sampling Strategies For LLMs During Constrained Generation • Developing a structured Python library for token-level constrained generation in LLMs • Integrating Monte Carlo methods to normalize tokenization subspace and marginalize over string surface forms
Timothy - Your Personal LLM-Driven Shop Assistant • Built a full-stack application with a modern chat frontend for conversational assistance in e-commerce • Automated the pipeline of web scraping, data cleaning, and storage with Selenium and MySQL
Zero-Shot Word Sense Disambiguation using Word Embeddings • Surveyed embedding-based approaches to Word Sense Disambiguation • Designed and evaluated new techniques for leveraging pre-trained word embeddings

OTHER

Honors	Scholar of German Academic Scholarship Foundation (www.studienstiftung.de) DAAD Full Scholarship for my Master's in Japan (www.daad.de) <i>Previous:</i> Germany Scholarship, Baden-Wuerttemberg Scholarship (merit-based, governmental)
Languages	English (fluent), German (native), Japanese (basic)