

Christoph Alt

RESEARCHER · DATA SCIENTIST

☎ (+49) xxxx xxx xxx x | ✉ xxx@xxx.de | 💻 ChristophAlt | 🌐 ChristophAlt | 📍 Berlin, GERMANY

Summary

I am a **researcher / software engineer** with 4+ years of experience (research and applied) in **machine learning & natural language processing & distributed computing**. Fluent in Python, Java, Python's scientific and machine learning stack, recent deep learning and natural language processing frameworks, Docker, and AWS. I am passionate about solving difficult problems with little data.

Experience



German Research Center for AI (DFKI)

Berlin

RESEARCHER AND PHD STUDENT

2017–(now)

- (Applied) research in **information extraction** with limited labeled data, **explainability**, and **dialogue systems**.
- I published and presented research at multiple top-tier conferences (e.g., ACL, EMNLP).
- I developed a dialogue system prototype that answers tens of millions of customer support requests per year.
- I developed an information extraction system that extracts thousands of supply-chain-related events (e.g., acquisitions) per day.



Zalando SE

DATA SCIENTIST

Berlin

2015–2017

- As part of the **query understanding and intent** team my goal was to improve customer search experience.
- I was partly responsible for architecture and system design during the transition towards a more intelligent search.
- I developed a query **autocomplete and autosuggest** system to assist customers with their search, which, by now, serves tens of millions of customers every month.



Technical University of Berlin

Berlin

RESEARCH ASSISTANT

2016–2017

- As a member of the Complex and Distributed IT Systems Group (CIT), I contributed to the design and development of "Exa", an experimental large scale **distributed machine learning** framework.
- I further implemented various distributed training strategies for commonly used machine learning algorithms.



Technical University of Berlin

Berlin

RESEARCH ASSISTANT

2015–2016

- As a member of the Database Systems and Information Management Group (DIMA), I **contributed algorithms** to the **Apache Flink** machine learning library and created content for the course "Scalable Data Mining, Concepts and Methods".
- I further developed a data processing pipeline to analyze large web corpora (e.g., Common Crawl) with Apache Spark and Flink.

Education



Technical University of Berlin

Berlin

DR.-ING. IN COMPUTER SCIENCE, MACHINE LEARNING 2017–(Dec 2020)

- **PhD thesis (submitted):** "Neural Sequential Transfer Learning for Relation Extraction"
- **Supervisors:** Prof. Sebastian Möller, Prof. Hans Uszkoreit
- I explored neural **transfer learning** for **information extraction**, specifically relation extraction from text (RE).
- I improved upon state-of-the-art RE methods, and identified shortcomings of recent approaches and remaining challenges of the task.



Technical University of Berlin

Berlin

MSC IN COMPUTER ENGINEERING

2014–2017

- **Master thesis:** "Adaptive Consistency Management for Distributed Machine Learning"
- **Supervisors:** Prof. Odej Kao, Prof. Volker Markl
- I explored different approaches to speed up distributed training of commonly used machine learning algorithms.
- Courses on general **machine learning**, **scalable data mining**, **natural language processing**, and **computer vision**.



Munich University of Applied Sciences

Munich

BENG IN COMPUTER ENGINEERING

2009–2013

- Courses on **embedded systems** development, **digital image processing**, and **network architectures**.

Projects

- RelEx** A comprehensive suite of baseline and state-of-the-art neural relation extraction methods; based on [AllenNLP](#). [🔗](#)
- REval** A framework to develop and evaluate probing tasks, e.g., to diagnose linguistic features captured by neural relation extraction models; based on the [SentEval toolkit](#). [🔗](#)
- tuna** Distributed hyperparameter search for AllenNLP; based on [Ray](#). The core framework now powers AllenAI's [allentune](#). [🔗](#)

Skills

- Programming Libraries** Python (6 years), Java (3 years), Rust (1 year), SQL PyTorch, TensorFlow, JAX, numpy, scikit-learn, pandas, AllenNLP, Spacy, StanfordNLP, etc.
- Tools** Docker, Elasticsearch, AWS, Apache Spark & Flink & Hadoop & Kafka & Lucene & Solr
- Languages** German, English, French

Publications

12. **Christoph Alt**, Aleksandra Gabryszak and Leonhard Hennig, “TACRED Revisited: A Thorough Evaluation of the TACRED Relation Extraction Task”, *58th Annual Meeting of the Association for Computational Linguistics*, 1558–1569, 2020. <https://www.aclweb.org/anthology/2020.acl-main.142>
11. **Christoph Alt**, Aleksandra Gabryszak and Leonhard Hennig, “Probing Linguistic Features of Sentence-level Representations in Neural Relation Extraction”, *58th Annual Meeting of the Association for Computational Linguistics*, 1534–1545, 2020. <https://www.aclweb.org/anthology/2020.acl-main.140>
10. David Harbecke and **Christoph Alt**, “Considering Likelihood in NLP Classification Explanations with Occlusion and Language Modeling”, *58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, 111–117, 2020. <https://www.aclweb.org/anthology/2020.acl-srw.16>
9. Hanchu Zhang, Leonhard Hennig, **Christoph Alt**, Changjian Hu, Yao Meng and Chao Wang, “Bootstrapping Named Entity Recognition in E-Commerce with Positive Unlabeled Learning”, *The 3rd Workshop on e-Commerce and NLP*, 1–6, 2020. <https://www.aclweb.org/anthology/2020.ecnlp-1.1>
8. **Christoph Alt**, Marc Hübner and Leonhard Hennig, “Fine-tuning Pre-Trained Transformer Language Models to Distantly Supervised Relation Extraction”, *57th Annual Meeting of the Association for Computational Linguistics*, 1388–1398, 2020. <https://www.aclweb.org/anthology/P19-1134>
7. **Christoph Alt***, Marc Hübner* and Leonhard Hennig, “Improving Relation Extraction by Pre-trained Language Representations”, *2019 Conference on Automated Knowledge Base Construction*, 1–18, 2019. <https://openreview.net/forum?id=BJgrxbqp67>
6. Robert Schwarzenberg, Marc Hübner, David Harbecke, **Christoph Alt** and Leonhard Hennig, “Layerwise Relevance Visualization in Convolutional Text Graph Classifiers”, *13th Workshop on Graph-Based Methods for Natural Language Processing*, 58–62, 2019. <https://www.aclweb.org/anthology/D19-5308>
5. Yang Li, Qingliang Miao, Ji Geng, **Christoph Alt**, Robert Schwarzenberg, Leonhard Hennig, Changjian Hu and Feiyu Xu, “Question answering for technical customer support”, *International Conference on Natural Language Processing and Chinese Computing, Information*, 10, 63, 2019.
4. Guoguang Zhao, Jianyu Zhao, Yang Li, **Christoph Alt**, Robert Schwarzenberg, Leonhard Hennig, Stefan Schaffer, Sven Schmeier, Changjian Hu and Feiyu Xu, “MOLI: Smart Conversation Agent for Mobile Customer Service”, *Information*, 10, 2, 2018.
3. David Harbecke, Robert Schwarzenberg and **Christoph Alt**, “Learning Explanations from Language Data”, *2018 EMNLP Workshop Black-boxNLP: Analyzing and Interpreting Neural Networks for NLP*, 316–318, 2018. <https://www.aclweb.org/anthology/W18-5434>
2. Roland Roller, **Christoph Alt**, Laura Seiffe and He Wang, “mEx - An Information Extraction Platform for German Medical Text”, *11th International Conference on Semantic Web Applications and Tools for Healthcare and Life Sciences*
1. Tobias Herb, Tim Jungnickel and **Christoph Alt**, “Weak Consistency and Stochastic Environments: Harmonization of Replicated Machine Learning Models”, *2nd Workshop on the Principles and Practice of Consistency for Distributed Data*, 8, 3, 2016. <https://doi.org/10.1145/2911151.2911161>