

Education

- 2021–2024 **PhD, Computer Science**, *University of Copenhagen*
Research Topics: Natural Language Processing, Machine Learning
Advisor: Prof. Anders Søgaard
- 2018–2020 **M.Sc., Computer Science**, *TU Darmstadt*, *Cumulative GPA: 1.34**
Focus: Natural Language Processing, Machine Learning, Artificial Intelligence
Thesis Supervisors: Jonas Pfeiffer, Prof. Iryna Gurevych
- 2015–2018 **B.Sc., Computer Science**, *DHBW Stuttgart*, *Cumulative GPA: 1.4* (ECTS Grade A)*

*On a scale of 1-5, 1.0 being the best possible grade

Publications and Manuscripts

- [1] **Phillip Rust**, Bowen Shi, Skyler Wang, Necati Cihan Camgöz, and Jean Maillard. Towards Privacy-Aware Sign Language Translation at Scale. *arXiv preprint*, 2024.
- [2] Nadav Borenstein, **Phillip Rust**, Desmond Elliott, and Isabelle Augenstein. PHD: Pixel-Based Language Modeling of Historical Documents. In *EMNLP*, 2023.
- [3] Jonas Lotz, Elizabeth Salesky, **Phillip Rust**, and Desmond Elliott. Text Rendering Strategies for Pixel Language Models. In *EMNLP*, 2023.
- [4] **Phillip Rust**, Jonas F. Lotz, Emanuele Bugliarello, Elizabeth Salesky, Miryam de Lhoneux, and Desmond Elliott. Language Modelling with Pixels. In *ICLR (notable-top-5%)*, 2023.
- [5] **Phillip Rust** and Anders Søgaard. Differential Privacy, Linguistic Fairness, and Training Data Influence: Impossibility and Possibility Theorems for Multilingual Language Models. In *ICML*, 2023.
- [6] Constanza Fierro*, Laura Cabello Piqueras*, Jonas F. Lotz*, **Phillip Rust***, Joen Rommedahl, Jeppe Klok Due, Christian Igel, Desmond Elliott, Carsten Bøcker Pedersen, Israfil Salazar, and Anders Søgaard. Date Recognition in Historical Parish Records. In *Frontiers in Handwriting Recognition (ICFHR)*, 2022.
- [7] Daniel Hershcovich, Stella Frank, Heather Lent, Miryam de Lhoneux, Mostafa Abdou, Stephanie Brandl, Emanuele Bugliarello, Laura Cabello Piqueras, Ilias Chalkidis, Ruixiang Cui, Constanza Fierro, Katerina Margatina, **Phillip Rust**, and Anders Søgaard. Challenges and Strategies in Cross-Cultural NLP. In *ACL*, 2022.
- [8] **Phillip Rust***, Jonas Pfeiffer*, Ivan Vulić, Sebastian Ruder, and Iryna Gurevych. How Good is Your Tokenizer? On the Monolingual Performance of Multilingual Language Models. In *ACL*, 2021.

- [9] Gözde Gül Şahin, Yova Kementchedjieva, **Phillip Rust**, and Iryna Gurevych. PuzzLing Machines: A Challenge on Learning From Small Data. In *ACL*, 2020.

Work Experience

- 05/2021– **PhD Fellow**, *University of Copenhagen*, Copenhagen, Denmark
Ongoing Doctoral research in natural language processing.
- 06/2023– **Research Scientist Intern**, *Fundamental AI Research (FAIR), Meta*, Menlo Park, USA
12/2023 Research internship with Jean Maillard on self-supervised learning for sign language translation.
- 10/2018– **Technical Consultant – DualStudy Master**, *DXC Technology*, Frankfurt, Germany
03/2021 Technical consulting and software development on petabyte-scale data and compute platforms (OpenShift, MapR) for client in the area of autonomous driving.
- 10/2015– **Intern – DualStudy Bachelor**, *Hewlett Packard Enterprise*, Bad Homburg, Germany
09/2018 Integrated program that combined studies in computer science at DHBW Stuttgart with six internships in various departments and locations at Hewlett Packard Enterprise and DXC Technology.

Summer/Winter Schools

- 2021 **Lisbon Machine Learning Summer School (LxMLS)**, *Virtual*
2021 **Advanced Language Processing Winter School (ALPS)**, *Virtual*
2023 **HPLT & NLPL Winter School**, *Skeikampen, Norway*

Invited Talks

- 08/2022 **Amazon AWS AI (Tel Aviv)**, *Language Modelling with Pixels*, Virtual
09/2022 **Pioneer Centre for AI**, *Language Modelling with Pixels*, Copenhagen, Denmark
01/2023 **Amazon AWS AI (New York)**, *Language Modelling with Pixels*, Virtual
01/2023 **COLT Group, UPF Barcelona**, *Language Modelling with Pixels*, Virtual
03/2023 **ML Collective Reading Group (NLP & CV)**, *Language Modelling with Pixels*, Virtual

Grants

- 2023 **EuroHPC Regular Access**, “Pixel Encoding of Language”, with Desmond Elliott

Service & Professional Activities

Reviewing: ARR, ACL, EMNLP, EACL, MRL @ EMNLP, TiE @ ECCV, TL4NLP @ NeurIPS

Program Committee: MRL @ EMNLP, TL4NLP @ NeurIPS

Organization: Weekly NLP Section Reading Group, University of Copenhagen

Teaching Assistance: Advanced Topics in NLP, Introduction to Data Science

Technical Skills

Programming: Python, Java, JavaScript

ML & NLP Frameworks: PyTorch, PyTorch XLA, HuggingFace Ecosystem, DeepSpeed, NumPy

Miscellaneous: Unix, Slurm, MapR, Kubernetes, Docker, Airflow, Elastic Stack, Ansible

Languages

German (native), English (fluent, TOEFL iBT 112), Japanese (elementary), Danish (elementary)