Philip John Gorinski, PhD

Curriculum Vitae

32A Grove Terrace
London NW5 1PL
United Kingdom
+44 75 0220 4046
p.j.gorinski@gmail.com

Experience & Qualifications

Vocational

01/2019 - ongoing	**Senior Research Scientist NLP**, Huawei Noah's Ark Lab, London.
	 As a research group, we drive Huawei's expertise in Natural Language Processing.
	 The role comprises key duties of
	   - Strengthening Huawei's contributions to and recognition by the research community through impactful publications in excellent venues.
	   - Fostering collaborations with universities and other research institutes.
	   - Contributing to the company's practical portfolio with inventions and patents.
	   - Delivering prototypes to internal customers based on cutting-edge research.
	   - Research into as well as development and application of Foundation Models, and into tuning methods for pretrained Large Language Models.

01/2018 - 01/2019	**Postdoc Researcher**, University of Edinburgh, School of Informatics ILCC. Work on NLP for Health Applications, with NHS Scotland. Developed EdIE-N Named Entity Recognition system for electronic health records, based on bi-directional LSTM and Conditional Random Fields.
06/2015 - 09/2015	**Research Intern**, Microsoft Research, Redmond, WA. Research in the Natural Language Processing group. Work on Text-to-Code / Intent-to-Code systems.
04/2013 - 09/2013	**Student Freelancer**, DIaLOGIKa Software GmbH, Saarbrücken. Software consulting and testing.
2007 - 2012	**Student Research Assistant**, Department of Computational Linguistics, Cluster of Excellence, Saarland University. "SemEval": Research on FrameNet Null Instantiations, programming work. "IDIX": Research on idioms in context, programming work and annotation. "SALSA": Semantic annotation of German newspaper texts, using German FrameNet.

Education

09/2013 - 11/2017	**PhD Program**, University of Edinburgh, School of Informatics ILCC. Program "Natural Language Processing". Submitted 11/2017; Defended 03/2018; PhD awarded 07/2018.
03/2011 - 08/2013	**M.Sc. Program**, Saarland University, Sarbrücken, Germany. Program "Language Science and Technology". Final grade: 1.4 (excellent).
09/2006 - 03/2011	**B.Sc. Program**, Saarland University, Saarbrücken, Germany. Program "Computational Linguistics". Final grade: 1.5 (excellent).
1998 - 2006	**Abitur**, Paul-Klee Gymnasium, Overath, Germany. Final grade: 2.0 (good).

Mentoring & Supervising

Internships	I supervised several interns on various research projects in Huawei Noah's Ark Lab, London. Duration ranged from 3 to 9 months, during which I was able to mentor and provide guidance in good research practices, project fulfilment, and scientific writing. Outcomes include peer-reviewed papers and patents.
M.Sc. Projects	Our team supervises University College of London M.Sc. projects yearly since 2019, as industry partners of UCL. All projects under my supervision were completed with honours. Outcomes for the team include recruiting students for follow-up internships, and support with writing follow-up papers based on projects.

Programming Languages, Tools, and Libraries

Python	**Very good**. Courses at Saarland University; frameworks for projects "Idix" and "SemEval"; majority of second half of PhD work; postdoc projects; primary language for projects in Huawei Noah's Ark NLP team.
pyTorch	**Very Good**. Used for all projects since PhD, from versions < 1.0 to 2.x; includes extensive use of DeepSpeed; extensive use and knowledge of HuggingFace libraries, and other relevant frameworks including Scikit Learn, NLTK.
Java	**Very good**. Courses at Saarland University; frameworks for projects "Idix" and "SemEval"; software projects on spoken dialogue systems; bimorphism-framework for bachelor's thesis; TAG-based machine translation framework for master's thesis; majority of first half of PhD programming work.
Core Skills	Machine Learning/Deep Learning for NLP: Focus on LLM training and fine-tuning; additional experience with Graph Neural Networks, Computer Vision and Speech models; Deep Reinforcement Learning for NLP applications. Algorithms & data structures.
Others	Experience with Perl (Good), Matlab (Basic), SML (Basic), C/++ (Very Basic); Huawei S3 (Similar Amazon S3); Huawei ROMA Cloud (Similar Amazon AWS)

Languages

German	**Native**.
English	**Fluent**. 8 years of education in secondary school including 2 years of advanced courses; stays abroad; Master's degree taught in English; PhD program and work in the UK.
French	**Basic**. 3 1/2 years of education in secondary school, with stays abroad.
Chinese	**Basic**. Business Chinese course by LSE at Huawei Noah's Ark; Courses Beginners 1+2, Elementary 1+2 by Practical Mandarin; language classes during 6 months deployment to Huawei campus in Dongguan, China.

Publications

Under Review	Mojtaba Valizadeh, **Philip John Gorinski**, Ignacio Iacobacci, Martin Berger. "The Regular Expression Inference Challenge". *Under review for AAAI 2024*.

Peer Reviewed	**Philip John Gorinski**, Matthieu Zimmer, Gerasimos Lampouras, Derrick Goh Xin Deik, Ignacio Iacobacci. "Automatic Unit Test Data Generation and Actor-Critic Reinforcement Learning for Code Synthesis". *Findings of EMNLP 2023*.

Alexander I. Cowen-Rivers, **Philip John Gorinski**, Aivar Sootla, Asif Khan, Liu Furui, Jun Wang, Jan Peters, Haitham Bou Ammar. "Structured Q-learning For Antibody Design". *RL4RealLife Workshop, NeurIPS 2022.*

Ieva Staliūnaitė, **Philip John Gorinski**, Ignacio Iacobacci. "Improving Commonsense Causal Reasoning by Adversarial Training and Data Augmentation." *Proceedings of AAAI 2021.*

Yusheng Tian, **Philip John Gorinski**. "Improving End-to-End Speech-to-Intent Classification with Reptile". *Proceedings of InterSpeech 2020.*

Gabriel Gordon-Hall, **Philip John Gorinski**, Shay B. Cohen. "Learning Dialog Policies from Weak Demonstrations". *Proceedings of ACL 2020.*

Gabriel Gordon-Hall, **Philip John Gorinski**, Gerasimos Lampouras, Ignacio Iacobacci. "Show Us the Way: Learning to Manage Dialog from Demonstrations". *DSTC8 Workshop at AAAI 2020.*

Philip John Gorinski, Honghan Wu, Claire Grover, Richard Tobin, Conn Talbot, Heather Whalley, Cathie Sudlow, William Whiteley, Beatrice Alex. "Named Entity Recognition for Electronic Health Records: A Comparison of Rule-based and Machine Learning Approaches". *HealTAC 2019.*

Philip John Gorinski and Mirella Lapata. "What's this Movie about? A Joint Neural Network Architecture for Movie Content Analysis". *Proceedings of NAACL-HLT 2018.*

Philip John Gorinski and Mirella Lapata. "Movie Script Summarization as Graph-based Scene Extraction". *Proceedings of NAACL-HLT 2015.*

Philip John Gorinski, Josef Ruppenhofer and Caroline Sporleder. "Towards Weakly Supervised Resolution of Null Instantiations". *Proceedings of IWCS 2013.*

Vera Demberg, Asad B. Sayeed, **Philip John Gorinski** and Nikolaos Engonopoulos. "Syntactic surprisal affects spoken word duration in conversational contexts". *Proceedings of AMLaP 2012.*

Josef Ruppenhofer, Caroline Sporleder and **Philip John Gorinski**. "In search of missing arguments: A linguistic approach". *Proceedings of RANLP 2011.*

Caroline Sporleder, Linlin Li, **Philip Gorinski** and Xaver Koch. "Idioms in Context: The IDIX Corpus". *Proceedings of LREC 2010.*

Preprints

Yunjie He, **Philip John Gorinski**, Ieva Staliunaite, Pontus Stenetorp. "Graph Attention with Hierarchies for Multi-hop Question Answering". *arXiv:2301.11792*

Ieva Staliūnaitė, **Philip John Gorinski**, Ignacio Iacobacci. "Relational Graph Convolutional Neural Networks for Multihop Reasoning: A Comparative Study". *arXiv:2210.06418*

Patents

"Code Synthesis through Reinforcement Learning and automatically generated Unit Tests". Filing Date 03/2023.

"Variable Importance Learning for Multi-task NLU Training". Filing Date 09/2020.

PhD Thesis

"Automatic Movie Analysis and Summarisation"
supervisors: Prof. Mirella Lapata, Dr. Rik Sarkar

Key contributions: (i) Collected and released ScriptBase Corpus, a new large-scale data set of movie scripts, annotated with additional meta-information. (ii) Developed a dynamic summarisation model for the screenplay domain, for extraction of highly informative and important scenes from scripts. (iii) Extended the summarisation model to capture additional modalities, introducing visual information obtained from the movie, allowing users to generate visual summaries of motion pictures. (iv) Devised a novel end-to-end neural network model for generating natural language screenplay overviews.

Master's Thesis	**"Rule Extraction for High Quality Tree-to-Tree Machine Translation"** graded 1.8 (good), supervisors: Dr. Yi Zhang, Dr. Rui Wang, Prof. Hans Uszkoreit. Investigated the application of synchronous Tree Adjoining Grammar for High Quality MT. This strand of MT is of special interested where it is important to have perfect translations, e.g. of legal and other documents the European Union parliament.
Bachelor's Thesis	**"A Bimorphism Approach to Training Tree Transducers"** graded 1.3 (excellent), supervisors: Prof. Alexander Koller, Prof. Manfred Pinkal. Investigated the treatment of probabilistic Tree Transducers as Bimorphisms, as well as how they can be automatically trained. Possible fields of application include, amongst others, Machine Translation and Syntax-Semantics-Interfaces.
B.Sc. Project	**"Spoken Dialogue Systems"** graded 1.3 (excellent), lecturers: Prof. Dietrich Klakow, Magdalena Wolska. Study and model strategies of COMMUNICATOR dialogue system for flight booking. Used Reinforcement Learning to learn an optimal dialogue strategy.