

Marek Rei

Email: marek@marekrei.com

Homepage: www.marekrei.com

Experience

Senior Research Associate, University of Cambridge

July 2016 - current

My research is on machine learning models and advanced neural network architectures, with a focus on structured prediction, language modeling and multi-objective optimization. I supervise MPhil and undergraduate students for their course projects, and teach as part of the MPhil course on Advanced Computer Science.

Scientific Advisor and Co-Founder, Transformative AI

March 2017 - current

Transformative AI is creating predictive healthcare technologies, focusing on the early detection and prevention of ventricular tachyarrhythmias in clinical care patients. I am a co-founder and a scientific advisor for the development of machine learning algorithms that power the predictive systems.

Research Associate, University of Cambridge

November 2014 - July 2016

Performed research on machine learning algorithms for automated language analysis. Among other things, I developed a general-purpose neural structured prediction framework, and an online representation learning algorithm for improved language modeling.

Visiting Lecturer, University of Tartu

March 2015 - June 2015

As a visiting lecturer, I created and delivered a new course on machine learning and language modeling. The lectures cover topics such as n-gram language modeling, smoothing techniques, language model evaluation, neural network language models, recurrent models, backpropagation, gradient descent, representation learning, and an overview of current state-of-the-art methods for language modeling.

Language Research Engineer, SwiftKey

November 2012 - November 2014

I was a member of the Research team, working on future technologies and prototyping new solutions. One of my main projects was the neural network version of SwiftKey, which was released to very positive reviews. Since then, SwiftKey has been acquired by Microsoft, largely based on the strength of its technology.

Supervision Tutor, University of Cambridge

January 2011 - March 2012

I organized tutoring sessions for undergraduate students in the subject of Information Retrieval. This involved regular tutorials in small groups, and providing them with theoretical and practical assignments.

Public Relations Manager, Board of European Students of Technology

September 2007 - July 2008

Organizing BEST General Assembly 2008 - an international student congress with representatives from 76 universities. Handled relations with the media, companies and universities, and designed various promotional materials. Worked in a core team of 5 people together with over 40 volunteers.

Software Analyst, Webmedia

June 2007 - August 2007

Internship as a software analyst. Worked on enhancing a self-service system for a mobile service provider, designed and performed usability tests and quality control, wrote specifications and created a prototype for a new user interface.

Project manager, Board of European Students of Technology

October 2006 - May 2007

Main organizer of Key to the Future (Võti Tulevikku) 2007, a national career project for graduates in Estonia with 74 participating companies. As the project manager I led a core team of 5 people plus many volunteers, and was involved in every aspect of the project.

Software Developer, Movies.ee

June 2006 - April 2007

Developed a custom-built content management system for managing movies, actors, news, galleries, contests and live cinema timetables.

Software Developer at Geomedia

February 2006 - June 2006

Developed software for conducting large-scale population surveys.

Education

PhD, University of Cambridge

2009 - 2013

My thesis is about minimally supervised dependency-based methods for natural language processing, under the supervision of Prof Ted Briscoe. I worked on topics such as distributional and semantic similarity, speculation detection, entailment detection, and unsupervised parse reranking. I applied various machine learning techniques and developed systems that achieved state-of-the-art performance.

MPhil, University of Cambridge

2008 - 2009

MPhil in Computer Speech, Text and Internet Technologies. Covered topics such as speech recognition, speech synthesis, parsing, discourse, dialogue systems, machine translation and question answering.

BSc, Tallinn University of Technology

2005 - 2008

Bachelor's Degree in Computer Science from the Tallinn University of Technology. Specialised in Network Applications and Logic, graduated *cum laude*.

Organization and Membership

I am a member of the following organizations:

- Area Chair for the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)
- Area Chair for the Eighth Joint Conference on Lexical and Computational Semantics (*SEM 2019)
- Organizer of the Workshop on Representation Learning for NLP (RepL4NLP 2019)
- Organizer of the Workshop on Machine Translation and Natural Language Processing at Makerere University 2018
- Member of the Association for the Advancement of Artificial Intelligence (AAAI)
- Member of the Association for Computational Linguistics (ACL)
- Member of the Special Interest Group on Representation Learning (SigREP)
- Member of the Special Interest Group on NLP and Education (SigEDU)

Program Committee

I am part of the program committee for the following conferences and workshops:

- The 36th International Conference on Machine Learning (ICML 2019)
- The 3rd Workshop on Evaluating Vector Space Representations for NLP (RepEval 2019)
- The Second Workshop on Fact Extraction and Verification (FEVER 2019)
- The Workshop on Structured Prediction for NLP (SPNLP 2019)
- The Workshop on Deep Learning Approaches for Low Resource Natural Language Processing (DeepLo 2019)
- The 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018)
- The SIGNLL Conference on Computational Natural Language Learning (CoNLL 2018)
- The 13th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2018)
- The Seventh Joint Conference on Lexical and Computational Semantics (*SEM 2018)
- The 3rd Workshop on Representation Learning for NLP (RepL4NLP 2018)
- The First Workshop on Fact Extraction and Verification (FEVER 2018)
- The Workshop on Figurative Language Processing (FigLang 2018)
- Workshop on Deep Learning Approaches for Low Resource Natural Language Processing (DeepLo 2018)
- Workshop on Subword and Character Level Models in NLP 2018 (SCLeM 2018)
- Workshop on Figurative Language Processing 2018 (FigLang 2018)
- The Sixth Joint Conference on Lexical and Computational Semantics (*SEM 2017)

The 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2017)
Workshop on Evaluating Vector Space Representations for NLP 2017 (RepEval 2017)
Workshop on Subword and Character LEvel Models in NLP 2017 (SCLeM 2017)
Workshop on Evaluating Vector Space Representations for NLP 2016 (RepEval 2016)
The 11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2016)

Reviewing

I review for the following journals, conferences and workshops:

The International Conference on Machine Learning (ICML)
Transactions of the Association for Computational Linguistics (TACL)
The Annual Conference of the Association for Computational Linguistics (ACL)
The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)
The European Chapter of the Association for Computational Linguistics Conference (EACL)
Conference on Empirical Methods in Natural Language Processing (EMNLP)
The Conference on Computational Linguistics (COLING)
Journal of Natural Language Engineering (JNLE)
Special Issue of the Natural Language Engineering
International Journal of Artificial Intelligence in Education (IJAIED)
The Joint Conference on Lexical and Computational Semantics (*SEM)
The SIGNLL Conference on Computational Natural Language Learning (CoNLL)
The Workshop on Representation Learning for NLP (RepL4NLP)
Workshop on Innovative Use of NLP for Building Educational Applications (BEA)
The Workshop on Fact Extraction and Verification (FEVER)
The Workshop on Figurative Language Processing (FigLang)
The Workshop on Structured Prediction for NLP (SPNLP)
Workshop on Metaphor in NLP
Workshop on Deep Learning Approaches for Low Resource Natural Language Processing (DeepLo)
Workshop on Evaluating Vector Space Representations for NLP (RepEval)
Workshop on Learning to Generate Natural Language (LGNL)
Language Sciences Symposium

Lecturing

The courses I have created and lectured:

Interpreting the Black Box: Explainable Neural Network Models, University of Cambridge, 2019

Data Science: Principles and Practice, University of Cambridge, 2018

Constructing and Evaluating Word Embeddings, University of Cambridge, 2017

Constructing and Evaluating Word Embeddings, University of Cambridge, 2016

Machine Learning for Language Modelling, University of Tartu, 2015

Supervision

Students whom I supervise for their thesis or course project:

Joint Text Classification on Multiple Levels with Multiple Labels

Miruna Pislari

MPhil in Advanced Computer Science, University of Cambridge, to finish in 2019.

Few-Shot Learning in Distributional Semantics

Jeroen Van Hautte

MPhil in Advanced Computer Science, University of Cambridge, to finish in 2019.

Supervising in collaboration with Dr Guy Emerson.

Language Modelling for Neural Error Detection

Samuel Bell

MPhil in Advanced Computer Science, University of Cambridge, 2018.

Neural Architectures for Visual Question Answering

Joshua Wong

Computer Science Tripos, University of Cambridge, 2018.

Supervised Attention for Neural Error Correction

Gladys Tien

MPhil in Advanced Computer Science, University of Cambridge, 2018.

Supervising in collaboration with Dr Zheng Yuan.

Domain Adaptation for Neural Named Entity Recognition

Stefan Hosein

MPhil in Advanced Computer Science, University of Cambridge, 2017.

Hybrid Language Modeling for Text Prediction

Devan Kuleindiren

Computer Science Tripos, University of Cambridge, 2017.

Convolutional Neural Networks for Automated Essay Assessment

Younna Farag

MPhil in Advanced Computer Science, University of Cambridge, 2016.

Predicting Word Difficulty for Lexical Acquisition

Shushan Arakelyan

MPhil in Advanced Computer Science, University of Cambridge, 2016.

Awards, Grants and Scholarships

King's College Junior Research Fellowship, 2017-2021
 Cambridge-Africa Alborada Research award, 2017-2018
 Language Sciences Research Incubator Fund, 2017
 EPSRC Doctoral Training award, 2009-2012
 Cambridge European Trust award, 2009-2012
 Computer Laboratory Local Industry Fund, 2009-2012
 EPSRC Collaborative Training award, 2008
 Cambridge European Trust award, 2008
 AS Elion Scholarship, 2007

Peer-reviewed Publications

Jointly Learning to Label Sentences and Tokens

Marek Rei and Anders Søgaard

In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI 2019).

Advance Prediction of Ventricular Tachyarrhythmias using Patient Metadata and Multi-Task Networks

Marek Rei, Josh Oppenheimer and Marek Sirendi

In Proceedings of the NeurIPS Workshop on Machine Learning for Health (ML4H 2018).

Sequence Classification with Human Attention

Maria Barrett, Joachim Bingel, Nora Hollenstein, Marek Rei and Anders Søgaard

In Proceedings of the SIGNLL Conference on Computational Natural Language Learning (CoNLL 2018).

Special award for the best paper on research inspired by human language learning and processing

Scoring Lexical Entailment with a Supervised Directional Similarity Network

Marek Rei, Daniela Gerz and Ivan Vulić

In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018).

Zero-shot Sequence Labeling: Transferring Knowledge from Sentences to Tokens

Marek Rei and Anders Søgaard

In Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2018).

Variable Typing: Assigning Meaning to Variables in Mathematical Text

Yiannos Stathopoulos, Simon Baker, Marek Rei and Simone Teufel

In Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2018).

A Supervised Similarity Network for Metaphor Detection

Marek Rei, Luana Bulat, Douwe Kiela and Katia Shutova

In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017).

Neural Sequence-Labeling Models for Grammatical Error Correction

Helen Yannakoudakis, Marek Rei, Øistein E. Andersen and Zheng Yuan

*In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017).***Artificial Error Generation with Machine Translation and Syntactic Patterns**

Marek Rei, Mariano Felice, Zheng Yuan and Ted Briscoe

*In Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2017).***Auxiliary Objectives for Neural Error Detection Models**

Marek Rei and Helen Yannakoudakis

*In Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2017).***An Error-Oriented Approach to Word Embedding Pre-Training**

Youmna Hussein, Marek Rei and Ted Briscoe

*In Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2017).***Detecting Off-topic Responses to Visual Prompts**

Marek Rei

*In Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2017).***Semi-supervised Multitask Learning for Sequence Labeling**

Marek Rei

*In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017).***Attending to Characters in Neural Sequence Labeling Models**

Marek Rei, Sampo Pyysalo and Gamal K.O. Crichton

*In Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016).***A Joint Model for Word Embedding and Word Morphology**

Kris Cao and Marek Rei

*In Proceedings of the 1st Workshop on Representation Learning for NLP (RepL4NLP 2016).***Compositional Sequence Labeling Models for Error Detection in Learner Writing**

Marek Rei and Helen Yannakoudakis

*In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016).***Automatic Text Scoring Using Neural Networks**

Dimitrios Alikaniotis, Helen Yannakoudakis and Marek Rei

*In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016).***Sentence Similarity Measures for Fine-Grained Estimation of Topical Relevance in Learner Essays**

Marek Rei and Ronan Cummins

*In Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2016).***Online Representation Learning in Recurrent Neural Language Models**

Marek Rei

*In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015).***Looking for hyponyms in vector space**

Marek Rei and Ted Briscoe

In Proceedings of the Eighteenth Conference on Computational Natural Language Learning (CoNLL 2014).

Parser lexicalisation through self-learning

Marek Rei and Ted Briscoe

*In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2013).***Unsupervised Entailment Detection between Dependency Graph Fragments**

Marek Rei and Ted Briscoe

*In Proceedings of the 2011 Workshop on Biomedical Natural Language Processing (BioNLP 2011).***Combining Manual Rules and Supervised Learning for Hedge Cue and Scope Detection**

Marek Rei and Ted Briscoe

The 14th Conference on Natural Language Learning (CoNLL 2010).

* Between 2012-2014 I was working as a researcher in SwiftKey. Research from that period was not published due to confidentiality restrictions, but a section of it has been made available as a patent.

Book Chapters

Intelligent Information Access from Scientific Papers

Ted Briscoe, Karl Harrison, Andrew Naish-Guzman, Andy Parker, Marek Rei, Advait Siddharthan, David Sinclair, Mark Slater and Rebecca Watson

Current Challenges in Patent Information Retrieval, edited by Mihai Lupu, Katja Mayer, John Tait and Anthony J. Trippe, 2011.

Patents

Patent applications for which I am the named inventor:

Improved Artificial Neural Network for Language Modelling and Prediction

Customised neural network structure for increasing the efficiency of language models, based on my research in SwiftKey.

*London, United Kingdom, 2016***Prediction of Cardiac Events**

Machine learning system for predicting cardiac events in patients and using it to guide their medical care, based on my research for Transformative AI.

London, United Kingdom, 2017

Theses

Minimally Supervised Dependency-based Methods for Natural Language Processing

PhD thesis, University of Cambridge

*Cambridge, United Kingdom, 2013***Adaptive Interactive Information Extraction**

MPhil thesis, University of Cambridge

*Cambridge, United Kingdom, 2009***Audio-visual Speech Synthesis**

Bachelor's thesis, Tallinn University of Technology

Tallinn, Estonia, 2008

Invited Talks

Semi-supervised and Zero-shot Methods for Sequence Labeling

TU Darmstadt, Germany. November 2018

Language Modelling and Machine Translation

University of Makerere, Uganda. June 2018

Human Interpretability of Machine Learning Models

University of Tartu, Estonia. May 2018

Understanding and Assessing Language with Neural Network Models

Grammarly, Ukraine. March 2018

Neural Architectures for Sequence Labelling

Grammarly, Ukraine. March 2018

Auxiliary Objectives for Neural Sequence Labelling Models

University of Copenhagen, Denmark. November 2017

Sequence Labeling Models for Error Detection in Learner Writing

Machine Learning for Further Language Learning, summer school in Chania, Greece. July 2017

Neural Architectures for Sequence Labelling

Apple Cambridge. June 2017

Neural Architectures for Sequence Labelling

University College London. May 2017

Neural Architectures for Sequence Labelling

University of Cambridge. May 2017

Character-based and Multi-task Learning for Neural Sequence Labeling

University of Sheffield. March 2017

Additional Academic Activities

Running a weekly reading group on Machine Learning in the Cambridge NLP research group.

Writing a blog on Machine Learning and NLP, with roughly 3,500 visitors per month.

Managing the official website of the Cambridge NLP group.

Creating and managing the Cambridge NLP Twitter account, publicising research and achievements coming from the group.

Creating and managing the Cambridge NLP Facebook group, for organizing joint social events and promoting informal discussion.

Programming Skills

Main experience: Python, Tensorflow, Theano, Java, JavaScript, PHP, SQL, HTML, CSS

Limited experience: PyTorch, C++, C, Android, Clojure, Perl, JSP, Lisp, Prolog, Haskell