

Todor Mihaylov

94404, Foster City, CA, USA
tbmihailov@gmail.com

RESEARCH INTERESTS *Natural Language Processing with Text and External Knowledge, Instruction Learning, Few-shot Learning with Large Language Models*

WORK EXPERIENCE *Research Scientist* Nov 2020 - Now
Facebook AI, Menlo Park, CA, USA

Doctoral Researcher April 2016 - Sep 2019
Department of Computational Linguistics
Heidelberg University, Heidelberg, Germany
Conducting research in natural language processing as part of doctoral studies.

Research Intern April 2018 - June 2018
The Allen Institute for Artificial Intelligence, Seattle, WA, USA
Work on science question answering at the Aristo team.

Research Scientist Intern June 2017 - September 2017
Amazon AI, AWS Deep Learning, Palo Alto, CA, USA
Work on transfer learning for question answering.

Senior Full-Stack Software Engineer / Team Lead October 2008 - March 2016
“Acstre“ Software Development Group
Faculty of Applied Mathematics and Informatics
Technical University of Sofia
Developing end-to-end web and server business applications.

EDUCATION *PhD in Computational Linguistics* April 2016 - 2022
Thesis: Knowledge-Enhanced Neural Networks for Machine Reading Comprehension
Department of Computational Linguistics
Heidelberg University, Germany
Research Training Group AIPHES

MSc in Computer Science October 2013 - July 2015
Information Retrieval and Knowledge Discovery
Sofia University, Sofia, Bulgaria

Bachelor of Computer Science, September 2008 - May 2012
Computer Systems and Technologies
Technical University of Sofia, Sofia, Bulgaria

TEACHING *Teaching Assistant (Machine Learning, Information Retrieval)* Oct 2014 - Feb 2016
Faculty of Mathematics and Informatics
Sofia University

- Lab teaching for the Data Mining Course.
- Lab teaching for the Information Retrieval Course.

- PUBLICATIONS** *Efficient Large Scale Language Modeling with Mixtures of Experts* Dec 2021
Pre-rint arXiv:2112.10684
Mikel Artetxe*, Shruti Bhosale*, Naman Goyal*, **Todor Mihaylov**, Myle Ott*, Sam Shleifer*, Xi Victoria Lin, Jingfei Du, Srinivasan Iyer, Ramakanth Pasunuru, Giri Anantharaman, Xian Li, Shuohui Chen, Halil Akin, Mandeep Baines, Louis Martin, Xing Zhou, Punit Singh Koura, Brian O'Horo, Jeff Wang, Luke Zettlemoyer, Mona Diab, Zornitsa Kozareva, Ves Stoyanov
- Few-shot Learning with Multilingual Language Models* Dec 2021
Pre-rint arXiv:2112.10668
Xi Victoria Lin*, **Todor Mihaylov**, Mikel Artetxe, Tianlu Wang, Shuohui Chen, Daniel Simig, Myle Ott, Naman Goyal, Shruti Bhosale, Jingfei Du, Ramakanth Pasunuru, Sam Shleifer, Punit Singh Koura, Vishrav Chaudhary, Brian O'Horo, Jeff Wang, Luke Zettlemoyer, Zornitsa Kozareva, Mona Diab, Veselin Stoyanov, Xian Li*
- EXAMS: A Multi-subject High School Examinations Dataset for Cross-lingual and Multilingual Question Answering* Nov 2020
Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020
Momchil Hardalov, **Todor Mihaylov**, Dimitrina Zlatkova, Yoan Dinkov, Ivan Koychev, Preslav Nakov
- Discourse-Aware Semantic Self-Attention for Narrative Reading Comprehension* Nov 2019
Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019
Todor Mihaylov, Anette Frank
- Can a Suit of Armor Conduct Electricity? A New Dataset for Open Book Question Answering* Oct 2018
Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018
Todor Mihaylov, Peter Clark, Tushar Khot, Ashish Sabharwal
- Knowledgeable Reader: Enhancing Cloze-Style Reading Comprehension with External Commonsense Knowledge* July 2018
Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018
Todor Mihaylov, Anette Frank
- What's Important in a Text? An Extensive Evaluation of Linguistic Annotations for Summarization* April 2018
Proceedings of the Second International Workshop on Advances in Natural Language Processing, Valencia, Spanien.
Markus Zopf, Teresa Botschen, Tobias Falke, Ana Marasovic, **Todor Mihaylov**, Avinash P.V.S, Eneldo Loza Menca, Johannes Frnkranz und Anette Frank
- The Dark Side of News Community Forums: Opinion Manipulation Trolls* March 2018
Internet Research, 2018
Todor Mihaylov, Tsvetomila Mihaylova, Preslav Nakov, Llus Mrquez, Georgi D. Georgiev, Ivan Koychev

Neural Skill Transfer from Supervised Language Tasks to Reading Comprehension November 2017
 Workshop on Learning with Limited Labeled Data (LLD) at NIPS 2017
Todor Mihaylov, Zornitsa Kozareva, Anette Frank

Large-Scale Goodness Polarity Lexicons for Community Question Answering August 2017
 In Proceedings of the 40th International Conference on Research and Development in Information Retrieval (SIGIR) 2017
Todor Mihaylov, Daniel Balchev, Yassen Kiproff, Ivan Koychev, Preslav Nakov

Story Cloze Ending Selection Baselines and Data Examination April 2017
Proceedings of the Linking Models of Lexical, Sentential and Discourse-level Semantics Shared Task 2017
Todor Mihaylov, Anette Frank

AIPHES-HD system at TAC KBP 2016: Neural Event Trigger Span Detection and Event Type and Realis Disambiguation with Word Embeddings November 2016
 Proceedings of the TAC Knowledge Base Population (KBP) 2016
Todor Mihaylov, Anette Frank

Discourse Relation Sense Classification Using Cross-argument Semantic Similarity Based on Word Embeddings August 2016
 Proceedings of the Twentieth Conference on Computational Natural Language Learning - Shared Task, 2016
Todor Mihaylov, Anette Frank

Hunting for Troll Comments in News Community Forums August 2016
 The 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016
Todor Mihaylov, Preslav Nakov, Georgi D Georgiev

SemanticZ at SemEval-2016 Task 3: Ranking Relevant Answers in Community Question Answering Using Semantic Similarity Based On Fine-tuned Word Embeddings June 2016
 International Workshop on Semantic Evaluation (SemEval) 2016
Todor Mihaylov, Preslav Nakov

Super Team at SemEval-2016 Task 3: Building a Hybrid System for Community Question Answering June 2016
 International Workshop on Semantic Evaluation 2016
 Tsvetomila Mihaylova, Pepa Gencheva, Martin Boyanov, Ivana Yovcheva
Todor Mihaylov, Momchil Hardalov, Yassen Kiproff, Daniel Balchev, Ivan Koychev, Preslav Nakov Ivelina Nikolova, Galia Angelova

Exposing Paid Opinion Manipulation Trolls September 2015
 Recent Advantages in Natural Language Processing (RANLP) 2015
Todor Mihaylov, Ivan Koychev, Georgi D Georgiev, Preslav Nakov

Finding Opinion Manipulation Trolls in News Community Forums July 2015
 The 19th Conference on Computational Natural Language Learning, CONLL 2015
Todor Mihaylov, Georgi D Georgiev, Preslav Nakov

SU-FMI: System Description for SemEval-2014 Task 9 on Sentiment August 2014

Analysis in Twitter, SemEval 2014
Boris Velichkov, Borislav Kapukaranov, Ivan Grozev, Jeni Karanesheva
Todor Mihaylov, Yassen Kiprof, Georgi Georgiev, Ivan Koychev, Preslav Nakov

REVIEWING *Conferences*: EMNLP 2016, ACL 2017, ACL 2018, NAACL-HLT 2019, EMNLP 2020, EMNLP 2021, ARR Nov 21 - Now
Workshops: SemEval 2014, SemEval 2016, CONLL-ST 2016, LSDSem-ST 2017, Rep4NLP 2018, SSS 2019, Rep4NLP 2019, COIN 2019, EurNLP 2019, Rep4NLP 2020, MRL 2021

TEACHING EXPERIENCE *Teaching Assistant (Machine Learning, Information Retrieval)* Oct 2014 - Feb 2016
Faculty of Mathematics and Informatics
Sofia University

HONORS AND AWARDS *3rd in LDSem 2017 Shared Task on the Story Cloze Test* Feb 2017
3rd best performing system in the official evaluation

1st in CoNLL-2016 Shared Task on Discourse Relation May 2016
Sense Classification

Best performing system in the overall evaluation for the Discourse Relation Sense Classification for English

3rd in SemEval-2016 Task 3: Community Question Answering Feb 2016
3rd place in Subtask A, English (Question-Comment Similarity)
3rd place in Subtask C, English (Question-External Comment Similarity)
1st place in Subtask C, English (Question-External Comment Similarity) (as a part of a team)

IIE's Summer Entrepreneurship Program (SEP) scholarship July 2014
International Institute of Education's Summer Entrepreneurship Program (SEP) scholarship part of the Bulgaria Young Leaders Program (BYLP) by the America for Bulgaria Foundation

Winner of BAIT Awards for 2013 , Students awards January 2014
Annual Awards of Bulgarian Association of Information Technologies

WEBLOZ Students'12 winner in category Web Applications February 2012
Student software development competition

TECHNOLOGY SKILLS *Machine learning and NLP tools*: PyTorch (4 years), TensorFlow (2 years), AllenNLP, scikit-learn, gensim, NLTK, Libsvm/Liblinear, GATE
Programming Languages: Python (6 years), C# (7 years), Delphi Pascal (7 years), Java
Other Technologies: Apache Lucene, Apache Solr, PostgreSQL (4 years), ASP.NET

MVC (5 years), ASP.NET Web API (6 years), ASP.NET Web Forms (4), WCF REST Services (6 years), WCF RIA Services (3 years), PostgreSQL (5 years), MS SQL Server, Firebird (8 years), MySQL, Ruby on Rails

LANGUAGES

Bulgarian: Native language

English: Full professional proficiency

German: Elementary proficiency

ADDITIONAL

Personal page: <https://tmihaylov.com>

Google Scholar profile: <https://scholar.google.de/citations?user=igempQwAAAAJ&hl=en>

Linkedin profile: <https://www.linkedin.com/in/todormihaylov/>

GitHub profile: <https://github.com/tbmihailov>

Last update on 26 Feb 2022