

GenBench 2023: The First Workshop on Generalisation (Benchmarking) in NLP

Held at ACL 2016

Singapore
6 December 2023

ISBN: 978-1-7138-8600-6

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2023) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006

Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>90% F1 Score in Relation Triple Extraction: Is it Real?</i> Pratik Saini, Samiran Pal, Tapas Nayak and Indrajit Bhattacharya	1
<i>GenCodeSearchNet: A Benchmark Test Suite for Evaluating Generalization in Programming Language Understanding</i> Andor Diera, Abdelhalim Dahou, Lukas Galke, Fabian Karl, Florian Sihler and Ansgar Scherp	12
<i>Adapt and Decompose: Efficient Generalization of Text-to-SQL via Domain Adapted Least-To-Most Prompting</i> Aseem Arora, Shabbirhussain Bhaisaheb, Harshit Nigam, Manasi Patwardhan, Lovekesh Vig and Gautam Shroff	25
<i>Evaluating Neural Language Models as Cognitive Models of Language Acquisition</i> Hector Javier Vazquez Martinez, Annika Lea Heuser, Charles Yang and Jordan Kodner	48
<i>Robust Code Summarization</i> Debanjan Mondal, Abhilasha Lodha, Ankita Sahoo and Beena Kumari	65
<i>Temporal Generalizability in Multimodal Misinformation Detection</i> Nataliya Stepanova and Björn Ross	76
<i>Robust Generalization Strategies for Morpheme Glossing in an Endangered Language Documentation Context</i> Michael Ginn and Alexis Palmer	89
<i>Walking a Tightrope – Evaluating Large Language Models in High-Risk Domains</i> Chia-Chien Hung, Wiem Ben Rim, Lindsay Frost, Lars Bruckner and Carolin Lawrence	99
<i>Latent Feature-based Data Splits to Improve Generalisation Evaluation: A Hate Speech Detection Case Study</i> Maïke Züfle, Verna Dankers and Ivan Titov	112
<i>Syntax-Guided Transformers: Elevating Compositional Generalization and Grounding in Multimodal Environments</i> Danial Kamali and Parisa Kordjamshidi	130
<i>mSCAN: A Dataset for Multilingual Compositional Generalisation Evaluation</i> Amélie Reymond and Shane Steinert-Threlkeld	143
<i>Inductive Bias Is in the Eye of the Beholder</i> Michael Wilson and Robert Frank	152
<i>Blackbird Language Matrices Tasks for Generalization</i> Paola Merlo, Chunyang Jiang, Giuseppe Samo and Vivi Nastase	163
<i>In-Context Learning for Text Classification with Many Labels</i> Aristides Miliotis, Siva Reddy and Dzmitry Bahdanau	173
<i>GQG: Generalized Quantifier Generalization - A Dataset for Evaluating Quantifier Semantics Understanding in Language Models</i> Leroy Zhifei Wang and Shane Steinert-Threlkeld	185

<i>Cross-Lingual Data Augmentation For Thai Question-Answering</i>	
Parinthapat Pengpun, Can Udomcharoenchaikit, Weerayut Buaphet and Peerat Limkonchotiwat	
193	
<i>On using distribution-based compositionality assessment to evaluate compositional generalisation in machine translation</i>	
Anssi Moiso, Mathias Creutz and Mikko Kurimo	204
<i>Shifted PAUQ: Distribution shift in text-to-SQL</i>	
Oleg Somov and Elena Tutubalina	214