

First Workshop on Efficient Benchmarking in NLP (NLP-Power 2022)

Dublin, Ireland
26 May 2022

ISBN: 978-1-7138-6749-4

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© (2022) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2023)

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>Raison d'être of the benchmark dataset: A Survey of Current Practices of Benchmark Dataset Sharing Platforms</i> Jaihyun Park and Sullam Jeoung	1
<i>Towards Stronger Adversarial Baselines Through Human-AI Collaboration</i> Wencong You and Daniel Lowd	11
<i>Benchmarking for Public Health Surveillance tasks on Social Media with a Domain-Specific Pretrained Language Model</i> Usman Naseem, Byoung Chan Lee, Matloob Khushi, Jinman Kim and Adam Dunn.....	22
<i>Why only Micro-F1? Class Weighting of Measures for Relation Classification</i> David Harbecke, Yuxuan Chen, Leonhard Hennig and Christoph Alt	32
<i>Automatically Discarding Straplines to Improve Data Quality for Abstractive News Summarization</i> Amr Keleg, Matthias Lindemann, Danyang Liu, Wanqiu Long and Bonnie L. Webber.....	42
<i>A global analysis of metrics used for measuring performance in natural language processing</i> Kathrin Blagec, Georg Dorffner, Milad Moradi, Simon Ott and Matthias Samwald.....	52
<i>Beyond Static models and test sets: Benchmarking the potential of pre-trained models across tasks and languages</i> Kabir Ahuja, Sandipan Dandapat, Sunayana Sitaram and Monojit Choudhury	64
<i>Checking HateCheck: a cross-functional analysis of behaviour-aware learning for hate speech detection</i> Pedro Henrique Luz de Araujo and Benjamin Roth	75
<i>Language Invariant Properties in Natural Language Processing</i> Federico Bianchi, Debora Nozza and Dirk Hovy.....	84
<i>DACT-BERT: Differentiable Adaptive Computation Time for an Efficient BERT Inference</i> Cristobal Eyzaguirre, Felipe del Rio, Vladimir Araujo and Alvaro Soto	93
<i>Benchmarking Post-Hoc Interpretability Approaches for Transformer-based Misogyny Detection</i> Giuseppe Attanasio, Debora Nozza, Eliana Pastor and Dirk Hovy	100
<i>Characterizing the Efficiency vs. Accuracy Trade-off for Long-Context NLP Models</i> Phyllis Ang, Bhuwan Dhingra and Lisa Wu Wills	113