5th Workshop on Insights from Negative Results in NLP (Insights 2025)

Albuquerque, New Mexico, USA
4 May 2025

ISBN: 979-8-3313-1919-9

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© (2025) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

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

Challenging Assumptions in Learning Generic Text Style Embeddings
Phil Ostheimer, Marius Kloft and Sophie Fellenz 1

In-Context Learning on a Budget: A Case Study in Token Classification
Uri Berger, Tal Baumel and Gabriel Stanovsky 7

Reassessing Graph Linearization for Sequence-to-sequence AMR Parsing: On the Advantages and Limitations of Triple-Based
Jeongwoo Kang, Maximin Coavoux, Didier Schwab and Cédric Lopez 15

Corrective In-Context Learning: Evaluating Self-Correction in Large Language Models
Mario S a n z - G u e r r e r o and Katharina Von Der Wense 24

Do Prevalent Bias Metrics Capture Allocational Harms from LLMs?
Hannah Cyberey, Yangfeng Ji and David Evans 34

Language-Specific Neurons Do Not Facilitate Cross-Lingual Transfer
Soumen Kumar Mondal, Sayambhu Sen, Abhishek Singhania and Preethi Jyothi 46

Monte Carlo Sampling for Analyzing In-Context Examples
Stephanie Schoch and Yangfeng Ji 63

Does Training on Synthetic Data Make Models Less Robust?
Lingze Zhang and Ellie Pavlick 79

Bridging the Faithfulness Gap in Prototypical Models
Andrew Koulogeorge, Sean Xie, Saeed Hassanpour and Soroush Vosoughi 86

Aligning Sizes of Intermediate Layers by LoRA Adapter for Knowledge Distillation
Takeshi Suzuki, Hiroaki Yamada and Takenobu Tokunaga 100

LLMs are not Zero-Shot Reasoners for Biomedical Information Extraction
Aishik Nagar, Viktor Schlegel, T h a n h - T u n g Nguyen, Hao Li, Yuping Wu, Kuluhan Binici and Stefan Winkler 106

Exploring Limitations of LLM Capabilities with Multi-Problem Evaluation
Zhengxiang Wang, Jordan Kodner and Owen Rambow 121

Exploring Multimodal Language Models for Sustainability Disclosure Extraction: A Comparative Study
Tanay Gupta, Tushar Goel and Ishan Verma 141

Self Knowledge-Tracing for Tool Use (SKT-Tool): Helping LLM Agents Understand Their Capabilities in Tool Use
Joshua Vigel, Renpei Cai, Eleanor Chen, Anish Neema, Austen Liao, Kevin Zhu and Sean O'brien 150

Error Reflection Prompting: Can Large Language Models Successfully Understand Errors?
Jason Li, Lauren Yraola, Kevin Zhu and Sean O'brien 157

Evaluating Robustness of LLMs to Numerical Variations in Mathematical Reasoning
Yuli Yang, Hiroaki Yamada and Takenobu Tokunaga 171