Ngoc Duyen Tanja Tu / Annelen Brunner / Christian Lang

Editorial

Failed experiments typically have no place in scientific discourse; they are discarded and not published. We believe that this practice results in a loss of potential knowledge gain. A systematic reflection on the causes of failures allows for the critical examination and/or improvement of methods used. Furthermore, when previously failed experiments are repeated and subsequently succeed, progress can be explicitly determined. From the perspective of methodological reflection, the discussion and documentation of failures thus provide added value for the scientific community. This is particularly true in a field like research on and with generative artificial intelligence (AI), which lacks a long-standing tradition and in which best practices are still in the process of being established.

This JLCL special issue focuses on linguistic and NLP experiments with generative AI that did not yield the desired results. All papers explore the extent in which their failed experiment can contribute to knowledge gain regarding the work with generative AI.

The first three papers test LLMs for various annotation and information extraction tasks.

Elena Leitner and Georg Rehm systematically evaluate LLMs on several classification tasks for German social media texts. In this context, they compare different fine-tuning and prompting techniques and point out weaknesses.

Barbara Heinisch describes challenges in the use of LLMs in terminology work, an area in which accuracy and replicability are of particular importance. She advocates a selective application of LLMs in terminology work, emphasizing the importance of evaluating their appropriateness for specific tasks rather than using them indiscriminately.

Elena Volkanovska tackles the question on how to systematically deal with LLM specific errors. She proposes an error classification framework complementary to established performance metrics for NER classifiers that accounts for additional possible outcomes in a few-shot, LLM-based NER task.

The following three papers focus on the struggles of LLMs when presented with tasks that require complex semantic analysis.

Yanming Li and Meaghan Fowlie test several prompting strategies for four GPT-models to perform Abstract Meaning Representation (AMR) parsing on natural language sentences and find the performance worse than that of state-of-the-art AMR parsers.

Natalia Skachkova, Simon Ostermann, Josef van Genabith and Bernd Kiefer investigate the ability of different (L)LMs in bridging generation. They challenge the models with two tasks: (1) generate texts containing bridging and (2) fill in missing (bridging) spans.

Sebastian Reimann and **Tatjana Scheffler** provide a series of zero- and few-shot experiments on the detection of linguistic metaphors and specifically on extended metaphors with LLaMa and GPT models.

The final two papers approach the topic from a more meta perspective.

John David Storment focuses on the ability of LLMs to provide linguistic acceptability judgments and shows that they struggle with texts that use emojis as morpho-grammatical components.

Finally, the contribution by **Simon Münker** widens the scope to LLMs as social agents: He investigates how personalized LLMs align with human responses on the Moral Foundation Theory Questionnaire. His results suggest that LLMs struggle to coherently represent ideologies, cautioning against using them to simulate social interactions.

We would like to thank the authors for their contributions, which enabled us to compile this thematically rich special issue. We also thank the reviewers for their thorough feedback. Last but not least, we want to thank the editor of the Journal for Language Technology and Computational Linguistics for his support in putting together this special issue. We wish the reader a pleasant and engaging reading experience!

The guest editors,
Ngoc Duyen Tanja Tu, Annelen Brunner and Christian Lang

Figure 1: LLMfails llama mascot - generated with GPT-4o

Reviewer index

Matthias Aßenmacher
Department of Statistics, LMU Munich
Munich Center for Machine Learning (MCML)
matthias@stat.uni-muenchen.de

Louis Cotgrove
Department of Lexical Studies
Leibniz Institute for the German Language
cotgrove@ids-mannheim.de

Nils Diewald
Department of Digital Linguistics
Leibniz Institute for the German Language
diewald@ids-mannheim.de

Keli Du
Trier Center for Digital Humanities
University of Trier
duk@uni-trier.de

Jennifer Ecker
Department of Digital Linguistics
Leibniz Institute for the German Language
ecker@ids-mannheim.de

Tim Feldmüller
Department of Digital Linguistics
Leibniz Institute for the German Language
feldmueller@ids-mannheim.de

Leonard Konle
Lehrstuhl für Computerphilologie
University of Würzburg
leonard.konle@uni-wuerzburg.de

Marc Kupietz
Department of Digital Linguistics
Leibniz Institute for the German Language
kupietz@ids-mannheim.de

Mark-Christoph Müller ⬤
Department of Pragmatics
Leibniz Institute for the German Language
mark-christoph.mueller@ids-mannheim.de

Peter Meyer ⬤
Department of Lexical Studies
Leibniz Institute for the German Language
meyer@ids-mannheim.de

Nina C. Rastinger ⬤
Literary & Textual Studies
Austrian Centre for Digital Humanities and Cultural Heritage
ninaclaudia.rastinger@oeaw.ac.at

Ines Rehbein
Data & Web Science Group
University of Mannheim
rehbein@uni-mannheim.de

Nils Reiter ⬤
Department of Digital Humanities
University of Cologne
nils.reiter@uni-koeln.de

Jan Oliver Rüdiger ⬤
Department of Lexical Studies
Leibniz Institute for the German Language
ruediger@ids-mannheim.de

Josef Ruppenhofer ⬤
Center of Advanced Technology for Assisted Learning and Predictive Analytics
FernUniversität in Hagen
josef.ruppenhofer@fernuni-hagen.de

Roman Schneider ⬤
Department of Grammar
Leibniz Institute for the German Language
schneider@ids-mannheim.de

Martina Scholger 🄳
Department of Digital Humanities
University of Graz
martina.scholger@uni-graz.at

Pia Schwarz 🄳
Department of Digital Linguistics
Leibniz Institute for the German Language
schwarz@ids-mannheim.de

Karolina Suchowolec 🄳
Institute of Translation and Multilingual Communication
TH Köln - University of Applied Sciences
karolina.suchowolec@th-koeln.de

Sascha Wolfer 🄳
Department of Lexical Studies
Leibniz Institute for the German Language
wolfer@ids-mannheim.de