

RetroEval 2026

**Proceedings of the 1st Symposium on Natural Language
Generation Evaluations**

Aberdeen, Scotland, United Kingdom

June 1-2, 2026

The RetroEval organizers gratefully acknowledge the support from the following sponsors.

Sponsored by



Supported by



Endorsed by



©2026 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-436-1

Preface

We are excited to present the Proceedings of the 1st Symposium on Natural Language Generation Evaluations (RetroEval 2026). This special symposium celebrates the career and accomplishments of Prof. Ehud Reiter (University of Aberdeen) by looking at evaluation practices in the field of Natural Language Generation past, present, and future.

The first RetroEval symposium will take place from 1-2 June, in Aberdeen, United Kingdom, endorsed by the Association of Computational Linguistics Special Interest Group in Generation (ACL SIGGEN). We thank the local organisation team, including Arpita Gado, Mengxuan Sun, Yujun Wang, and Jakub Zbrzezny; the symposium would not have been possible without their efforts. We also thank Debbie Meharg, the head of the Department of Computing Science at the University of Aberdeen, for arranging financial support from the department for this event; and we thank Chrissy Sanachan from the Scottish Informatics and Computer Science Alliance (SICSA) for logistical support.

In addition to a number of papers that were selected following an open call and a light review process, we are excited to present three keynotes, which discuss pertinent issues in evaluation within Natural Language Processing:

- Sina Zarrieß, Universität Bielefeld, Germany
- Beatrice Alex, Heriot-Watt University, United Kingdom
- Albert Gatt, Universiteit Utrecht, the Netherlands

Taken jointly, the contributions to RetroEval illustrate how Ehud has helped to make our field what it is today, both in academia and in industry. They highlight Ehud's role as an enthusiastic inspirator and a kind mentor, and the role he has played as a builder and evaluator of NLG systems. Last but not least, they highlight the way in which Ehud has made the NLG community, and the NLP community more broadly, aware of the many pitfalls that can stand in the way of truly scientific system evaluation. It is to his work on evaluation that this symposium is devoted.

Your RetroEval 2026 Organizers,

Saad Mahamood, David M. Howcroft, Kees van Deemter, Adarsa Sivaprasad, Barkavi Sundararajan, Jose Maria Alonso-Moral, and Simone Balloccu

Programme Committee

Jose Maria Alonso-Moral, CiTIUS, University of Santiago de Compostela

Simone Balloccu, TU Darmstadt

Alberto Bugarín-Diz, CiTIUS, University of Santiago de Compostela

Silvia Casola, Ludwig Maximilian University of Munich

Ondřej Dušek, Charles University

Albert Gatt, Universiteit Utrecht

Dimitra Gkatzia, Napier University

David M. Howcroft, University of Aberdeen

Mateusz Lango, Charles University

Chenghua Lin, University of Manchester

Saad Mahamood, Shopware

Patrícia Schmidtová, Charles University

Kees van Deemter, Universiteit Utrecht

Mengxuan Sun, University of Aberdeen

Craig Thomson, Dublin City University

Table of Contents

<i>Decomposition Does Not Help: Evidence from Semantic Clustering in LLM-based Causal Graph Discovery</i>	
Nikolay Babakov and Alberto Bugarín-Diz	1
<i>NLG Evaluation: Past, Present, Future</i>	
Ehud Reiter	8
<i>Evaluation and Assessment as Complementary Frameworks</i>	
Elie Antoine	16
<i>The Arabic Bible as an Evaluation Tool: The Case Study of the Khalīlī Arabic Dialect</i>	
Jakub Zbrzeźny, Ehud Reiter and Wei Zhao	24
<i>RAG as a collapsed NLG pipeline</i>	
Adarsa Sivaprasad, Barkavi Sundararajan and David M. Howcroft	33
<i>A Comparative Evaluation of End-to-End and Pipeline Approaches for Summarisation</i>	
Fahime Same, Saad Mahamood and Srinivas Ramesh Kamath	39
<i>Solving the Task but Not the Problem: A Customer Support Case Study on Why Extrinsic Evaluation Matters</i>	
Daniel Braun	53
<i>Never Truly Out of Fashion: A Retrospective Look at Evaluation in NLG</i>	
Patrícia Schmidtová, Saad Mahamood and Ondřej Dušek	63