GEM 2022

2nd Workshop on Natural Language Generation, Evaluation, and Metrics

Proceedings of the Workshop

December 7, 2022

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-959429-12-8

Introduction

Welcome to the second workshop on Natural Language Generation, Evaluation, and Metrics (GEM), to be held on December 7, 2022 as part of EMNLP in Abu Dhabi. The workshop aims to bring together researchers interested in model audits, new evaluatyion approaches and meta evaluations. The workshop is privileged to present several invited talks this year and the results of the shared task on generation with limited resources.

We are grateful to the program committee for their careful and thoughtful reviews of the papers submitted this year. Likewise, we are thankful to the shared task organizers for their hard work in preparing the shared tasks. We are looking forward to a workshop covering a wide range of topics, and we hope for lively discussions.

GEM Workshop Organizers.

Organizing Committee

Organizers

Antoine Bosselut, EPFL
Khyathi Chandu, Meta AI
Kaustubh Dhole, Emory University
Varun Gangal, Carnegie Mellon University
Sebastian Gehrmann, Google Research
Yacine Jernite, Hugging Face
Jekaterina Novikova, Winterlight Labs
Laura Perez-Beltrachini, University of Edinburgh

Program Committee

Reviewers

Tosin Adewumi, Ameeta Agrawal, Fatih Amasyali, David Aparicio, Samuel Arcadinho, Shima Asaadi

Simone Balloccu, Indrajit Bhattacharya, Şafak Bilici, Bernd Bohnet

Samuel Cahyawijaya, Pengshan Cai, Eduardo Calò, Ronald Cardenas, Boaz Carmeli, Silvia Casola, Khyathi Raghavi Chandu, Elizabeth Clark, Jordan Clive

Kordula De Kuthy, Mingkai Deng, Yuntian Deng, Daniel Deutsch, Mark Dingemanse, Esin Durmus, Ondřej Dušek

Moussa Kamal Eddine

Federico Fancellu, Raquel Fernandez

Albert Gatt, John Glover, Olga Golovneva

Tahmid Hasan, Behnam Hedayatnia, David M. Howcroft, Shulin Huang, Rudali Huidrom

Nikolai Ilinykh

Yacine Jernite, Mayank Jobanputra, Shailza Jolly

Emil Kalbaliyev, Mihir Kale, Marzena Karpinska, Noriaki Kawamae, Dimitar Kazakov, George Kour, Sergey Kovalchuk, Kalpesh Krishna, Rishu Kumar, Vanya Bannihatti Kumar

Harsh Lara, Alberto Lavelli, Hwanhee Lee, Jing Yang Lee, Yinghui Li, Paul Pu Liang, Andreas Liesenfeld, Sijia Liu, Yinhong Liu, Yixin Liu, Zhengzhong Liu, Ehsan Lotfi

Shirong Ma, Aman Madaan, Khyati Mahajan, Abinaya Mahendiran, Andreas Marfurt, Simon Mille, Sebastien Montella

Tapas Nayak, Vitaly Nikolaev, Tadashi Nomoto

Salomey Osei

Alexandros Papangelis, Cheoneum Park, Eunil Park, Tatiana Passali, Diogo Pernes, Sandro Pezzelle, Maja Popović, Jiashu Pu

Vipul Raheja, Anand A. Rajasekar, Vikas Raunak, Ehud Reiter, Leonardo F. R. Ribeiro, Daniele Riboni, Giuseppe Riccardi, Michael Ridenour, Terry Ruas

Rifat Shahriyar, Tianhao Shen, Anna Shvets, Arabella Sinclair, Marco Antonio Sobrevilla Cabezudo, Somayajulu Sripada, Yixuan Su, Barkavi Sundararajan

Bowen Tan, Katherine Thai, Craig Thomson, Grigorios Tsoumakas

Ashish Upadhyay

Jan Philip Wahle, Shira Wein, Michael White, Genta Winata, Zixiu Wu

Yadong Xi, Deyi Xiong, Xinnuo Xu, Yumo Xu

Li Yangning

Alessandra Zarcone, Jiawei Zhou, Yongxin Zhou, Qi Zhu

Keynote Talk: Challenges in evaluating safety for LLMs

Emily Dinan
FAIR (Meta AI)

Abstract: While research on large language models (LLMs) continues to accelerate, much recent work has called attention to anticipated risks and harms from their use in society. We will discuss challenges in evaluating the relative safety of these models as well as current approaches for doing so. Finally, we will highlight avenues for future research into evaluating and mitigating these harms.

Bio: Emily Dinan is a Research Engineer at FAIR (Meta AI) in New York. Her research interests include conversational AI, natural language processing, and safety and responsibility in these fields. Recently she has focused on methods for preventing conversational agents from reproducing biased, toxic, or otherwise harmful language. Prior to joining FAIR, she received her master's degree in Mathematics from the University of Washington.

Keynote Talk: Instructable and Collaborative Language Models

Timo Schick
FAIR (Meta AI)

Abstract: Textual content is often the output of a collaborative writing process — which includes writing text, making comments and changes, finding references, and asking others for help —, but today's NLP models are only trained to generate the final output of this process. In this talk, we will discuss an alternative approach where models are trained to imitate the entire writing process. We will look at examples of how this enables models to plan and explain their actions, to correct their own mistakes, and to better collaborate with humans. We will also discuss how to make such models better at following human-written instructions.

Bio: Timo Schick is a research scientist at FAIR working on few-shot learning in NLP. Previously, he did his PhD at the Center for Information and Language Processing (CIS) in Munich and worked in industry as a data scientist for several years. Timo's current research focuses on instruction-based learning and teaching language models to collaborate with other entities.

Keynote Talk: Reflections on Trusting Untrustworthy Language Generators

Sean Welleck
University of Washington

Abstract: In his 1984 Turing Award Lecture "Reflections on Trusting Trust", Ken Thompson famously said "You can't trust code that you did not totally create yourself". These words are especially relevant today, as powerful and flexible language models generate natural language and code that is increasingly human-like. However, these same systems challenge our trust, exhibiting odd degeneracies, amplifying biases, and producing flawed reasoning. In this talk, I will introduce two directions for harnessing the potential of these language models while mitigating the risks. First, I will discuss unlearning: removing undesirable behaviors by integrating feedback and learning. Second, I will discuss how integrating language models with trustworthy symbolic systems can open the door to tackling challenging mathematical reasoning tasks. Join me as we explore the path towards trusting untrustworthy language generators.

Bio: Sean Welleck is a Postdoctoral Scholar at the University of Washington and the Allen Institute for Artificial Intelligence, working with Yejin Choi. His research focuses on algorithms for natural language generation and machine reasoning, with the aim of minimizing the effort needed to trust the output of AI systems. He has developed unlearning, decoding, and evaluation algorithms for controllable neural language generation, and methods for integrating language models with symbolic systems, with a particular focus on mathematical reasoning. He received his Ph.D. from New York University, where he was advised by Kyunghyun Cho. Outside of his research activities, he hosts the Thesis Review Podcast and enjoys running long distances.

Table of Contents

Program

Wednesday, December 7, 2022

09:00 - 10:30 *Opening Remarks and Keynote (Sean Welleck)*

10:30 - 11:00 *Coffee Break*

11:00 - 12:30 *Talk Session*

12:30 - 14:00 *Lunch Break*

14:00 - 15:30 *Poster Session*

15:30 - 16:00 *Coffee Break*

16:00 - 17:00 *Keynote (Timo Schick)*

17:00 - 18:30 *Talk Session*

20:00 - 21:00 *Virtual Keynote (Emily Dinan)*

21:00 - 22:30 *Virtual Poster Session*