

Automatic and Human-AI Interactive Text Generation (with a focus on Text Simplification and Revision)

Yao Dou, Philippe Laban, Claire Gardent and Wei Xu

In this tutorial, we focus on text-to-text generation, a class of natural language generation (NLG) tasks, that takes a piece of text as input and then generates a revision that is improved according to some specific criteria (e.g., readability or linguistic styles), while largely retaining the original meaning and the length of the text. This includes many useful applications, such as text simplification, paraphrase generation, style transfer, etc. In contrast to text summarization and open-ended text completion (e.g., story), the text-to-text generation tasks we discuss in this tutorial are more constrained in terms of semantic consistency and targeted language styles. This level of control makes these tasks ideal testbeds for studying the ability of models to generate text that is both semantically adequate and stylistically appropriate. Moreover, these tasks are interesting from a technical standpoint, as they require complex combinations of lexical and syntactical transformations, stylistic control, and adherence to factual knowledge, – all at once. With a special focus on text simplification and revision, this tutorial aims to provide an overview of the state-of-the-art natural language generation research from four major aspects – Data, Models, Human-AI Collaboration, and Evaluation – and to discuss and showcase a few significant and recent advances: (1) the use of non-retrogressive approaches; (2) the shift from fine-tuning to prompting with large language models; (3) the development of new learnable metric and fine-grained human evaluation framework; (4) a growing body of studies and datasets on non-English languages; (5) the rise of HCI+NLP+Accessibility interdisciplinary research to create real-world writing assistant systems.

Yao Dou, Ph.D. student in the College of Computing at the Georgia Institute of Technology

email: douy@gatech.edu

website: <https://yao-dou.github.io/>

Yao Dou is a Ph.D. student in the College of Computing at the Georgia Institute of Technology, advised by Prof. Wei Xu. His research interests lie in natural language processing and machine learning. His recent work focuses on text simplification, text evaluation, and social media privacy.

Philippe Laban, Research Scientist at Salesforce Research

email: plaban@salesforce.com

website: <https://tingofurro.github.io/>

Philippe Laban is a Research Scientist at Salesforce Research. His research is at the intersection of NLP and HCI, focusing on several tasks within text generation, including text simplification and summarization. He received his Ph.D. in Computer Science from UC Berkeley in 2021. His thesis is titled “Unsupervised Text Generation and its Application to News Interfaces”. His recent work has focused on expanding the scope of text simplification to the paragraph and document-level and evaluating textediting interfaces. He publishes in both *ACL and HCI conferences, including work on interactive user interface design for NLP applications.

Claire Gardent, Senior Research Scientist at the French National Center for Scientific Research (CNRS)

email: claire.gardent@loria.fr

website: <https://members.loria.fr/CGardent/>

Claire Gardent is a Senior Research Scientist at the French National Center for Scientific Research (CNRS), based at the LORIA Computer Science research unit in Nancy, France. In 2022, she was selected as an ACL Fellow and was awarded the CNRS Silver Medal. She works in the field of NLP with a particular interest in Natural Language Generation. In 2017, she launched the WebNLG challenge, a shared task where the goal is to generate text from Knowledge Base fragments. She has proposed neural models for simplification and summarization; for the generation of long-form documents such as multi-document summaries and Wikipedia articles; for multilingual generation from Abstract Meaning Representations and for response generation in dialog. She currently heads the AI XNLG Chair on multi-lingual, multi-source NLG and the CNRS LIFT Research Network on Computational, Formal and Field Linguistics

Wei Xu, Assistant Professor in the College of Computing at the Georgia Institute of Technology

email: wei.xu@cc.gatech.edu

website: <https://cocoxu.github.io/>

Wei Xu is an Assistant Professor in the College of Computing at the Georgia Institute of Technology. Her recent research focuses on text generation (including data construction, controllable model, human and automatic evaluation), stylistics, analyzing and evaluating large language models (including multilingual capability, cross-lingual transfer learning, cultural bias, and cost efficiency). She is a recipient of the NSF CAREER Award, CrowdFlower AI for Everyone Award, best paper award from COLING 2018, and honorable mention from ACL 2023. She is an NAACL executive board member and regularly serves as a (senior) area chair for *ACL conferences. She frequently gives invited talks at universities and companies. She has given tutorials on “NLP for Social Media and Text Analysis” and has organized multiple workshops, including WNUT, GEM, and TSAR.