Tanya Goyal

Email: tanyagoyal@cornell.edu
Website: https://tagoyal.github.io/

Academic Positions

Assistant Professor, Cornell University   Fall 2024 -
Department of Computer Science

Postdoctoral Scholar, Princeton University   2023-2024
Princeton Language and Intelligence + Department of Computer Science

Education

University of Texas at Austin   Austin, Texas
PhD in Computer Science (Advisor: Greg Durrett)   2019 – 2023
MS in Computer Science   2017 – 2022
Awarded the **Bert Kay Dissertation Award 2023**. (awarded to 1 PhD graduate per year)

Indian Institute of Technology, Guwahati   Guwahati, India
B.Tech in Mathematics and Computing   2011 – 2015

Select Publications

D2PO: Discriminator-Guided DPO with Response Evaluation Models
Prasann Singhal, Nathan Lambert, Scott Niekum, **Tanya Goyal**, Greg Durrett
COLM 2024

FABLES: Evaluating faithfulness and content selection in book-length summarization
Yekyung Kim, Yapei Chang, Marzena Karpinska, Aparna Garimella, Varun Manjunatha, Kyle Lo,
Tanya Goyal, Mohit Iyyer
COLM 2024

A Long Way To Go: Investigating Length Correlations in RLHF
Prasann Singhal, **Tanya Goyal**, Jiacheng Xu, Greg Durrett.
COLM 2024

BooookScore: A systematic exploration of book-length summarization in the era of LLMs
Yapei Chang, Kyle Lo, **Tanya Goyal**, Mohit Iyyer.
ICLR 2024

Evaluating Large Language Models at Evaluating Instruction Following
Zhiyuan Zeng, Jiatong Yu, Tianyu Gao, Yu Meng, **Tanya Goyal**, Danqi Chen.
ICLR 2024

WICE: Real-World Entailment for Claims in Wikipedia
Ryo Kamoi, **Tanya Goyal**, Juan Diego Rodriguez, and Greg Durrett.
EMNLP 2023

Understanding Factual Errors in Summarization: Errors, Summarizers, Datasets, Error Detectors
Liyan Tang, **Tanya Goyal**, Alexander Fabbri, Philippe Laban, Jiacheng Xu, Semih Yavuz, Wojciech Kryściński, Justin F. Rousseau, Greg Durrett.
ACL 2023

Shortcomings of Question Answering Based Factuality Frameworks for Error Localization

Ryo Kamoi, **Tanya Goyal**, Greg Durrett.
EACL 2023

SNaC: Coherence Error Detection for Narrative Summarization
Tanya Goyal, Jessy Li, Greg Durrett.
EMNLP 2022

HᴏᴅʀᴀSᴜᴍ: Disentangling Style Features in Text Summarization with Multi-Decoder Models
Tanya Goyal, Nazneen Rajani, Wenhao Liu, Wojciech Kryściński.
EMNLP 2022

FALTE: A Toolkit for Fine-grained Annotation for Long Text Evaluation
Tanya Goyal, Jessy Li, Greg Durrett.
EMNLP-Demo 2022

Training and Fine-tuning Dynamics for Summarization
Tanya Goyal, Jiacheng Xu, Jessy Li, Greg Durrett.
ACL Findings 2022

Annotating and Modeling Fine-grained Factuality in Summarization
Tanya Goyal, Greg Durrett.
NAACL 2021.

Evaluating Factuality in Generation with Dependency-level Entailment
Tanya Goyal, Greg Durrett.
EMNLP Findings 2020.

Neural Syntactic Preordering for Controlled Paraphrase Generation
Tanya Goyal, Greg Durrett.
ACL 2020.

Preprints

One Thousand and One Pairs: A "novel" challenge for long-context language models
Marzena Karpinska, Katherine Thai, Kyle Lo, **Tanya Goyal**, Mohit Iyyer.
arxiv 2024

LitSearch: A Retrieval Benchmark for Scientific Literature Search
Anirudh Ajith, Mengzhou Xia, Alexis Chevalier, **Tanya Goyal**, Danqi Chen, Tianyu Gao
arxiv 2024

News Summarization and Evaluation in the Era of GPT-3
Tanya Goyal, Jessy Li, Greg Durrett.
arxiv 2022

Work Experience

Salesforce Research   Summer 2021
Research Intern
Host: Wojciech Kryściński, worked on text summarization. Filed 1 US patent based on this research

A9 Labs, Amazon   Summer 2018
Research Intern
Project: Predictive modelling for customer engagement

Adobe Research, India   May 2015 - July 2017

Member of Technical Staff
Prototyped predictive models for e-commerce objectives for Adobe Document Cloud
Granted 12 US patents based on my work here.

Service

Area Chair
COLM (2024), LREC-COLING (summarization and generation track, 2024), ARR (multiple cycles, 2023-present), EMNLP 2023 (summarization track), EACL 2023 (summarization track)

Reviewer (*outstanding reviewer)
EMNLP (2020*, 2021, 2022), CoNLL (2020), NAACL (2021, 2022), ACL (2021, 2022), AKBC (2021), ARR (multiple cycles, 2021-present), TACL (standing reviewer, 2023-present)

Mentorship

Research Mentorship
- Chengyu Huang (PhD at Cornell University)   Fall 2024 - Present
- Nora Gera (PhD at Cornell University)   Fall 2024 - Present
- Prasann Singhal (Undergrad at UT Austin)   Fall 2023 - Present
- Ryo Kamoi (MS at UT Austin, next PhD UPitt)   Spring 2022 - Present
- Liyan Tang (PhD at UT Austin)   Spring 2022
- Sachin Kelkar (Research Intern at Adobe Research, next MS at Columbia)   Summer 2016
- Manas Agarwal (Research Intern at Adobe Research, next MS at Columbia)   Summer 2016
- Jeenu Grover (Research Intern at Adobe Research, next Google India)   Summer 2016

Undergraduate Reading Group, UT Austin   Fall 2021, Spring 2022
Lead a small-sized reading group with ~5 undergraduate students covering basics of NLP.

Mentor for Graduate Application Assistance Program, UT Austin   2021-2022
Program to provide assistance during graduate school applications to underrepresented groups

Invited Talks

Clinical and Translational Science Center, Weill Cornell Medicine   June 2024
Title: Automatic Evaluation of Chatbots in the era of LLMs

Keynote @ Singapore Symposium on NLP (SSNLP 2023)   December 2023
Title: Evaluation in the GPT-3 Era

LLM panel @ IEEE International Conference on Healthcare Informatics   June 2023
Title: Evaluating Text Generation in the GPT-3 Era

Microsoft Research   March 2023
Title: Building Reliable Text Generation Models

NLP Seminar Series at AI Sweden (link)   December 2022
Title: Text summarization and evaluation in the era of GPT-3

Adobe Research India   November 2022
Title: Bridging the gap between generation and evaluation capabilities of large language models

University of Alberta   July 2019
Title: Syntax-controlled generation using deep neural models

Other Publications

NL-Augmenter: A framework for task-sensitive natural language augmentation
Kaustubh D Dhole, et al.
NEJLT 2023

Contemporary NLP Modeling in Six Comprehensive Programming Assignments
Greg Durrett, Jifan Chen, Shrey Desai, **Tanya Goyal**, Lucas Kabela, Yasumasa Onoe, Jiacheng Xu.
Workshop on Teaching NLP 2021

Embedding time expressions for deep temporal ordering models
Tanya Goyal, Greg Durrett.
ACL 2019.

Your behavior signals your reliability: Modeling crowd behavioral traces to ensure quality relevance annotations
Tanya Goyal, Tyler McDonnell, Mucahid Kutlu, Tamer Elsayed, Matthew Lease.
HCOMP 2018.

An Empirical Analysis of Edit Importance between Document Versions
Tanya Goyal, Sachin Kelkar, Manas Agarwal, Jeenu Grover.
EMNLP 2017

Preventing inadvertent information disclosures via automatic security policies
Tanya Goyal, Sanket Mehta, Balaji Vasan Srinivasan.
PAKDD 2017

Environment Specific Content Rendering and Transformation
Balaji Vasan Srinivasan, **Tanya Goyal**, Varun Syal, Shubhankar Singh, Vineet Sharma.
IUI 2016 Companion

Patents *Granted by US PTO*

Filed by Salesforce Research

Multi-attribute control for text summarization using multiple decoder heads	*US Patent 11,755,637*

Filed by Adobe Research

Generating data-driven geo-fences	*US Patent 9,838,843*
Method to expand seed keywords into a relevant social query	*US Patent 9,965,766*
Generating predictive models for authoring short messages	*US Patent 10,528,652*
Content to layout template mapping and transformation	*US Patent 10,521,494*
Classifying and ranking changes between document revisions	*US Patent 10,713,432*
Constructing enterprise-specific knowledge graphs	*US Patent 10,915,577*
Determination of Paywall Metrics	*US Patent 10,853,887*
Prediction of tone of interpersonal text communications	*US Patent 10,796,095*
Tagging documents with security policies	*US Patent 11,748,501*
Content optimization for audiences	*US Patent 10,922,492*
Bundling online content fragments for presentation based on content-specific metrics and inter-content constraints	*US Patent 10,891,667*
Method and apparatus for generating predictive insights for authoring short messages	*US Patent 10,073,822*