

## RESEARCH AREAS

---

**Primary Interest:** Natural Language Generation (NLG), with specific interests in style transfer, low-resource & creative NLG, narrative generation and data-to-text generation.

**Secondary Interest:** Data Augmentation (DA), with specific interests in DA for generation, DA for better evaluating models and assessing their robustness to domain shift.

## EDUCATION

---

- **Language Technologies Institute, CMU** Pittsburgh, Pennsylvania  
PhD in Language Technologies; **GPA:** 3.97/4.33 Sep. 2016 – Ongoing
  - **Advisor:** Eduard Hovy
  - **Key Courses:** *Grammars & Lexicons, Machine Translation, Language Grounding For Vision & Control, Neural Networks For NLP, Algorithms for NLP, Deep Reinforcement Learning, Structured Prediction for NLP*
- **Indian Institute of Technology, Madras (IIT-M)** Chennai, India  
B.Tech/M.Tech. in Computer Science and Engineering; **CGPA:** 9.27/10 Aug. 2011 – May. 2016
  - **Advisor:** Balaraman Ravindran
  - **Key Electives:** *Machine Learning, NLP, Reinforcement Learning, Memory Based Reasoning, Searching & Indexing, Knowledge Representation, Distributed Algorithms, Ontologies, Graph Theory*
  - **GRE:** 337/340 ( **Quant:** 170/170 **Verbal:** 167/170 **Analytical Writing:** 5.5/6)
  - **TOEFL:** 118/120 ( **Reading:** 30/30 **Writing:** 30/30 **Listening:** 30/30 **Speaking:** 28/30)

## PUBLICATIONS

---

- [C.19] NAREOR: The Narrative Reordering Problem [ArXiv]  
*Varun Gangal\**, Steven Y Feng\*, Malihe Alikhani, Teruko Mitamura, Eduard Hovy  
Accepted for **AAAI 2022**
- [C.18] Retrieve, Caption, Generate: Visual Grounding for Enhancing Commonsense in Text Generators [ArXiv] [Slides]  
Steven Feng, Kevin Lu, Zhuofu Tao, Malihe Alikhani, Teruko Mitamura, Eduard Hovy, *Varun Gangal*  
Accepted for **AAAI 2022**
- [C.17] Investigating Robustness of Dialog Models to Popular Figurative Language Constructs [PDF]  
Harsh Jhamtani\*, *Varun Gangal\**, Eduard Hovy, Taylor Berg-Kirkpatrick  
Accepted for **EMNLP 2021**
- [C.16] Coarse2Fine: Fine-grained Text Classification on Coarsely-grained Annotated Data [PDF] [ArXiv]  
Dheeraj Mekala, *Varun Gangal*, Jingbo Shang  
Accepted for **EMNLP 2021**
- [C.15] SAPPHERE: Approaches for Enhanced Concept-to-Text Generation [PDF] [ArXiv] [CODE] [POSTER]  
Steven Feng, Jessica Huynh, Chaitanya Narisetty, Eduard Hovy, *Varun Gangal*  
Accepted for **INLG 2021** [🏆 **Best Long Paper**]
- [C.14] Automatic Construction of Evaluation Suites for Natural Language Generation Datasets [PDF] [ArXiv] [CODE] [POSTER]  
Simon Mille, Kaustubh D Dhole, Saad Mahamood, Laura Perez, *Varun Gangal*, Mihir Kale, Emiel van Miltenburg, Sebastian Gehrmann  
Accepted for **NeurIPS 2021 Datasets and Benchmarks Track**
- [C.13] Improving Automated Evaluation of Open Domain Dialog via Diverse Reference Augmentation [PDF] [ArXiv] [CODE] [POSTER]  
*Varun Gangal\**, Harsh Jhamtani\*, Eduard Hovy, Taylor Berg-Kirkpatrick  
Accepted for **Findings of ACL 2021**
- [C.12] A Survey of Data Augmentation Approaches for NLP [PDF] [ArXiv] [REPO] [TALK (for Google Research)]  
Steven Feng\*, *Varun Gangal\**, Jason Wei, Sarath Chandar, Soroush Vosoughi, Teruko Mitamura, Eduard Hovy  
Accepted for **Findings of ACL 2021**

- [C.11] BERTing RAMS: What and How Much does BERT Already Know About Event Arguments? – A Study on the RAMS Dataset [[ArXiv](#)][[PDF](#)]  
*Varun Gangal, Eduard Hovy*  
Accepted for **BlackBoxNLP 2020 Workshop, EMNLP 2020**
- [C.10] GenAug: Data Augmentation for Finetuning Text Generators [[ArXiv](#)] [[PDF](#)] [[CODE](#)] [[TALK](#)]  
*Steven Feng\*, Varun Gangal\*, Dongyeop Kang, Teruko Mitamura, Eduard Hovy*  
Accepted for **Deep Learning Inside Out Workshop, EMNLP 2020**
- [C.9] SCDE: Sentence Cloze Dataset with High Quality Distractors From Examinations [[PDF](#)] [[DATA](#)] [[CODE](#)]  
*Xiang Kong\*, Varun Gangal\*, Eduard Hovy*  
Accepted for **ACL 2020**
- [C.8] Likelihood Ratios and Generative Classifiers For Unsupervised OOD Detection In Task-Based Dialog [[PDF](#)] [[DATA](#)]  
*Varun Gangal, Abhinav Arora, Arash Einolghozati, Sonal Gupta*  
Accepted for **AAAI 2020**
- [C.7] (Male, Bachelor) and (Female, Ph.D) have different connotations: Parallely Annotated Stylistic Language Dataset with Multiple Personas [[PDF](#)] [[CODE+DATA](#)] [[PRESENTATION](#)]  
*Dongyeop Kang, Varun Gangal, Eduard Hovy*  
Accepted for **EMNLP 2019**
- [C.6] Learning to Generate Move-by-Move Commentary for Chess Games [[PDF](#)] [[POSTER](#)]  
*Harsh Jhamtani\*, Varun Gangal\*, Eduard Hovy, Graham Neubig, Taylor Berg-Kirkpatrick*  
Accepted for **ACL 2018**
- [C.5] Charmanteau: Character Embedding Models For Portmanteau Creation [[ARXIV](#)][[SLIDES](#)][[CODE](#)][[DATA](#)]  
*Varun Gangal\*, Harsh Jhamtani\*, Graham Neubig, Eduard Hovy, Eric Nyberg*  
Accepted for **EMNLP 2017**
- [C.4] Detecting and Explaining Causes From Text For a Time Series Event [[ARXIV](#)]  
*Dongyeop Kang, Varun Gangal, Ang Lu, Zheng Chen, Eduard Hovy*  
Accepted for **EMNLP 2017**
- [C.3] Shakespearizing Modern Language Using Copy-Enriched Sequence-to-Sequence Models [[ARXIV](#)][[CODE](#)]  
*Harsh Jhamtani\*, Varun Gangal\*, Eduard Hovy, Eric Nyberg*  
Accepted for **EMNLP 2017 Workshop on Stylistic Variation**
- [C.2] HEMI: Hyperedge Majority Influence Maximization [[PDF](#)]  
*Varun Gangal, Balaraman Ravindran, Ramasuri Narayanam*  
Accepted for **The Second IJCAI Workshop on Social Influence Analysis (SocInf 2016)**
- [C.1] Trust And Distrust Across Coalitions: Shapley Value Centrality Measures For Signed Networks [[PDF](#)]  
*Varun Gangal, Abhishek Narwekar, Balaraman Ravindran, Ramasuri Narayanam*  
Accepted for **NIPS 2015 Workshop on Networks In the Social And Information Sciences**

## ABSTRACTS

---

- [A.2] Personifications are Cunning: Exploring Approaches For Personification Identification [[PDF](#)]  
*Kevin Lu, Steven Feng, Varun Gangal, Harsh Jhamtani, Eduard Hovy*  
Accepted for **New Directions in Analyzing Text as Data (TADA) 2021**
- [A.1] Trust And Distrust Across Coalitions: Shapley Value Centrality Measures For Signed Networks [[PDF](#)]  
*Varun Gangal, Abhishek Narwekar, Balaraman Ravindran, Ramasuri Narayanam*  
Accepted for **AAAI Student Abstract 2016**

## PREPRINTS

---

- [P.1] The GEM benchmark: Natural Language Generation, its Evaluation and Metrics [[ArXiv](#)] [[CODE](#)]  
*Sebastian Gehrmann, Tosin Adeyemi\*, ... Varun Gangal\*, ... (Multiple Authors)*

## ORGANIZATIONAL EXPERIENCE

---

- **GEM: Natural Language Generation, its Evaluation and Metrics Workshop**  
Role : Organizer

ACL 2021  
Jan - July 2021

- This workshop aimed to create and discuss better and standardized evaluation and comparison of NLG models and systems - a parallel to **GLUE** for NLG. Was closely involved in choosing tasks, designing challenging test sets, developing basic tutorial notebooks, reviewing, and inviting panelists/speakers.
- **CtrlGen: Controllable Generative Modeling in Language and Vision Workshop** NEURIPS 2021  
 Role : Organizer Jan - December 2021
  - Our workshop explores disentanglement and control for generative models in CV and NL. Co-conceptualized idea with co-organizer **Steven Feng**, assembled co-organizer team, involved in proposal drafting, scheduling, inviting speakers, formulating Call for Papers and publicity.

## TALKS & LECTURES

---

- NOVEMBER 13, 2021: 1-hour invited Lecture as a part of the CS1699: Intro To NLP course at **University Of Pittsburgh**. Spoke on “*Style In NLG*”. [VIDEO]
- OCTOBER 15, 2021: Invited 1-hour Seminar talk at **University Of Pittsburgh**. Spoke on “*Endowing Generators With Style, Creativity and Commonsense*”. [VIDEO]
- AUGUST 30, 2021: Invited talk and discussion session at **Google Research** with co-author **Steven** on our “*A Survey of Data Augmentation Approaches for NLP*” work ([C.12]) [VIDEO]
- OCTOBER 22, 2020: Invited 1-hour talk at **University of Utah Data Science Seminar**. Presented my work on data augmentation for conditional generation ([C.10]) and probing extra-sentential abilities of BERT ([C.9],[C.11]) [VIDEO]
- FEBRUARY 8, 2016: Co-taught a two-part tutorial on Linear Algebra as part of the Introduction to Machine Learning MOOC with 135 participants, designed by IIT-Madras for the NPTEL platform. The videos have received 38K+ and 22K+ views respectively [PART I VIDEO][PART II VIDEO]

## REVIEWING

---

**Conferences:** ACL '19, EMNLP '19, AAAI '20, ACL '20, COLING '20, ACL-RR '21, AAAI '22  
**Workshops:** HAMLETS@ NEURIPS '20

## PROFESSIONAL EXPERIENCE

---

- **Unsupervised OOD Detection For Task Based Dialog** Facebook Conversational AI, Menlo Park, CA  
 Mentors : **Sonal Gupta**, **Arash Einolghozati**, Abhinav Arora May – August 2019
  - Task-based dialog systems on deployment often get user inputs which aren't actually intents pertaining to any domain, such as rhetorical remarks, subjective questions and ill-specified search queries.
  - If not filtered, these inputs can wreak havoc on downstream components like slot detection. Furthermore, it is infeasible to curate training data for these “*OOD*” inputs. Hence, we need unsupervised approaches to detect these at test-time jointly with intent classification.
  - We explore **likelihood ratio** with a **background** likelihood as an alternative to plain likelihood. We find this to consistently improve OOD detection for multiple types of likelihood functions.
  - We propose learning a **generative classifier** and computing a marginal likelihood (ratio) for OOD detection. This outperforms approaches based on simple likelihood as well as discriminative classifiers.
  - **Accepted at AAAI 2020** [C.8]
- **Centrality and Influence in Unconventional Social Networks** IBM Research, India  
 Mentor : **Ramasuri Narayanam** May – July 2015
  - Worked on centrality measures for signed networks [C.1], multiplex networks and hypergraphs [C.2].
- **Building Ad Customer Profile Using ML (CUPRUM)** Bing Ads Team, Microsoft India  
 Mentor : **Prashant Rajoria** May – July 2014
  - Worked on building models of various aspects of Bing advertisers such as spamminess, malware and serveability using supervised machine learning methods, based on features from their ad pages.

## RESEARCH MENTORING EXPERIENCE

---

- **Kevin Lu - Univ of Waterloo Undergrad, CS, Class of 2026, ([C.18] [A.2])** June-Present 2021
- **Dheeraj Mekala - Univ of San Diego Masters, CS, Class of 2021, ([C.16])** Jan-Sept 2021

- |   |                   |
|---|-------------------|
| • Zhuofu Tao - UCLA 1st Year PhD, CS, ([C.16])                  | June-August 2021  |
| • Sedrick Keh - CMU Masters, Machine Learning, Class of 2023    | Sept-Present 2021 |
| • Petros Karypis - UMN Masters, Computer Science, Class of 2022 | Sept-Present 2021 |
| • Steven Y. Feng - CMU Masters, LTI, Class of 2022, ([C.10])    | April-Sept 2020   |

## TEACHING EXPERIENCE

---

- |   |                    |
|---|--------------------|
| • Teaching Assistant - Neural Networks For NLP, Spring 2018 | CMU                |
| • Teaching Assistant - Grammars & Lexicons, Fall 2017       | CMU                |
| • Teaching Assistant - Machine Learning MOOC, Spring 2016   | NPTEL & IIT Madras |
| • Teaching Assistant - Reinforcement Learning, Spring 2016  | IIT Madras         |
| • Teaching Assistant - Machine Learning, Fall 2016          | IIT Madras         |