

# Dheeraj Rajagopal

[HTTPS://SHORTURL.AT/E3wGx](https://shorturl.at/E3wGx)

□ (+1) 412-636-2651 | □ dheeraj.gopal@gmail.com | □ dheerajrajagopal | □ dheerajgopal

## Education

---

### Carnegie Mellon University

Pittsburgh, Pennsylvania

#### PHD IN LANGUAGE TECHNOLOGIES (GPA: 3.85/4)

- Advisor: Prof. Eduard Hovy
- Primary Interests: Deep Learning for NLP, Knowledge Representation and Reasoning, Question and Answering with Background Knowledge, Learning with Partial or Limited Knowledge

September 2017 - July 2022

### Carnegie Mellon University

Pittsburgh, Pennsylvania

#### MASTERS IN LANGUAGE TECHNOLOGIES (GPA: 3.71/4)

- Advisors: Prof. Eduard Hovy and Prof. Teruko Mitamura
- Research Fellowship for the entire course duration (Monthly Stipend + Full tuition Fee Waiver)
- Courses: Algorithms for NLP, Introduction to Machine Learning, Computational Semantics for NLP, Advanced Multimodal Machine Learning, Topics in Deep Learning, Deep Reinforcement Learning, Machine Translation and Sequence-to-Sequence Models, Language and Statistics

August 2015 - July 2017

### VIT University

Vellore, India

#### BACHELORS IN COMPUTER SCIENCE AND ENGINEERING (GPA: 8.7/10)

- Recipient of Scholarship awarded from VIT University for ranking among top 10 in the state

August 2008 - May 2012

## Work Experience

---

### Fastino AI

Seattle, WA

#### MEMBER OF TECHNICAL STAFF

March 2025 - Present

- Leading a team to build proactive ML Agents and benchmarking proactivity. Submitted a paper to ICLR 2026
- Built an inhouse memory system that has better performance than memo using RAG and graph-based memory retrieval
- Currently working on a diffusion architecture for generation with the aim of achieving lower hallucination rates

### Google DeepMind

Seattle, WA

#### RESEARCH SCIENTIST, PEOPLE AND AI RESEARCH TEAM

August 2022 - March 2025

- Building novel interpretability techniques
- Built the TDA (Training Data Attribution) Pipeline to identify most influential training data samples for a prediction
- Impacted in several dataset mixture changes in Gemini and BARD SFT mixtures

### Google Research

Remote

#### GRADUATE RESEARCH INTERN WITH DR. CHUNG-CHING CHANG, DR. SIAMAK SHAKERI

September 2021 - December 2021

- Explored various Augmentation approaches for Factuality
- Implemented two approaches - Entity and Ontology Based
- Improved results by 3 points for Factuality

### Allen Institute for AI

Seattle, WA

#### GRADUATE RESEARCH INTERN WITH DR. NIKET TANDON

May - August 2020

- Created a general purpose procedural model using T5
- Developed a unit testing strategy to evaluate different model capabilities
- Implemented multiple pretraining strategies for procedure understanding

### Allen Institute for AI

Seattle, WA

#### GRADUATE RESEARCH INTERN WITH DR. NIKET TANDON

May - August 2019

- Created a model for explanations with the WIQA dataset
- Developed a framework with explanations as first-class citizens
- Showed that models with explanations learn better than models without

## Microsoft Research

GRADUATE RESEARCH INTERN WITH DR. MICHAEL GAMON

- Created a dataset for understanding the evolution of language of documents over time
- Developed a classification strategy for nature of edits in a document
- Implemented a model that predicts the nature of the edit with about 90% accuracy
- Expanded the model as a semi-supervised approach to classify and summarize edits automatically

Redmond, WA

May - August 2018

## Institute for High Performance Computing-A\*STAR

Singapore

RESEARCH ENGINEER

- Implemented a fully functional Graph-Based Knowledge-Base system using Neo4j
- Built a system for unsupervised Knowledge Based tagging for Social Data
- Investigated different Knowledge Representations for commonsense reasoning in Artificial Intelligence Systems

November 2014 - July 2015

## National University of Singapore - Temasek Laboratories

Singapore

ASSOCIATE SCIENTIST

October 2012 - October 2014

- Research Area: Commonsense Reasoning and its Applications
- Developed tools for enabling commonsense reasoning in Natural Language Processing applications
- Built applications for Information Extraction, Topic Modeling and Document Categorization using Knowledge-bases
- Built the back-end for a game for extracting domain-specific commonsense knowledge (disaster relief and humanitarian response)

## Technical Skills

---

- 1 **Programming Languages** : Python(Advanced), Java(Intermediate), C(Beginner)
- 2 **Deep Learning Libraries** : PyTorch
- 3 **Database Technologies** : Neo4j, MySQL, SQLite
- 4 **Operating Systems** : OSX, Linux, Windows

## Selected Publications

---

### PAPERS

2025	<b>Arxiv Paper</b> : G Pasternak, D Rajagopal, J White, D Atreja, M Thomas, G Hurn-Maloney, A Lewis. Beyond Reactivity: Measuring Proactive Problem Solving in LLM Agents.	Arxiv
2025	<b>Conference Paper</b> : TA Chang, D Rajagopal, T Bolukbasi, L Dixon, I Tenney. Scalable influence and fact tracing for large language model pretraining. ICLR	Singapore
2025	<b>Conference Paper</b> : PQ Da Silva, H Sethuraman, D Rajagopal, H Hajishirzi, S Kumar. Steering off Course: Reliability Challenges in Steering Language Models. ACL	Vienna
2023	<b>Conference Paper</b> : Rajagopal .D., Khetan .V., Sacaleanu .B., Gershman .A., Fano .A., Hovy .E.: Cross-Domain Reasoning via Template Filling. IJCNLP-AACL	Bali
2021	<b>Conference Paper</b> : Rajagopal .D., Balachandran .V., Tsvetkov .Y., Hovy .E.: SelfExplain: A Self-Explaining Architecture for Neural Language Classifiers. EMNLP	Punta Cana
2021	<b>Conference Paper</b> : Madaan .A., Tandon. N., Rajagopal .D., Yang .Y., Hovy .E.: Think about it! Improving defeasible reasoning by first modeling the question scenario. EMNLP	Punta Cana
2021	<b>Conference Paper</b> : Rajagopal .D. *, Madaan .A. *, Tandon. N. *, Yang .Y., Hovy .E.: Could you give me a hint? Generating inference graphs for defeasible reasoning. ACL Findings	Remote
2020	<b>Conference Paper</b> : Tandon .N., Sakaguchi .K., Dalvi .B., Rajagopal .D., Guerquin .M., Richardson .K., Clarke .P., Hovy .E. : A Dataset for Tracking Entities in Open Domain Procedural Text. EMNLP	Remote
2020	<b>Conference Paper</b> : Rajagopal .D., Tandon .N., Clarke .P., Dalvi .B., Hovy .E. : What-if I ask you to explain: Explaining the effects of perturbations in procedural text. EMNLP	Remote
2021	<b>Conference Paper</b> : Balachandran .V., Pagnoni .A., Lee .J.Y., Rajagopal .D., Carbonell .J., Tsvetkov .Y. : StructSum: Incorporating Latent and Explicit Sentence Dependencies for Single Document Summarization. EACL	Remote
2018	<b>Conference Paper</b> : Dhingra .B. *, Pruthi .D. *, Rajagopal .D. *: Simple and Effective Semi-Supervised Question Answering. NAACL [ * = Equal Contribution]	New Orleans, USA

# Projects

---

## Retrofitting Knowledge Bases to Word Embeddings

LTI, CMU

TECHNOLOGIES USED: PYTHON

Jan 2016 - April 2016

- Examined if we could retrofit the KB concepts based on individual relationships to word embeddings
- Observed that some relationships improved and some decreased in the Question and Answering scenario
- Studied the validity of the approach to intrinsically evaluate Knowledge Bases

## Citation Classification and PageRank for research papers

IIIT, Hyderabad

TECHNOLOGIES USED: C++

December 2010

- Extracted citations from a huge set of research papers, classification by “Cue Phrases” using page-rank
- Citations were classified as Comparative, Base or citation based on previous work
- Won the 3rd Best Research Project award

# References

---

### Dr. Eduard Hovy

Carnegie Mellon University

RESEARCH PROFESSOR

hovy@cmu.edu

### Dr. Yulia Tsvetkov

University of Washington

ASSISTANT PROFESSOR

yuliats@cs.washington.edu

### Dr. Peter Clark

Allen Institute for AI

SENIOR RESEARCH MANAGER

peterc@allenai.org