Raghuveer Thirukovalluru

http://raghavlite.github.io
rt195@duke.edu | 413.409.1632

EDUCATION

UMASS AMHERST | MS in Computer Science
Graduated July 2021 | Amherst, Massachusetts
- Cum. GPA: 4.0/4.0

DUKE UNIVERSITY | PhD in Computer Science
2021-Present | Durham, North Carolina
Advisors: Dr. Bhuwan Dhingra
- Cum. GPA: 4.0/4.0

IIT KANPUR | B.Tech in Electrical Engg.
Graduated July, 2016 | Kanpur, India
- Cum. GPA: 8.9 / 10.0

SELECTED PUBLICATIONS & PATENTS

- R. Thirukovalluru, X. Han, B. Dhingra, E. Dinan, and M. Elbayad. Text-guided semantic image encoder. In *Arxiv (Nov 2025)*

- R. Thirukovalluru, R. Meng, Y. Liu, M. Su, P. Nie, S. Yavuz, Y. Zhou, W. Chen, and B. Dhingra. Breaking the batch barrier (b3) of contrastive learning via smart batch mining. In *NeurIPS 2025, (SPOTLIGHT)*

- J. Chen*, R. Thirukovalluru*, J. Wang, K. Luo, and B. Dhingra. Atomic consistency preference optimization for long-form question answering. In *IJCNLP-AACL, 2025 (ORAL)*

- R. Thirukovalluru, Y. Huang, and B. Dhingra. GenEOL: Harnessing the Generative Power of LLMs for Training-Free Sentence Embeddings. In *Findings of the Association for Computational Linguistics: NAACL 2025*

- R. Thirukovalluru, Y. Huang, and B. Dhingra. Atomic self-consistency for better long form generations. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024. (Also ORAL at SouthNLP Symposium)*

- R.Thirukovalluru, X. Wang, J. Chen, S. Li, J. Lei, R. Jin, and B. Dhingra. SumCSE: Summary as a transformation for Contrastive Learning. In *Findings of the Association for Computational Linguistics: NAACL 2024*

- R.Thirukovalluru, N. Monath, B. Dhingra, and S. Wiseman. Sequence Reducible Holdout Loss for Language Model Pretraining. In *Proceedings of the (LREC-COLING 2024)*

- R. Thirukovalluru, N. Monath, K. Shridhar, M. Zaheer, M. Sachan, and A. McCallum. Scaling Within Document Coreference to Long Texts. In *Findings of the Association for Computational Linguistics: ACL 2021*

- R. Thirukovalluru*, M. Sridhar*, D. Thai*, S. Chanumolu, N. Monath, S. Ananthakrishnan, and A. McCallum. Knowledge informed semantic parsing for conversational question answering. In *Proceedings of the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021)*

- D. Thai*, R. Thirukovalluru*, T. Bansal *, and A. McCallum. Simultaneously Self-Attending to Text and Entities for Knowledge-Informed Text Representations. In *Proceedings of the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021)*

- R.Thirukovalluru, R.Mariyappan, and S.Roy. Method and System for Real-time Summary Generation of Conversation. *USPTO*, US9881614

TALKS

- 'B3' lightning talk at **Salesforce** booth, NeurIPS 2025. [Dec. 2025]
- Invited talk for 'B3' at **Microsoft Research India**. [Oct. 2025]
- Presented 'B3' in AI Orgs Reading @ **Meta AI**. [Jul. 2025]

EXPERIENCE

FAIR, META AI | Research Scientist Intern
Summers 2025 | Menlo Park

- Worked on a text guided semantic image encoder.

META AI | Research Scientist Intern
Summers 2023 | Sunnyvale

- Worked on generating unsupervised sentence embeddings for sentence similarity.

INFORMATION EXTRACTION & SYNTHESIS LAB | Research Assistant
Summers 2019 | Amherst

- Worked on multiple high impact NLP problems- Coreference Resolution, Knowledge graphs etc. Details in Research Section.

AI IN FINANCE | Co-Founder
August 2018 - May 2019 | Bangalore (Non-Incorporated)

- Designed and developed a credit card alternative with better creditworthiness estimation, lower NPAs and higher user incentives. Worked on both *pricing & growth models* and also *hardware & software* of payment devices.

XEROX RESEARCH CENTRE, INDIA | Budding Scientist, Text & Graph Analytics
July 2016 – July 2018 | Bangalore

- Worked on multiple research problems from conversation summarization to trending topic analysis.

RESEARCH

MULTIMODAL EMBEDDINGS | Mentor : Dr. Bhuwan Dhingra, Dr. Wenhu Chen
Dec 2024 – May 2025 | Duke University

- *Problem Statement*: Propose a methodology to improve Multimodal embeddings.
- *Approach*: Proposed a graph based batch mining technique to achieve state-of-the-art performance without using any hard negatives.

IMAGE ENCODER | Mentor : Dr. Maha Elbayad, Dr. Xiaochuang Han, Emily Dinan
May 2025 – Nov 2025 | Meta AI, Menlo Park

- *Problem Statement*: Visual encoders are traditionally trained contrastively and lack text grounding.
- *Approach*: We use text-guided supervision during training to shape visual embeddings, yielding a semantic image encoder with improved alignment and VQA performance.

LONG FORM QUESTION ANSWERING | Mentor : Dr. Bhuwan Dhingra
Oct 2023 – May 2024 | Duke University

- *Problem Statement*: Improve overall correctness of LLMs for long form question answering.
- *Approach*: While most LLM generation techniques like self-consistency, self evaluation improve precision of LLMs, we improve recall of generations in this project. We use a clustering based solution over candidate resposes from an LLM.

LANGUAGE MODEL PRETRAINING | Mentor : Dr. Bhuwan Dhingra, Dr. Sam Wiseman
Sept 2021 – April 2023 | Duke University

- *Problem Statement*: Propose a methodology to improve language model pretraining time.
- *Approach*: Proposed an auxiliary model which can filter out some less important examples for the language model at a given state during training. These examples while not contributing much to pretraining of the language model, eat up time.

COURSEWORK

• Machine Learning • Reinforcement Learning • Algorithms for Data Science • Probabilistic Graphical Models • Advanced NLP • Adv. Algorithms • Theory Software Engineering • Software Architecture • Discrete Mathematics • Game Theory

AWARDS/ACHIEVEMENTS

- **Best Preliminary Research Exam, Duke University CS, 2025**.
- Academic Excellence Award, IIT Kanpur.
- All India Rank 26 in Kishore Vaigyanik Protsahan Yojana (KVPY 2012).
- All India Rank 588 (amongst 5,00,000 students) in IIT-JEE 2012.
- All India Rank 87 (amongst 10,00,000 students) in AIEEE.BA 2012.

PROGRAMMING

ML Libraries: TensorFlow • PyTorch
Programming Languages: Python • JS • C++ • Java
Frameworks: Android • NodeJS

EX. CURRICULAR

- **Chess** - Regular player of chess. Rated 1700+(rapid) on Chess.com.
- **Table Tennis** - Regular player. Won multiple office contests.

REVIEWER

• Neurips - '23, '25; AAAI '24, '25
• ARR - Dec '23, Feb '24, June '24, Oct '24
• ICML '23, '24, '25; ICLR '26