

# Nandan Thakur

200 University Ave W, Waterloo, ON, N2L 3G1

✉ [nandan.thakur@uwaterloo.ca](mailto:nandan.thakur@uwaterloo.ca)

🌐 [thakur-nandan.github.io](https://github.com/thakur-nandan)

---

RESEARCH INTERESTS	I develop real-world evaluation benchmarks and curated training datasets to advance search and retrieval-augmented generation (RAG) systems. My research focuses on enabling retrieval models to accurately retrieve and reason over information in niche domains, complex problem settings, and diverse languages.
EDUCATION	<b>Univesity of Waterloo</b> September 2021 – April 2026 Ph.D. in Computer Science Waterloo, ON, Canada Advised by <a href="#">Prof. Jimmy Lin</a> Thesis: “ <i>Benchmarks, Data, and Evaluation for Robust Retrieval and Retrieval-Augmented Generation on Heterogeneous Domains and Languages</i> ” <b>Birla Institute of Technology &amp; Science (BITS) Pilani</b> July 2014 – July 2018 B.E. (Hons.) in Electronics & Instrumentation, Minor in Finance Goa, India
PAST EMPLOYMENT	<b>Databricks &amp; Mosaic Research</b> August – December 2024 <i>Research Intern under <a href="#">Omar Khattab</a> and <a href="#">Prof. Michael Carbin</a></i> Automatic framework for RAG benchmark construction [1]. San Francisco, CA <b>Vectara</b> February – July 2024 <i>Research Intern under <a href="#">Amin Ahmad</a></i> Multilingual RAG benchmarking [7] & LLM hallucinations [9]. Palo Alto, CA <b>Google Research</b> September 2022 – May 2023 <i>Student Researcher under <a href="#">Daniel Cer</a> and <a href="#">Jianmo Ni</a></i> Synthetic construction of multilingual retrieval datasets using LLMs [10]. Mountain View, CA <b>UKP Lab, Technical University of Darmstadt</b> November 2019 – August 2021 <i>Research Assistant under <a href="#">Nils Reimers</a> and <a href="#">Prof. Iryna Gurevych</a></i> Zero-shot IR benchmarking [17] & data augmentation [18] [16]. Darmstadt, Germany <b>KNOLSKAPE</b> September 2018 – October 2019 <i>Data Scientist</i> Constructed Krawler.ai, an enterprise multimodal search product. Bangalore, India <b>(EMBL) European Molecular Biology Laboratory</b> June – August 2018 <i>Research Trainee under <a href="#">Manjeet Kumar</a> and <a href="#">Prof. Toby Gibson</a></i> ML Prediction toolkit to predict phosphorylation sites within protein sequences. Heidelberg, Germany <b>Belong.co</b> July – December 2017 <i>Data Science Intern under <a href="#">Vinodh K. Ravindranath</a></i> Semi-supervised topic modeling and keyword extraction with GuidedLDA. Bangalore, India
SELECTED AWARDS & GRANTS	<a href="#">Tinker Research Grant</a> of \$5,000 API credits 2026 <a href="#">David R. Cheriton Graduate Scholarship</a> of \$20,000 for two academic years 2024 <a href="#">Snowflake AI Research &amp; University of Waterloo Collaborative Grant</a> 2024 <a href="#">Huawei Technologies &amp; University of Waterloo Collaborative Grant</a> 2022 Received a fully-funded ML fellowship to work at <a href="#">EMBL Heidelberg</a> 2018

INVITED TALKS	<p><b>IISc Bangalore</b> <i>Beyond Models: Rethinking Benchmarks, Data, and Evaluation for RAG</i> October 2025 Bangalore, India</p> <p><b>MILA, Hamel Husain's Course &amp; Weaviate Podcast</b> <i>Modern IR Evaluation for RAG (Over 4000+ signups)</i> September 2025 Virtual</p> <p><b>Microsoft Research India</b> <i>Accelerating Multilingual RAG Systems</i> January 2025 Virtual</p> <p><b>IIT Delhi &amp; IIIT Delhi</b> <i>Heterogenous IR Benchmarking across Domains and Languages</i> January 2024 Delhi, India</p> <p><b>Koç University</b> <i>A Tutorial on Advanced Information Retrieval</i> July 2023 Virtual</p> <p><b>Stanford University</b> <i>Heterogenous Benchmarking in IR Research</i> November 2022 Palo Alto, CA</p> <p><b>OpenNLP Meetup, Deepset.ai</b> <i>BEIR, An Open-Source Benchmark for IR Systems</i> August 2021 Virtual</p>
------------------	---

ACADEMIC  
SERVICE

- Lead Organizer of the TREC's inaugural RAG track: [TREC-RAG 2024 & 2025](#).
- Lead Organizer on multilingual IR competition task (MIRACL) at [WSDM Cup 2023](#).
- **Reviewer (NLP conferences)**: ACL ARR Oct-Nov (2021), Jan-Apr (2022), 2024–Present
- **Reviewer (ML & IR conferences)**: NeurIPS 2023, SIGIR 2023 & 2025, ECIR 2024 & 2026, ICLR 2026.

PEER-REVIEWED  
CONFERENCE  
PUBLICATIONS

Check my Google Scholar (<https://scholar.google.com/citations?user=CE9GJoMAAAAJ>) for all of my publications. NeurIPS, ACL, EMNLP, NAACL, TACL, SIGIR etc. are top-tier peer-reviewed conferences in NLP/IR with an acceptance rate between 20-30%. (\* denotes equal contributions.)

- [1] [FreshStack: Building Realistic Benchmarks for Evaluating Retrieval on Technical Documents](#). N. Thakur, J. Lin, S. Havens, M. Carbin, O. Khattab, A. Drozdov. *Conference on Neural Information Processing Systems: Datasets and Benchmarks Track (NeurIPS D&B Track)*. 2025.
- [2] [Hard Negatives, Hard Lessons: Revisiting Training Data Quality for Robust Information Retrieval with LLMs](#). N. Thakur\*, X. Zhang\*, X. Ma, J. Lin. *Findings of the Association for Computational Linguistics (EMNLP Findings)*. 2025.
- [3] [Assessing Support for the TREC 2024 RAG Track: A Large-Scale Comparative Study of LLM and Human Evaluations](#). N. Thakur, R. Pradeep, S. Upadhyay, D. Campos, N. Craswell, J. Lin. *Conference on Research and Development in Information Retrieval (SIGIR)*. 2025.
- [4] [The Great Nugget Recall: Automating Fact Extraction and RAG Evaluation with Large Language Models](#). R. Pradeep, N. Thakur, S. Upadhyay, D. Campos, N. Craswell, J. Lin. *Conference on Research and Development in Information Retrieval (SIGIR)*. 2025.
- [5] [A Large-Scale Study of Relevance Assessments with Large Language Models Using UMBRELA](#). S. Upadhyay, R. Pradeep, N. Thakur, D. Campos, N. Craswell, I. Soboroff, H. T. Dang, J. Lin. *Conference on Innovative Concepts and Theories in Information Retrieval (ICTIR)*. 2025.

- [6] [Ragnarök: A Reusable Framework and Baselines for TREC 2024 Retrieval-Augmented Generation Track](#). R. Pradeep\*, **N. Thakur\***, S. Sharifymoghaddam, E. Zhang, R. Nguyen, D. Campos, N. Craswell, J. Lin.  
*Findings of the European Conference on Information Retrieval (ECIR Findings)*. 2025.
- [7] [MIRAGE-Bench: Automatic Multilingual Benchmark Arena for Retrieval-Augmented Generation Systems](#). **N. Thakur**, S. Kazi, G. Luo, J. Lin, A. Ahmad.  
*Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*. 2025.
- [8] [MMTEB: Massive Multilingual Text Embedding Benchmark](#).  
K. Enevoldsen, I. Chung, ..., **N. Thakur**, ... (80 additional authors).  
*International Conference on Learning Representations (ICLR)*. 2025.
- [9] [“Knowing When You Don’t Know”: A Multilingual Relevance Assessment Dataset for Robust Retrieval-Augmented Generation](#). **N. Thakur**, L. Bonifacio, X. Zhang, O. Ogundepo, E. Kamaloo, D. Alfonso-Hermelo, X. Li, Q. Liu, B. Chen, M. Rezagholizadeh, J. Lin.  
*Findings of the Association for Computational Linguistics (EMNLP Findings)*. 2024.
- [10] [Leveraging LLMs for Synthesizing Training Data Across Many Languages in Multilingual Dense Retrieval](#). **N. Thakur**, J. Ni, G. Hernández Ábrego, J. F. Wieting, J. Lin, D. Cer.  
*Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*. 2024.
- [11] [Systematic Evaluation of Neural Retrieval Models on the Touché 2020 Argument Retrieval Subset of BEIR](#). **N. Thakur**, L. Bonifacio, M. Fröbe, A. Bondarenko, E. Kamaloo, M. Potthast, M. Hagen, J. Lin.  
*Conference on Research and Development in Information Retrieval (SIGIR Resource & Reproducibility Track)*. 2024. **Oral Presentation**.
- [12] [Resources for Brewing BEIR: Reproducible Reference Models and Statistical Analyses](#).  
E. Kamaloo, **N. Thakur**, C. Lassance, X. Ma, J. H. Yang, J. Lin.  
*Conference on Research and Development in Information Retrieval (SIGIR Resource & Reproducibility Track)*. 2024. **Oral Presentation**.
- [13] [MIRACL: A Multilingual Retrieval Dataset Covering 18 Diverse Languages](#).  
X. Zhang\*, **N. Thakur\***, O. Ogundepo, E. Kamaloo, D. Alfonso-Hermelo, X. Li, Q. Liu, M. Rezagholizadeh, J. Lin.  
*Transactions of the Association for Computational Linguistics (TACL)*. 2023
- [14] [SPRINT: A Unified Toolkit for Evaluating and Demystifying Zero-shot Neural Sparse Retrieval](#). **N. Thakur**, K. Wang, I. Gurevych, J. Lin.  
*Conference on Research and Development in Information Retrieval (SIGIR Resource & Reproducibility Track)*. 2023.
- [15] [Evaluating Embedding APIs for Information Retrieval](#)  
E. Kamaloo, X. Zhang, O. Ogundepo, **N. Thakur**, D. Alfonso-Hermelo, M. Rezagholizadeh, J. Lin.  
*Annual Conference of the Association for Computational Linguistics (ACL Industry Track)*. 2023.
- [16] [GPL: Generative Pseudo Labeling for Unsupervised Domain Adaptation of Dense Retrieval](#) K. Wang, **N. Thakur**, N. Reimers, I. Gurevych.  
*Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*. 2022.

[17] [BEIR: A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models](#). **N. Thakur**, N. Reimers, A. Rücklé, A. Srivastava, I. Gurevych. *Conference on Neural Information Processing Systems: Datasets and Benchmarks Track (NeurIPS D&B Track)*. 2021. [Widely used benchmark by OpenAI, Google, IBM, Microsoft, Elastic Search, Cohere and many others. Over 2K GitHub stars & 1.5K citations.](#)

[18] [Augmented SBERT: Data Augmentation for Improving Bi-Encoders for Pairwise Sentence Scoring Tasks](#). **N. Thakur**, N. Reimers, J. Daxenberger, I. Gurevych. *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*. 2021.

WORKSHOP  
PUBLICATIONS

[19] [BrowseComp-Plus: A More Fair and Transparent Evaluation Benchmark of Deep-Research Agent](#). Z. Chen, X. Ma, ..., **N. Thakur**, ..., J. Lin. *Workshop on Multi-Turn Interactions in Large Language Models (MTI-LLM) @ NeurIPS 2025*. [Spotlight Presentation](#).

[20] [Injecting Domain Adaptation with Learning-to-hash for Effective and Efficient Zero-shot Dense Retrieval](#). **N. Thakur**, N. Reimers, J. Lin. *Workshop on Reaching Efficiency in Neural Information Retrieval (ReNeuIR) @ SIGIR 2023*. [Oral Presentation](#).

ARXIV  
PREPRINTS

[21] [ORBIT: Scalable and Verifiable Data Generation for Search Agents on a Tight Budget](#). **N. Thakur**, Z. Chen, X. Ma, J. Lin. 2026 (under review).

[22] [Overview of the TREC 2025 Retrieval Augmented Generation \(RAG\) Track](#). S. Upadhyay, **N. Thakur**, R. Pradeep, N. Craswell, D. Campos, J. Lin. 2026.

[23] [Still Fresh? Evaluating Temporal Drift in Retrieval Benchmarks](#). N. Kuissi, S. Subrahmanyam **N. Thakur (Mentor)**, J. Lin. 2026 (under review).

[24] [Chatbot Arena Meets Nuggets: Towards Explanations & Diagnostics in the Evaluation of LLM Responses](#) S. Sharifymoghaddam\*, S. Upadhyay\*, **N. Thakur\***, R. Pradeep, J. Lin. 2025 (under review)

[25] [UMBRELA: UMBrela is the \(Open-Source Reproduction of the\) Bing RElevance Assessor](#) S. Upadhyay, R. Pradeep, **N. Thakur**, N. Craswell, J. Lin. 2024.

[26] [A Human-LLM Collaborative Dataset for Generative Information-Seeking with Attribution](#). E. Kamaloo, A. Jafari, X. Zhang, **N. Thakur**, J. Lin. 2023.

[27] [Simple Yet Effective Neural Ranking and Reranking Baselines for Cross-Lingual Information Retrieval](#). J. Lin, D. Alfonso-Hermelo, V. Jeronymo, E. Kamaloo, C. Lassance, R. Nogueira, O. Ogundepo, M. Rezagholizadeh, **N. Thakur**, X. Zhang. 2023.

TEACHING  
EXPERIENCE

**Head TA** at University of Waterloo

- CS 370 Numerical Computation Winter, Spring & Fall 2025,  
Fall 2023, Spring 2024
- CS 486/686 Introduction to Artificial Intelligence Winter 2023
- CS 116 Introduction to Computer Science 2 Winter 2024
- CS 136 Elementary Algorithm Design Spring 2023, Winter 2022
- CS 241 Foundations of Sequential Programs Spring 2022
- CS 135 Designing Functional Programs Fall 2021

ADVISING & MENTORING	<p><b>Masters/Ph.D. Students:</b></p> <ul style="list-style-type: none"> <li>• Linsen Gao (open source reproduction of SWIM-IR [10])</li> </ul> <p><b>Undergraduate Students:</b></p> <ul style="list-style-type: none"> <li>• Nathan Kuissi, Suraj Subrahmanyam (Mentored Temporal Drift in FreshStack [23])</li> <li>• Jonathan Zhao (extension of GPL [16])</li> </ul>
PRESS & MEDIA	<p>2025: <a href="#">FreshStack: Honourable Mention for Best 2025 Search Project</a> <span style="float: right;"><i>BCS</i></span></p> <p>2025: <a href="#">FreshStack has been included in the RTEB benchmark</a> <span style="float: right;"><i>Hugging Face</i></span></p> <p>2025: <a href="#">Modern RAG evaluation is part of the Hamel Husain's RAG mini book</a> <span style="float: right;"><i>Maven</i></span></p> <p>2025: <a href="#">Invited Guest Speaker (#124) at Weaviate Podcast</a> <span style="float: right;"><i>Weaviate</i></span></p> <p>2025: <a href="#">TREC RAG research has been incorporated in Vectara's product</a> <span style="float: right;"><i>Vectara</i></span></p> <p>2023: <a href="#">Domain Adaptation with Generative Pseudo-Labeling (GPL)</a> <span style="float: right;"><i>Pinecone.ai</i></span></p> <p>2022: <a href="#">Extending Neural Retrieval Models to New Domains and Languages</a> <span style="float: right;"><i>Zeta Alpha</i></span></p> <p>2022: <a href="#">BEIR benchmark in Stanford's CS224U Teaching Material</a> <span style="float: right;"><i>Stanford University</i></span></p> <p>2022: <a href="#">BEIR benchmark as a helpful ML library</a> <span style="float: right;"><i>ML News by Yannic Kilcher</i></span></p> <p>2021: <a href="#">Making the Most of Data: Augmentation with BERT</a> <span style="float: right;"><i>Pinecone.ai</i></span></p>
COURSEWORK	<p><b>University of Waterloo (2021–Present):</b> CS 680: Introduction to Machine Learning, CS 889: Data Sources for Emerging Technology, CS 886: Graph Neural Networks, CS 886: Robustness of Machine Learning, CS 848: Information Retrieval, CS 679: Neural Networks, CS 854: Experimental Performance Evaluation, CS 649: Human-Computer Interaction.</p> <p><b>BITS Pilani (2014–2018):</b> Machine Learning, Neural Networks &amp; Fuzzy Logic, Data Structures &amp; Algorithms, Probability &amp; Statistics, Linear Algebra, Econometric Methods, Discrete Mathematics.</p>
CO-CURRICULAR	<p><a href="#">Mime Club Coordinator</a>, BITS Pilani <span style="float: right;">2016–2017</span></p> <p>Led a team of 30 student performers. Involved in acting, sound mixing, designing slides and creating stories for more than 10 shows over a span of 4 years. (<a href="#">YouTube</a>)</p>
REFEREES	<p><a href="#">Prof. Jimmy Lin</a>, Full Professor, University of Waterloo</p> <p><a href="#">Prof. Iryna Gurevych</a>, Full Professor, TU Darmstadt; Adjunct Professor, MBZUAI</p> <p><a href="#">Omar Khattab</a>, Assistant Professor, MIT</p> <p><a href="#">Daniel Cer</a>, Senior Research Scientist, Google Research</p> <p><a href="#">Nick Craswell</a>, Principal Architect, Microsoft</p>