Bhavik Agarwal

California, United States | +1 (667) 899-1927 | bhavik2506@gmail.com
linkedin.com/in/bhavik-agarwal-7a5432178/ | Google Scholar | Website

I build small, tool-using language models with multi-agent and knowledge-graph retrieval for reliable, verifiable NLP in regulated domains.

Research Interests

Small language models and tool-using LMs; multi-agent orchestration; retrieval-augmented and knowledge-graph–augmented generation; verifiable structured (JSON/program) outputs for NLP in regulated domains.

Education

Johns Hopkins University **Aug 2023 – May 2025**
Masters, Computer Science Engineering (Thesis) — *GPA:* ***3.85/4.0*** *Baltimore, MD*

IIIT-Delhi **Aug 2018 – Apr 2022**
Bachelors, Computer Science Engineering, Biosciences Engineering — *GPA:* ***3.80/4.0*** *New Delhi, India*

Research Experience

MasterControl **Fall 2024 – Present**
Machine Learning Research Scientist, Advisor: Dr. Viktoria Rojkova *United States*

- First author, **RAGulating Compliance** (International Semantic Web Conference 2025): triplet-first knowledge graph and multi-agent retrieval for evidence-grounded regulatory QA (basis for U.S. Patent App. 19/355,798) and invited talk/poster at PDA Good Digital Manufacturing Conference (2025).
- First author, **ThinkJSON** (Hugging Face Spotlight Paper, Under review at EACL 2025): multi-reward GRPO for strict JSON with separated `<think>` and `<answer>` modes; a 7B+LoRA model achieves **82%** strict-parse success on noisy JSON-schema benchmarks on a single A100 GPU.
- Designed and deployed budget-aware orchestration with fallback policies and response auditing for regulated enterprise workloads.

Microsoft Research & Johns Hopkins University **Fall 2023 – Spring 2025**
Graduate Research Assistant, Advisor: Prof. Casey Overby Taylor *Baltimore, MD*

- Built an Azure-based RAG pipeline for genomics/pharmacogenomics using token-aware chunking, a provenance store, Azure Cognitive Search, and Azure OpenAI; evaluated vector-RAG, keyword-RAG, and GPT-4o baselines for patient-centric QA.
- This work forms the basis of the Master's thesis; tuition fully funded in the final two semesters.

Google DeepMind **Fall 2023 – Spring 2024**
Research Mentee, Advisor: Dr. Esteban Real *California, United States*

- Studied compact vision encoders and attention budget allocation (self/cross-attention) for small models and applied these ideas to tool-using LMs and dense RAG retrievers.

SAP Labs **Summer 2024**
AI/ML Scientist Intern, Generative AI (SAP Joule) *California, United States*

- Developed multi-agent embedding-space API orchestration: domain clustering with leader election and two-stage routing using cosine similarity and a reasoning LLM, improving internal task-success rate from **42%** to **78%**.
- First-author U.S. Patent App. 19/022,406 on multi-agent API orchestration; collaborated with NVIDIA and Mistral on API-calling constraints.

Master's Thesis

Bhavik Agarwal. "Design and Evaluation of GPT and RAG-Based QA Systems for Genomic and Pharmacogenomic Patient Queries." Master's Thesis, *Johns Hopkins University*, 2025. Advisor: **Casey Overby Taylor**. LLM-based RAG evaluation for genomics/pharmacogenomics question answering from a patient perspective. [link]

Open Source

Released **13+** models with cumulative **4M+** downloads on Hugging Face. Orgs: MasterControlAIML; username: bhaviktheslider.

Publications

Bhavik Agarwal, Hemant Sunil Jomraj, Simone Kaplunov, Jack Krolick, and Viktoria Rojkova. "RAGulating Compliance: A Multi-Agent Knowledge Graph for Regulatory QA." *International Semantic Web Conference (ISWC)*, 2025. [paper]

Bhavik Agarwal, Ishan Joshi, and Viktoria Rojkova. "Think Inside the JSON: Reinforcement Strategy for Strict LLM Schema Adherence." *Hugging Face Spotlight Paper* [spotlight] *(Under review at EACL, 2025)*. [paper]

Natalie Wang, Sukrit Treewaree, Ayah Zirikly, Yuzhi L. Lu, Michelle H. Nguyen, **Bhavik Agarwal**, Jash Shah, James Michael Stevenson, and Casey Overby Taylor. "Taxonomy-based prompt engineering to generate synthetic drug-related patient portal messages." *Journal of Biomedical Informatics*, vol. 160, 104752, 2024. [paper]

Nimisha Malik and **Bhavik Agarwal**. "Time Series Nowcasting of India's GDP with Machine Learning." In *Proceedings of the 2022 International Conference on Artificial Intelligence of Things (ICAIoT)*, Istanbul, Turkey, 2022. [paper]

Aryaman Babbar and **Bhavik Agarwal**. "Exploration of Negative Mass as a Harbinger of Dark Energy." *IOSR Journal of Applied Physics*, vol. 12, no. 5, pp. 53–59, 2020. [paper]

Patents (U.S. Applications)

Bhavik Agarwal, Julien Yu, Anil Babu Ankisettipalli, and Sebastien Schrieber. "Multi-Agent Embedding-Space API Orchestration for Enterprise Assistants." U.S. Patent Application 19/022,406. Originated at SAP.

Bhavik Agarwal, Hemant Sunil Jomraj, Simone Kaplunov, Jack Krolick, and Viktoria Rojkova. "Multi-Agent Knowledge Graph Retrieval with Evidence Grounding." U.S. Patent Application 19/355,798. Originated at MasterControl.

Teaching

Johns Hopkins University **Spring 2024 – Spring 2025**
Lead Teaching Assistant *Baltimore, MD*
– Machine Learning (Spring 2025); Deep Learning with Unstructured Data (Spring 2024); Big Data Machine Learning (Spring 2024).

IIIT-Delhi **Spring 2021 – Fall 2021**
Teaching Assistant *New Delhi, India*
– Statistical Machine Learning (Fall 2021); Biophysics (Spring 2021).

Mentorship

Johns Hopkins University **Fall 2024 – Spring 2025**
Graduate Mentor, M.S. CS Cohort *Baltimore, MD*
– Mentored **8** incoming M.S. students on research scoping, reproducibility, and internship preparation.

IIIT-Delhi **Fall 2021 – Spring 2022**
Undergraduate Mentor, CS Mentoring Program *New Delhi, India*
– Advised **5** mentees on projects and research skills; held weekly office hours and interview preparation sessions.

Industry Experience

BlueStacks / now.gg **Jul 2022 – Jul 2023**
Senior AI/ML Engineer (Advisor: Dr. Rosen Sharma) *Gurgaon, India*
– Led a team of 10 to deploy LSTM and N-BEATS models for global traffic forecasting, improving capacity planning and infrastructure utilization (multi-million USD annual savings) while maintaining **99%** availability.

Western Digital (SanDisk) **May 2021 – Jun 2021**
Machine Learning Firmware Engineering Intern *Remote*
– Implemented telemetry-driven reliability checks and ML-assisted firmware validation; contributed scripts and analyses to accelerate testing.

Honors & Awards

ACM Tapia Scholar (Richard Tapia Celebration of Diversity in Computing), Fall 2025.

Google DeepMind CS Research Mentorship Program (two semesters), Fall 2023, Spring 2024.

Dean's List of Academic Excellence, IIIT-Delhi, Spring 2022.

Top **1%** in Joint Entrance Examination (IITJEE - India) among 1.5M candidates.

Only master's student invited to speak about the program experience on the Johns Hopkins YouTube channel, Spring 2025.