Ganesh Chandrasekar

cb.ganesh666@gmail.com +1 (438)-921-8453 https://www.linkedin.com/in/cbsag/
https://github.com/cbsag portfolio-cbsag-2026.vercel.app

PROFILE

Machine learning, NLP, and AI engineer with experience in fine-tuning, building, and evaluating transformer-based classification, information retrieval, RAG, and LLM systems. My work combines applied AI research with software engineering and has contributed to shared-task systems, model evaluation workflows, publications, and practical analysis tools.

EDUCATION

Master of Computer Science (Thesis) 09/2023 – 07/2026
Concordia University Montreal, Canada
Thesis: *Evidence-Grounded Biomedical Question Answering: Design, Analysis, and Evaluation*
Expected completion: July 2026

Bachelor of Engineering in Computer Science and Engineering 06/2018 – 04/2022
KCG College of Technology, Anna University Chennai, India
CGPA: **8.81/10**, First Class with Distinction

PROFESSIONAL EXPERIENCE

Research Assistant 09/2023 – Present
CLaC Lab, Concordia University Montreal, Canada

Applied NLP and Machine Learning

- Developed classification, sequence-labeling, span-prediction, and information extraction systems using Python, PyTorch, Transformers, CRFsuite, and scikit-learn, including hierarchical and PPO-based approaches for class-imbalanced negation detection.
- Collaborated with Concordia journalism researchers to build a named-entity extraction pipeline for identifying people, organizations, and locations in news articles.

Retrieval, RAG, and LLM Systems

- Designed sparse and dense retrieval pipelines using BM25, Pyserini/Lucene, MedCPT, FAISS, cross-encoder reranking, and TF-IDF/MMR.
- Processed a 26.8-million-document PubMed corpus and optimized quantized Qwen inference on GPU-based HPC infrastructure using vLLM and Slurm.
- Developed shared-task baselines, model evaluation workflows, and analysis dashboards for measuring answer accuracy, retrieval quality, evidence grounding, and citation reliability, contributing to two published question-answering systems.

Teaching Assistant 09/2024 – 05/2026
Concordia University Montreal, Canada

- Supported undergraduate and graduate courses in COMP 479 (Information Retrieval and Web Search) and COMP 232 (Discrete Mathematics) through tutorials, office hours, assignment guidance, grading, and technical concept clarification.

Solution Developer 07/2022 – 08/2023
OrangeScape Technologies Pvt. Ltd Chennai, India

- Developed application features and workflow automation for Kissflow Procurement Cloud, supporting purchasing, approvals, supplier management, and source-to-pay processes.
- Designed workflow rules, forms, validations, and approval paths from business requirements using Kissflow's low-code application platform.
- Built API services on Google Cloud Platform and collaborated with backend and product teams to integrate procurement data, deliver tested features, and resolve workflow issues.

Software Engineering Intern 05/2022 – 07/2022
Basik Marketing Pvt. Ltd. Chennai, India

- Developed a real-time esports broadcast overlay using Python and Node.js, with backend logic that dynamically updated match and event counters during live streams.

TECHNICAL SKILLS

- **Languages:** Python, JavaScript, Java, SQL
- **ML and NLP:** PyTorch, Hugging Face Transformers, scikit-learn, Pandas, CRFsuite, Sequence Labeling, Named Entity Recognition, Reinforcement Learning, PPO
- **LLMs and Retrieval:** RAG, RLHF, RLAIF, LLM-as-a-Judge, Human-in-the-Loop Evaluation, BM25, Pyserini/Lucene, MedCPT, FAISS, Cross-Encoder Reranking, TF-IDF/MMR, vLLM
- **Software and Knowledge Systems:** Node.js, NestJS, Flask, React, REST APIs, RDF, SPARQL, LangChain
- **Infrastructure:** Google Cloud Platform, AWS, Docker, Linux, Slurm, Git

PROJECTS

TREC BioGen: Evidence-Grounded RAG and LLM Evaluation System | Dashboard
- Built and evaluated multiple sparse and dense RAG pipelines over 26.8 million PubMed documents using BM25, Pyserini, MedCPT, cross-encoder reranking, TF-IDF/MMR, and Qwen.
- Applied human and AI feedback in RLHF/RLAIF feedback loops to refine query reformulation, retrieval ranking, and answer generation strategies.

Biomedical Negation Detection with Transformers, CRF, and PPO
- Fine-tuned BERT, BioBERT, and ClinicalBERT for negation cue and scope detection using token-level BIO sequence labeling.
- Built CRF and hierarchical classification pipelines to model label dependencies and address class imbalance in biomedical text.
- Trained a PPO-based sequence-labeling policy using rewards derived from human-annotated negation spans, improving minority-label detection over the supervised baseline.

Agentic and Non-Agentic Multi-Hop Medical Question Answering
- Developed three multi-hop question-answering systems combining open-source LLMs with Wikipedia and PubMed retrieval.
- Implemented an agentic workflow using smolagents and Qwen2.5-Coder-32B, alongside controlled non-agentic pipelines for evidence retrieval and answer extraction.
- Compared short-answer accuracy, evidence use, and reliability; the Wikipedia-and-PubMed pipeline performed best on the development set, with results published at BioCreative IX, IJCAI 2025.

Insomnia Classification and Evidence Extraction | SMM4H-HeaRD Shared Task
- Collaborated with shared-task organizers to build baseline systems for identifying insomnia cases and supporting evidence spans in clinical notes.
- Evaluated Llama 3 8B/70B, Qwen2.5-32B-GPTQ, and Qwen3-8B for patient-level classification and evidence extraction.

Knowledge-Graph and Hybrid Retrieval for Course Enrollment
- Built a question-answering system for Concordia course enrollment using RDF triples and SPARQL to retrieve course, prerequisite, and program information.
- Integrated LangChain and FAISS to combine structured knowledge-graph queries with semantic vector retrieval.
- Compared graph-based, vector-based, and hybrid retrieval approaches for answering natural-language enrollment questions.

PUBLICATIONS

Modular Sparse and Dense Retrieval for Evidence-Constrained Biomedical Question Answering 2025
G. Chandrasekar, B. L. Follo, A. Vinokhodov, and S. Bergler. TREC 2025, BioGen Track, NIST, 2025.

Agentic and Non-Agentic Multi-Hop Systems for Medical Question Answering 2025
G. Chandrasekar, H. G. Saisudha, and S. Bergler. BioCreative IX Workshop at IJCAI, 2025. DOI: 10.5281/zenodo.16875905.

Exploration of Performance of Dynamic Branch Predictors Used in Mitigating Cost of Branching 2022
G. Chandrasekar et al. ICICICT, 2022. DOI: 10.1109/ICICICT54557.2022.9917915.