Bhiman Kumar Baghel

+1 (412) 844 1085 | 130 N Bellefield, Pittsburgh, PA, USA, 15213 | bkb45@pitt.edu | bhimanbaghel.github.io
linkedin.com/in/bhiman-kumar-baghel | Google Scholar

Professional Summary

Ph.D. researcher and former Samsung Lead NLP Engineer with 7+ years building production-grade AI systems. Focused on interpretable, parameter-efficient reasoning in LLMs, with expertise in model editing, LoRA/PEFT, and scalable ML pipelines (AWS, Docker). Developed a framework that improves editing reliability by 38 points. Published 5 peer-reviewed papers, hold 4 patents, and deployed smart-home AI to 100M+ devices. Seeking an Applied Scientist / ML Engineer role to bridge foundational reasoning research with real-world impact.

Skills

ML / LLM Frameworks: PyTorch, Hugging Face Transformers, PEFT / LoRA, vLLM, PyTorch Lightning, TensorFlow
Cloud & MLOps: AWS (EC2, S3), Docker, MLflow, Weights & Biases, Git + CI/CD
Languages & Databases: Python, C++, Pandas, NumPy, SQL, MongoDB, Neo4j

Education

University of Pittsburgh, PA, USA August 2023 – April 2027
PhD, Computer Science (GPA: 3.57/5)

Indian Institute Of Technology, Kharagpur, India July 2017 – May 2019
Master's, Computer Science (GPA: 8.82/10)

Professional Experience

Incoming Applied Scientist Intern, Amazon – Seattle, WA, USA Fall 2025

- Selected for a competitive internship with the People eXperience and Technology Central Science (PXTCS) team.
- Internship will focus on Generative AI applications and infrastructure with emphasis on enterprise knowledge integration, while ensuring fairness and privacy are maintained.

Graduate Research Assistant, University Of Pittsburgh – PA, USA August 2024 – Present

- Engineered a plug-and-play iterative editing pipeline that enhanced edit-success rate by 38 percentage points over prior SOTA on LLaMA-3/2 and GPT-J, enabling rapid knowledge updates without full-model fine-tuning.
- Developed a Shapley- and cartography-based framework to identify influential training examples, revealing key differences in generalization behavior of LoRA on legal reasoning tasks compared to other tuning methods.
- Conducted a gender-bias audit of GPT-3.5 and BART summaries over 19,579 student reflections; used Jensen–Shannon divergence to reveal a 10% male-topic skew and uncovered under-represented female topics.
- Built a 2,900-meme multimodal dataset; my manual audit revealed stereotype bias in 40% of LLaVA and MiniGPT-4 explanations, traced to visual/named-entity stereotypes, and text–image representation imbalance.

Lead NLP Engineer, Samsung Research – Bangalore, India June 2019 – August 2023

- Spearheaded CoSMIC, a BERT-based multi-intent NLU engine for SmartThings; shipped to 100 M + devices, reaching 96% intent accuracy and cutting live NLU errors by 67%.
- Localized and scaled CoSMIC for the Korean market, mentoring a cross-site team and re-engineering tokenization to lift intent-slot F_1 by 25%.
- Architected production conversational-AI models (intent, slot, OOD) that raised multi-intent F_1 from 87% $\rightarrow$ 92% and achieved 90% OOD recall across all public benchmarks.

Machine Learning Intern, IBM – Bangalore, India May 2018 – July 2018

- Prototyped an LSTM-based anomaly-prediction engine that monitors 33 infrastructure health metrics and launches auto-remediation scripts, forecasting critical failures with 97% precision.

Projects

Chat-Enabled AI Agent for Multi-Step Flight Search Demo

- Engineered a modular framework that lets GPT-4o reason over BrowserGym observations and user goals, solving

multi-step flight-search tasks, demonstrating temporal & spatial reasoning for real-world UI automation.

Automatic Concept-Map Generation from Wikipedia Github Link

- Designed an NLP pipeline (PySpotlight, FastText, Stanford CoreNLP) that extracts entities & semantic relations, rendering interactive concept maps that compress 10 K-word articles into 50 node graphs.

Publications

Resolving UnderEdit & OverEdit with Iterative & Neighbor-Assisted Model Editing EMNLP Findings *(2025)*
Bhiman Kumar Baghel, Scott M. Jordan, Zheyuan Ryan Shi, Xiang Lorraine [PDF]

A Fairness Analysis of Human and AI-Generated Student Reflection Summaries GeBNLP Workshop, ACL *(2024)*
Bhiman Kumar Baghel, Arun Balajiee Lekshmi Narayanan, Michael Miller Yoder [PDF] [Talk]

Multimodal Understanding of Memes with Fair Explanations MULA Workshop, CVPR *(2024)*
Yang Zhong, ***Bhiman Kumar Baghel*** [PDF] [Talk]

Intent–Focused Semantic Parsing and Zero-Shot Learning for Out-of-Domain Detection in Spoken Language Understanding IEEE Access *(2021)*
Niraj Kumar, ***Bhiman Kumar Baghel*** [PDF]

Smart Stacking of Deep Learning Models for Granular Joint Intent–Slot Extraction for Multi-Intent SLU IEEE Access *(2021)*
Niraj Kumar, ***Bhiman Kumar Baghel*** [PDF]

Patents

Method and system for time-based personalization management in multi-device environment WO2025018568A1 *(2024)*
Sourabh Tiwari, ***Bhiman Kumar Baghel***, Jalaj Sharma, Manish Chauhan, Boddu Venkata Krishna Vinay, Syed Khaja Moinuddin [Link]

Methods and systems for enabling seamless indirect interactions US18517995 *(2023)*
Venkata Krishna Boddu Vinay, ***Bhiman Kumar Baghel***, Gorang Maniar, Syed Khaja Moinuddin, Sudhansu Ranjan Acharya [Link]

Method and system for mitigating physical risks in an IoT environment US18202687 *(2023)*
Niraj Kumar, ***Bhiman Kumar Baghel*** [Link]

Methods and systems for determining missing slots associated with a voice command for an advanced voice interaction US17835387 *(2023)*
Niraj Kumar, ***Bhiman Kumar Baghel*** [Link]

Honors & Awards

Samsung High Performance Bonus (3×), Samsung Research – Bangalore, India 2023

Samsung Excellence Award (5×), Samsung Research – Bangalore, India 2023
Recognized for SmartThings CLab innovation finalist and 4 US A1 patent filings.

2nd Runner-Up, Audience Poll, IBM Extreme Blue Expo – Bangalore, India 2018
Voted top-3 of 24 projects by 100+ expo attendees.

Academic Service

Program Committee Member, Explainable Automated Software Engineering (ExASE) Workshop, ASE 2025

Reviewer, Multimodal Learning and Applications (MULA) Workshop, CVPR 2025

Reviewer, Gender Bias in NLP (GeBNLP) Workshop, ACL 2025

Reviewer, Representation Learning for NLP (RepL4NLP) Workshop, NAACL 2025