

# Ganga Meghanath

New York, NY | [linkedin.com/in/gangamegha/](https://www.linkedin.com/in/gangamegha/) | [gangamegha.github.io](https://github.com/GangaMegha) | [github.com/GangaMegha](https://github.com/GangaMegha)

## SUMMARY

Data Scientist with 4.5 years of experience building AI and LLM-based systems for security at Amazon and developing experimentation and causal inference frameworks at Microsoft.

**Research Interests:** AI for Security · LLM Retrieval Systems · Machine Learning · Causal Inference · Experimentation Platforms

## WORK EXPERIENCE

### Data Scientist — Proactive Security Organization

Amazon

May 2025 – Present | AI/LLM, Security, Retrieval, Classification

New York, NY

- Built a **CWE Tagger** to classify ~10K monthly security findings into relevant CWE categories using a retrieval–ranking–classification pipeline on **Bedrock**; combined **Cohere** embeddings for retrieval and reranking with **Claude Sonnet** / **Haiku** for classification and critique, achieving ~80% precision and ~95% top-10 retrieval accuracy for CWE categories in scope (~940 total); benchmarked **Llama**, **Palmyra**, **Nova**, **DeepSeek R1** and sentence-transformer baselines.
- Designed a **Review Priority Score** statistical framework integrating business impact, vulnerability likelihood, and prior certification using a non-linear weighted model for application triage.

### Data Scientist — Windows Experimentation

Microsoft

Feb. 2023 – Apr. 2025 | Causal Analysis, A/B Testing, Machine Learning

Redmond, WA

- **Windows 11 2024 rollout (GE/GA):** Led causal-discovery and pattern-mining on upgrade telemetry to identify high-risk device configurations (rollbacks, bluescreens); produced a **5M + 40M device blocklist** and collaborated with engineering to remediate critical issues before release.
- Developed a **Double Machine Learning (DML)** pipeline in Azure ML for end-to-end causal inference with automated feature transformation and imputation for reproducible analyses.
- Served as **causal methodology consultant**, analyzing TSCOOBE and ChurnToCompete signals to identify causal drivers of engagement and retention.
- Built a **Power BI dashboard** to query and visualize historical Windows experiment scorecard data, enabling experimenters to see which tests moved key business metrics, by how much, and in which direction—driving data-informed hypothesis generation and improving visibility into experiment impact.
- Developed the **WExp CoPilot (Chatbot)** PoC—an internal RAG assistant (GPT-Turbo + Ada, Semantic Kernel + Azure Cognitive Search) to answer experimentation FAQs; delivered PoC and production plan.

### Graduate Research Assistant — Computational Neuroscience Lab

Purdue University

May 2021 – Dec. 2022 | Neural Modeling, NLP, Deep Learning

West Lafayette, IN

- Modeled macaque motor-cortex neural population dynamics using RNN, Transformer, and **TERN (Transformer Encoder over RNN)** architectures on Neural Latents Benchmark datasets—achieving state-of-the-art inference of unobserved neuron activity, contributing to top leaderboard results, and published in *Journal of Neural Engineering* (2023).
- Built a **seq2seq GRU-based neural decoder** for ECoG speech data achieving ~3% Word Error Rate.

### Data & Applied Scientist — Bing Ads

Microsoft

Jun. 2019 – Dec. 2020 | Computer Vision, NLP, Data Mining, Anomaly Detection

Bengaluru, India

- Developed compromised-account detection models using domain embeddings, reputation, and index scores to proactively flag malicious advertiser accounts.
- Built scalable data-mining pipelines labeling 1 in  $10^6$  ads across sensitive categories (Adult, Weapon, Tobacco, etc.), reducing manual labeling effort; deployed **MobileNet**, **EfficientNet**, **BERT** models for editorial policy violation detection.

## ADDITIONAL SELECTED PROJECTS

---

- Credit Assignment Using Reviews** | *Prof. Gao; NLP, Sentiment Analysis, PyTorch* Nov. 2021 – Dec. 2021
- Developed sentiment analysis models to capture positive and negative sentiments associated with Amazon product reviews.
  - Compared models built using count-based word vectors versus pre-trained sentence embeddings and transformer-based models.
  - Attributed credit to product features influencing product success or failure using customer reviews. ([Report](#))
- Identify Political Framing from Tweets** | *Prof. Goldwasser; NLP, Data Mining* Apr. 2021 – May. 2021
- Encoded 3.2M tweets from US politicians using Distilled RoBERTa embeddings and sentence similarity metrics.
  - Annotated and mined unlabelled tweets for framing patterns, improving macro F1-score by 15.5% and micro F1-score by 10.8%. ([Report](#))

## SELECTED UNDERGRADUATE PROJECTS — IIT MADRAS

---

- Modeling Ecological Populations** — Developed an N-player Hawk–Dove game in Python to study evolutionary dynamics; analyzed equilibrium convergence and demonstrated cooperative codes of conduct yielding higher collective payoffs.
- Memory-Based Multi-Task A3C** — Proposed and implemented a memory-augmented multi-task reinforcement learning framework in TensorFlow using ALE; empirically showed reduced catastrophic forgetting across Atari games.
- Hierarchical Multi-Tasking** — Extended Feudal and Option-Critic RL architectures by integrating active sampling strategies (UCB, Doubling DQN) for hierarchical policy learning in OpenAI Gym.
- Weather Data Summarizer** — Designed an encoder–decoder BiLSTM with attention for automatic summarization of weather tables; built a Malayalam text corpus and trained word embeddings to evaluate multilingual performance.
- Team Anveshak (URC 2017)** — Engineered robotic control and vision modules in ROS and OpenCV for a Mars rover prototype under Prof. T. Asokan; ranked 29th of 70+ global teams in the University Rover Challenge.

## EDUCATION

---

- |                                                                                                    |                                                           |
|----------------------------------------------------------------------------------------------------|-----------------------------------------------------------|
| <b>Purdue University (GPA 4.0/4.0)</b><br><i>MS, School of Electrical and Computer Engineering</i> | <b>Jan. 2021 – Dec. 2022</b><br><i>West Lafayette, IN</i> |
| <b>IIT Madras (CGPA 8.79/10.0)</b><br><i>B.Tech Electrical Engineering, Minor in Data Science</i>  | <b>Aug. 2015 – May 2019</b><br><i>Chennai, India</i>      |

## PUBLICATIONS

---

- [1] Ganga Meghanath, Bryan Jimenez, and Joseph G Makin. “Inferring population dynamics in macaque cortex”. In: *Journal of Neural Engineering* 20.5 (2023), p. 056041.
- [2] Ran Xu et al. “ApproxNet: Content and contention-aware video object classification system for embedded clients”. In: *ACM Transactions on Sensor Networks (TOSN)* 18.1 (2021), pp. 1–27.

## WORKSHOPS & TALKS

---

- End-to-End Causal Analysis Workshop** — Women in Data Science (WiDS), Puget Sound Chapter, 2024. Delivered a hands-on tutorial on causal discovery and causal inference frameworks using causica, econml and lingam. [GitHub: CausalAnalysisTutorial](#)

## LEADERSHIP & ACTIVITIES

---

- Graduate Teaching Assistant — Purdue ECE (Python labs for 44 students; co-designed assignments and quizzes).
- Microsoft Garage Hackathon — Winner (Sponsored Hack category: “MicroManagers” microaggressions game).
- Elected Class Representative — IIT Madras (represented 130 EE students; initiated curriculum additions).
- Academic Mentor — IIT Madras (guided 4 freshmen; coordinated academic and social support).
- Trained Bharatanatyam dancer; Runner-up — American Express Big Data Challenge (Shaastra); Volunteer — NSS India, Tech-Soc events.