Anjani Ganapathy

tel: +1 678 894 6168 • **email**: anjani.ganapathy@gmail.com • https://aganapath.github.io/

EDUCATION

UNIVERSITY COLLEGE LONDON, Faculty of Brain Sciences — **London, United Kingdom**
Masters in Computational Linguistics — *Sept 2025*

GEORGIA INSTITUTE OF TECHNOLOGY, College of Engineering — **Atlanta, Georgia**
Bachelor of Science in Industrial and Systems Engineering — *May 2019*

EXPERIENCE

Capital One Financial Corporation — **Washington, DC** | *Aug 2019 - Oct 2023*
Data Analysis Manager (Data Product Owner) | Card Data Infrastructure
Snowflake Depersonalized Data Ecosystem (DepEco)

- **Defined product direction with stakeholder input:** Translated business needs to actionable tasks for product and tech teams and regularly communicated product updates to senior leadership (VP+).
  - Refined product end state vision and managed execution for key improvements like real-time updates to DepEco datasets.
  - Directed stakeholder collaboration to identify gaps and redundancies in product requirements, leading to project plan updates such as reducing datasets 20% which accelerated timelines by 1 month while maintaining analysis potential.
- **Coordinated cross-functional delivery:** Facilitated collaboration and aligned priorities across product, tech, and data science teams.
  - Mitigated dependencies and risks in DepEco by escalating and resolving issues like a data transfer gap that resulted in a 2% improvement in DepEco data completeness.
- **Established data validation & customer support:** Led team of 5 to ensure data quality, proactively identify issues, and escalate customer needs.
  - Created and managed DQ checks in AWS Quicksight, enabling escalation and recovery of missing data (5%) in critical datasets.
  - Created executive dashboards in Quicksight to track DepEco adoption metrics and customer satisfaction, informing further improvements
- **Led product adoption to unlock business value**: Enabled adoption of DepEco, which facilitated use of external data in critical valuation and decisioning models that were estimated to generate >$250M in revenue.
  - Conducted trainings to onboard >100 analysts to the product and created official documentation and guidelines for usage.

Principal Data Analyst (Product Focus) | Card Data Infrastructure
Credit Policy Analysis Simulation Tool

- **Led product development & user onboarding**: Wrangled large datasets to create a central tool using Python for more consistent and faster simulation of credit policies on card applicants and experimental A/B testing for credit policy changes; onboarded >75 analysts.
  - Owned entire product lifecycle, including requirement gathering, creation, user onboarding, and maintenance.
- **Designed and executed central reporting suite:** Created a robust automated pipeline to generate reports on the performance of 150+ experimental credit card customer segments, along with alerts for underperforming segments.
  - Monitoring informed larger business decisions such as launching products for customers with no FICO score.

Senior Data Analyst | Bank Marketing Experience
Digital-First Email Marketing Experimentation

- **Optimized campaigns with data-driven insights**: Produced Tableau dashboards to evaluate customer engagement with email marketing campaigns and identify drop-off points.
  - Designed experiments to test email changes that improved consumer adoption of digital-first banking strategies paperless billing increased by ~40%).

PROJECTS

Vector-Quantized VAEs for Unsupervised Part-of-Speech Induction — **UCL** | *Apr 2024 - Aug 2024*
Dissertation Research in Computational Linguistics

- **Developed first neural unsupervised POS tagging baseline for Hindi:** Achieved 47.41% Many-to-One accuracy using custom VQ-VAEs with Gumbel-Softmax discretization
  - Designed novel architecture combining MuRIL BERT embeddings with character-level learned representations to capture morphological features in Hindi's free word order structure
  - Implemented custom diversity loss term, preventing codebook collapse and optimizing latent tag usage

    - Built dual-decoder system with word-level (FFN) and character-level (LSTM) reconstruction, improving convergence by 50% compared to baseline
- **Conducted ablation studies to test efficacy of standard assumptions and model components:** Disproved standard Bi-LSTM assumption for languages with flexible syntax; proved feed-forward decoders more effective for Hindi (25% accuracy improvement).
    - Results showed that Hindi BERT embeddings with character embeddings capture Hindi parts of speech more effectively.
- Technologies: PyTorch, HuggingFace Transformers, Universal Dependencies treebanks, custom VQ-VAE implementation

Neural Machine Translation for Code-Mixed Hinglish **UCL** | *Mar 2024 - May 2024*

NLP Seminar Project

- **Built end-to-end seq2seq NMT system using LSTM architecture**: Achieved low training losses (4.28 to 0.82) through thoughtful architecture design, including proper sequence token handling and dropout regularization
    - Engineered beam search decoding, attention masking for padding tokens, and smoothing functions for BLEU score calculation
    - Processed 190k sentence parallel corpus combining human-annotated and synthetic data, implementing custom tokenization for dual-language input
- **Unique challenges of code-mixed NMT led to unfavorable results**: The combination of dataset size, romanization variations, informal social media text, and lack of standardized orthography resulted in a difficult to train model.
- Technologies: PyTorch, NLTK, custom attention mechanisms, cross-entropy loss optimization

SKILLS

- **Programming Languages + Tools:** Python, Pandas, PyTorch, TensorFlow, NLTK, SQL, R, HuggingFace, Sklearn, matplotlib, seaborn
- **Software & Platforms:** Github, Snowflake, Tableau, AWS Quicksight & S3, Jira
- **Languages:** English – native, Tamil – native, Hindi – native, Spanish – advanced