

KARTIK VASUDEV SHENOY

(323) 986-9136 | kartikshenoy.com | kartik.shenoy@gmail.com | [Linkedin: kartik2112](https://www.linkedin.com/in/kartik2112) | [Github: kartik2112](https://github.com/kartik2112) | San Jose, CA

EDUCATION

University of Southern California

Master of Science in Computer Science with Honors

Courses: Advanced Natural Language Processing, Advanced Computer Vision, Algorithms

Jan 2021 – Dec 2022

GPA: 4 / 4

University of Mumbai, India

Bachelor of Technology in Computer Engineering; **Department Rank: 1**

Aug 2014 – Jun 2018

GPA: 9.56 / 10

EXPERIENCE

Senior Machine Learning Engineer – Bill.com | San Jose, CA

Aug 2024 – Present

- Contributed to the development of an end-to-end pipeline for field extraction for invoices using a rules-based engine to get low-latency and high-quality results. This pipeline achieves a p90 latency of under 350ms and a precision of 95%.
- Created an XGB ranker algorithm using several invoice and customer features that achieves a top-1 accuracy of 89% for entity search on document text-embeddings based retrieval.

Machine Learning Engineer 2 – Bill.com | San Jose, CA

Feb 2023 – Aug 2024

- Designed a proprietary logo extraction model from invoices using YOLOv8 model with a mAP of 0.9 and a similarity model - ConvNext with 87% accuracy to assist in identifying similar invoices with logos.
- Enhanced a document similarity model to improve clustering of merchants in the database using Siamese-BERT model which has increased the transaction payment volume by 10% through enhanced payment rails.
- Leveraged fuzzy string matching based on merchant name and addresses and delivered around 3k leads to the sales team for virtual card payment delivery mechanism.

Machine Learning Engineer Intern – Bill.com | San Jose, CA (Remote)

May 2022 – Aug 2022

- Generated a dataset of 1 million invoices and statements from user data after sanity checks for improving the amount OCR model.
- Improved the amount prediction test accuracy from 75.5% to 77.67% with 50% confidence threshold.

Research Assistant – Information Sciences Institute, USC | Dr. Jon May | Los Angeles, CA

Jan 2022 – May 2022

- Worked on designing bots capable of understanding communication and strategizing for the board game – diplomacy.
- Increased the win rates by an average of 5% for the powers in the game by generating and understanding DAIDE messages and using rules on top of the pre-trained reinforcement learning based Dipnet bot.

Research Assistant – Centre of Knowledge Graphs, USC | Dr. Filip Ilievski | Los Angeles, CA

Feb 2021 – Jan 2022

- Constructed a repository of scripts for identifying low quality statements in Wikidata knowledge graph amongst 1.1 billion statements on the basis of deleted, deprecated statements and constraint violations using kgkt library.
- Enhanced the graph embeddings of nodes using retrofitting based on BERT embeddings and structural, textual properties extracted from Wikidata, Probase and DBPedia datasets increasing Spearman correlation from 0.66 to 0.73 on WordSim353 benchmark.

Software Developer – Barclays Global Service Centre | Pune, India

Jul 2018 – Dec 2020

- Devised a prototype fraud detection system to identify mule accounts by building a pipeline of Kafka queues, Cassandra DB and PySpark servers having an ensemble of Machine Learning models achieving a processing speed of 20ms per transaction.
- Created a classifier application utilizing topics identified with the help of unsupervised algorithm Latent Dirichlet Allocation to extract insights by generating ontologies from iOS and Android application reviews and customer complaints.

PROJECTS

- Alexa Social Bot** - Contributed to the development of Alexa social bot by including conversational elements such as Sentence BERT, offensive speech classification, SNIPS NLU intent classifier, FSMs, DialogPT and other neural text generation modules.
- Generation Augmented Retrieval for Question Answering** achieving 92.6% top-1000 retrieval accuracy and 43.9% EM accuracy on Natural Questions dataset by fine-tuning BART-large model and using OTS Dense Passage Retriever models.
- VeriSign** [Demo Link] [Github Link] - Deployed a web application to verify signatures and to detect forgery which achieved an accuracy of 96% by training a Convolutional Siamese Network using the concept of One-Shot Learning.

PUBLICATIONS

- "MULTIMODAL EMBEDDING SYSTEM FOR BUSINESS RECOMMENDATIONS." Patent Application number 19/043,191, filed on 1/31/2025, with an estimated publication date of 7/31/2026.
- "A study of concept similarity in Wikidata", Semantic Web Journal, October 2023. [Paper Link]
- "Creating and querying personalized versions of wikidata on a laptop, Proceedings of the 2nd Wikidata Workshop (Wikidata 2021). [Paper Link]
- "Does Wikidata Support Analogical Reasoning?" Iberoamerican Conference on Knowledge Graphs and Semantic Web, 2021. [Paper Link]
- "Viola: A Topic Agnostic Generate-and-Rank Dialogue System", 4th Proceedings of Alexa Prize (Alexa Prize 2020). [Paper Link]

- “A Study of the Quality of Wikidata”, Journal of Web Semantics, Elsevier, July 2021. [[Paper Link](#)] [[Github Link](#)]
- “Real-time Indian Sign Language (ISL) Translation”, 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT). Published in IEEE Xplore, October 2018. [[Paper Link](#)] [[Github Link](#)]
- “An Effective Pixel-Wise Approach for Skin Colour Segmentation - Using Pixel Neighbourhood Technique”, International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), 2018. [[paper link](#)]

SKILLS

Programming Languages: Python, C++, Java, JS

Machine Learning: PyTorch, Keras, Pandas, TensorFlow, Scikit-learn, NLTK, Spacy, Knowledge Graphs

Frameworks: OpenCV, Numpy, PySpark, Kafka

Web: LAMP stack, Angular JS, NodeJS, Flask

Database: MySQL, MongoDB, CassandraDB

Tools: AWS (EC2, RDS, S3, EKS, Comprehend, Athena, SageMaker), Github, Firebase, Matlab, Google Colab, Docker

ACHIEVEMENTS & INVOLVEMENTS

- Technical Volunteer for the **LLM Workshop** at Bill assisting in teaching essential LLM concepts and API usage to engineers outside the AI team to make them better equipped.
- Spearheaded the charter of and was the ‘**President**’ of Bill Dot Toast Toastmasters Club (Apr 2023 – Dec 2023).
- Won the **best project** award at **Bill AI Hackathon** for BILL Pulse.
- **Semifinalist** team of **Alexa Socialbot Grand Challenge 4** (2020 – 2021) as part of Team Viola.
- **Guided 900+ students** as a **Teaching Assistant** for the courses Artificial Intelligence and Algorithms (Jan 2022 – Present) at USC.
- Volunteered at **Graffiti Removal** and **Urban Farm** events through USC Volunteer Center.
- Bagged the **Barclays Award of Stewardship** for automating the generation of real-time delivery metrics of more than 30 teams from Agile Central and Jira data sources. These dashboards have been saving around 150 man-hours annually.
- **Sir Dorabji Tata Trust scholarship** for securing 1st Rank in a class of 120 students during undergraduate studies.
- ‘**Vice-President - Membership**’ of Barclays Pune 2 Toastmasters Club (January 2019 – June 2019).
- Active member of Junior Chambers International Club – contributing to **social service** activities such as creating seed balls for afforestation in areas devastated by floods and blood donation.
- **Taught basic computer skills** of Microsoft Office to underprivileged students over a span of 10 days.
- **Placement Coordinator** during under-graduation, managing logistics for around 40 companies and 800 students.
- Indian Classical Music Level 3 certified singer and Harmonium player and a beginner guitar player.