

Kartikay Bagla

✉ hello@karb.cc | 🌐 karb.cc | 📞 9560309524 | 📄 kartikay-bagla | 🌐 LinkedIn

EXPERIENCE

Senior Software Engineer – Flexera **Dec 2025 – Present**
Scaling AI agents for data cloud cost optimization (Databricks, Snowflake) from 1x prototype to 10x production. Serving Fortune 100 enterprises with auditable decision traces.

Machine Learning Engineer II – Chaos Genius **Jun 2024 – Dec 2025**
Building AI agents for automatic cost optimization across Databricks & Snowflake. Developing MCP servers using SOTA models (open-weight and closed source) to analyze cluster/warehouse history for intelligent optimization.

Machine Learning Engineer – Chaos Genius **Jun 2022 – May 2024**
Built LLM-powered Spark log parser for inefficiency detection. Created Databricks Cost Observability pipeline with ARIMA, Prophet, Neural Prophet forecasting. Developed Snowflake Observability backend (FastAPI + AWS Cognito SSO).

Machine Learning Intern – Chaos Genius **May 2021 – Jun 2022**
Built open source platform for automated Root Cause Analysis, Anomaly Detection, and Data Quality at enterprise scale.

NLP Intern – RuleZero **Jul 2020 – May 2021**
Developed Chatbot and Q&A system for startup funding queries (shares, options, securities) collaborating with lawyers and VCs.

Co-Founder – IndiaAgainstCovid **May 2020 – Dec 2020**
Founded NGO relief coordination platform during COVID-19. Built web platform to optimize supply distribution. Led mental health, job placement, and fundraising initiatives.

NLP Intern – Practical NLP Book (O'Reilly) **Dec 2019 – Jun 2020**
Contributed to O'Reilly book by Anuj Gupta & Harshit Surana. Edited drafts and QA'd code. Endorsed by Zachary Lipton, Sebastian Ruder & Marc Najork.

ML Intern – DeepFlux **Jun 2019 – Jan 2020**
Built marketing impact prediction models. Developed custom GAMs (Generalized Additive Models) for interpretability.

PROJECTS

QuickClone Extension **Aug 2021**
Browser extensions (Chrome, Firefox) and VSCode extension to clone and open GitHub projects directly in VSCode.

Real-time Background Replacement **Jul – Dec 2020**
Real-time background removal for video streaming using custom Mask R-CNN/tinyYOLO architecture. Achieved 30fps on CPU.

Bump Plot Library for Matplotlib/Seaborn **Aug 2019**
Python library adding Bump Plot support for visualizing rank changes over time.

Face Follower **Jul 2019**
Face-tracking camera using Arduino servos + Raspberry Pi 2B+ with OpenCV and ROS. Achieved 5-10 FPS.

Q-Table Learning Visualization **Jun 2019**
Interactive visualization of Q-Learning algorithm for pathfinding with real-time Q-Table updates.

PUBLICATIONS

Noisy Text Data: Foible of Popular Transformer Based NLP Models **2024**
AIMLSystems '23: Third International Conference on AI-ML Systems
Shows that popular transformer-based NLP models perform poorly on noisy text data across tasks like classification, NER, QA, and summarization, with performance degrading as noise increases.

DarthYOLO: Using YOLO for Real-Time Image Segmentation **2022**
ICAPIE 2022, Advances in Transdisciplinary Engineering
Novel architecture for near-real-time image segmentation using YOLOv4 for lightweight object detection, enabling deployment on consumer-grade hardware.

EDUCATION

B.Tech. Mathematics & Computing **2018 – 2022**

Delhi Technological University

AISSCE (Class 12) **2018**

Amity International School, Noida

AISSE (Class 10) **2016**

Amity International School, Noida