PRONOB KUMAR BARMAN

Halethorpe, MD | test861999@gmail.com | +1 301-615-2587

SUMMARY

AI/ML Engineer and Ph.D. candidate with 7+ years of experience designing and deploying intelligent systems across healthcare, cybersecurity, and financial domains. Expert in LLMs, GenAI, deep learning, MLOps, and recommendation systems with a proven ability to transition models from research to production. Adept in deploying transformer models, integrating AI fairness, and leading scalable AI architecture using cloud-native and secure pipelines. Published researcher and mentor dedicated to building transparent and impactful AI solutions.

CORE COMPETENCIES

- **Programming & Framework:** Python, R, SQL, Java, PyTorch, TensorFlow, scikit-learn, Keras, NumPy, Pandas, FastAPI, Flask
- **LLM/GenAI:** Prompt Engineering, GPT-3/4, LangChain, RAG Pipelines, PEFT, LoRA, Hugging Face
- **Machine Learning:** Deep Learning, NLP, XGBoost, BERT, Recommendation Engines, Reinforcement Learning
- **MLOps & Infrastructure:** AWS SageMaker, GCP, MLflow, Kubernetes, Docker, Vertex AI, CI/CD Pipelines
- **AI Fairness & Safety:** SHAP, LIME, Fairlearn, Opacus, Differential Privacy, Explainability Toolkits
- **Security AI:** Threat Detection, Anomaly Scoring, SIEM Analysis, Secure API Deployment
- **Data & Analytics:** Pinecone, PostgreSQL, ChromaDB, Power BI, Tableau, Kafka, FastAPI
- **Collaboration & Tools:** Trello, Slack, Jira, Confluence, Notion, MS Teams, Google Workspace, VS Code, Jupyter Notebook, Postman, Swagger, Zoom, Figma

PROFESSIONAL EXPERIENCE

Machine Learning Engineer | Technuf LLC, Rockville, MD | May 2025 – Present

Leading design and deployment of HIPAA-compliant AI systems for behavioral health platforms using cloud-native MLOps infrastructure.

- Built and maintained a **custom MLOps platform** on **AWS SageMaker Studio** integrating **TensorBoard, MLflow, and Kubernetes**, enabling lifecycle tracking, auto-scaling, and drift monitoring for transformer-based **DistilBERT and RoBERTa** models.
- Designed and deployed a **Reinforcement Learning for Recommendation (RL4Rec)** module using **Proximal Policy Optimization (PPO)** to deliver daily peer-matching recommendations within a digital therapy platform, increasing daily session retention by 15%.
- Developed an **AI Fairness Evaluation module** with **SHAP, LIME, and Fairlearn**, integrating demographic parity checks for psychographic patient clusters, improving fairness metrics (DPG, EOD) by over 25%.
- Engineered an **Inference-as-a-Service API Gateway** using **FastAPI**, deployed on **EKS (Elastic Kubernetes Service)**, and backed by **Pinecone vector DB** to return real-time personalized treatment recommendations; average response time dropped to under 350ms.
- Created a **secure ingestion pipeline** using **Kafka, Apache Airflow**, and **BERT Tokenizer** to sanitize, classify, and vectorize over 1M behavioral health notes per day, integrating with an EMR via **FHIR-compliant endpoints**.

Research Assistant – LLM Recommendation Systems | UMBC, Baltimore, MD | Aug 2023 – Present

Researching scalable hybrid recommendation systems and bias-aware LLM architectures for healthcare personalization.

- Developed a **Neural Collaborative Filtering (NCF)** recommender system with **DistilBERT sentence encoders** and **user survey embeddings** to predict top-K support group matches; improved **NDCG@10** by 22% vs matrix factorization baselines.
- Built a **cold-start resolution module** combining **contrastive learning (SimCLR)** and **multi-task loss** using content metadata and sparse survey responses, improving **HR@10** by 19%.
- Released gSTM (R package for structural topic modeling) and gDMR (Python package for dynamic matrix reduction), used in reproducible evaluation pipelines across four internal research projects.
- Designed a **dual-tower transformer encoder** combining user psychographics (MBTI, stress level) and group content (topic labels, sentiment), embedded into a **shared latent space** via cosine similarity for fine-grained usergroup matching.
- Led fairness and explainability evaluation for 50+ ablation configurations, visualized via **UMAP/T-SNE** and published at **MedInfo 2025** and **SIAM SDM 2025**.

Machine Learning Intern – Prompt Engineering & GenAI | Technuf LLC, Rockville, MD | Jun 2024 – Aug 2024
Focused on designing secure, explainable GenAI pipelines for security monitoring.

- Developed a **LangChain RAG system** leveraging **ChromaDB and FAISS retrievers** for summarizing and contextualizing SIEM logs from **Splunk** and **Wazuh**, processed over 50K log entries/day.
- Implemented a **hybrid threat scoring engine** combining **LightGBM** on structured log features and **LSTM** on tokenized alert streams, reducing false alert volume by 20%.
- Designed **templated GPT-3.5 prompts** with memory context windows and injected system roles for classifying alerts by type (DDoS, privilege escalation); integrated with **Power BI dashboards** for SOC consumption.
- Built and deployed **spaCy Named Entity Recognition (NER)** pipelines fine-tuned on cybersecurity corpora (MITRE ATT&CK) to extract IOCs (IP, MAC, CVE ID) in real time.
- Created a hardened FastAPI layer using **OWASP ASVS** Level 2 guidelines for delivering threat briefings via secure endpoints with **JWT-based authentication**.

ML Engineer – Conversational AI | Technuf LLC, Baltimore, MD | Aug 2022 – Dec 2022 Worked on
chatbot architecture and PHI protection modules for healthcare clients.

- Architected a **multi-turn dialog manager** using **BERT embeddings + GRU decoder** with slot-filling for appointment, prescription, and insurance inquiries; achieved 87% intent classification accuracy.
- Built a **PHI redaction pipeline** using **Regex + BERT NER** ensemble with FastText embeddings to ensure compliance with HIPAA storage rules across incoming chats.
- Optimized **Redshift-SQL pipelines** handling 200K+ daily user conversations, enabling Power BI dashboards to load in under 5 seconds for real-time sentiment and CSAT visualization.
- Developed an **escalation risk classifier** using XGBoost trained on user tone and token sequences, increasing autoescalation accuracy by 18%.
- Created an internal **Engagement KPI Dashboard** using **Power BI** linked via **REST APIs** to the dialog platform's backend metrics (avg. session time, first-response resolution, negative sentiment triggers).

Deputy Director – AI Systems & Policy Modeling | Bangladesh Bank, Dhaka, Bangladesh | Apr 2018 – Dec 2021 Led
AI policy analytics and digital transformation across financial systems.

- Rolled out a **Bengali NLP chatbot** using **FastText + LSTM encoder-decoder** for real-time user query handling, reducing call center load by 60%.
- Built an **anomaly detection pipeline** using an **Autoencoder + Isolation Forest ensemble** to monitor transactional data for AML (Anti-Money Laundering) compliance.
- Created a **Tableau-powered Policy Analytics Dashboard** combining macroeconomic indicators and chatbot usage stats for weekly review by executives.
- Migrated legacy banking datasets (COBOL + Excel) to a **PostgreSQL lakehouse** using **Python's Pandas and pyodbc**, reducing data refresh lag by 80%.
- Developed a **policy forecasting simulation** using time-series modeling in **R (ARIMA, Prophet)** integrated into Tableau for scenario planning.

EDUCATION University of Maryland, Baltimore County (UMBC)
Ph.D. in Information Systems | GPA: 3.85 | *Expected Dec 2025* Focus:
Human-Centered GenAI, Recommender Systems

University of Kentucky
M.S. in Statistics | GPA: 3.85 | May 2021

University of Dhaka
M.S. & B.S. in Statistics | GPA: 3.46 | Jan 2017

SELECTED PUBLICATIONS

- Barman, P. K. et al. "Facilitating Online Healthcare Support Group Formation using Topic Modeling," *MedInfo 2025*.
- Barman, P. K. et al. "Enhancing Online Support Group Formation with Topic Modeling," *SIAM SDM 2025*.
- Barman, P. K. et al. "User-generated Data in Mobile-Personal Health Records," *AMIA Symposium 2024*.
- Barman, P. K. "Detecting Abnormal Health Conditions in Smart Home Using a Drone," *arXiv:2310.05012*.

AWARDS & GRANTS

- UMBC GSA Research Grant ($1,000), 2024 – Healthcare recommender systems
- NSF Travel Grant ($750), 2025 – SIAM SDM Doctoral Forum presenter