SURBHI AGRAWAL

646-309-7816 sa4480@columbia.edu Surbhi Agrawal

EDUCATION

Columbia University, New York **Aug 2024 – Dec 2025**
*M.Sc in Mechanical Engineering (**Specialization - Robotics and Controls**)* *New York, USA*
Key Courses: Reinforcement Learning, Probabilistic Robotics, Robot Learning, Speech Processing, Adv. Probability Theory
Teaching Assistant: Nonlinear and Adaptive Control (Fall'25), Advanced Machine Dynamics (Spring'25)

Indian Institute of Technology (IIT), Delhi **June 2017 – June 2021**
Bachelor of Technology in Textile Engineering *Delhi, India*

WORK EXPERIENCE

Applied NLP Engineer – Master's Thesis Research **Jan 2025 – Present**
Multi-modal Semantic Search & Retrieval System *New York, NY*

- Architected multi-modal semantic search fine-tuning LLM and **CLIP** models with distributed training across 8 GPUs, improving cross-modal retrieval by 18%
- Engineered production **RAG** pipeline with knowledge graph integration and vector database indexing, reducing LLM hallucination by 31% in production scenarios
- Built **RLHF** pipeline using PPO with distributed reward modeling on human preference datasets, improving by 23%
- Deployed realtime classification with XGBoost on multi-modal streams, achieving 92% F1-score at <100ms latency

ML/AI Research Engineer – Production Systems | Honda R&D, Tokyo **Oct 2021 – Aug 2024**
Real-Time Mapping and Object Tracking in Birds Eye View Representation *Tokyo, Japan*

- Designed scalable ML workflow using **CI/CD pipelines, Docker, Kubernetes**, and GitHub Actions for automated training/deployment
- Implemented distributed model optimization through knowledge distillation and quantization, achieving 2× inference speedup with 1% accuracy improvement
- Deployed production models via FastAPI with TensorRT backend on Kubernetes cluster, serving 30 FPS inference with less than 50ms P99 latency
- Built distributed data pipeline processing 1TB+ of sensor data using **PySpark** and cloud storage

KEY PROJECTS

RL Based Recommendation System with Dynamic CTR Optimization **Jan 2025 – May 2025**
Prof. Javad Ghaderi *Columbia University, New York*

∗ Built search pipeline using TF-IDF (scikit-learn), **BERT embeddings** (Hugging Face), **FAISS indexing**, and **PyTorch** DQN agent with experience replay to learn optimal ranking policies maximizing CTR and engagement
∗ Implemented contextual multi-armed bandit with Thompson Sampling, integrated **Apache Kafka** for real-time streaming and **Redis caching**, achieving 32% CTR improvement over LambdaMART baselines
∗ Designed multi-objective reward function in OpenAI Gym, trained with Ray RLlib distributed computing, tracked via MLflow/WB, improving satisfaction by 18% validated through **A/B testing**

Prediction of Consumer Satisfaction of Medium Articles **April 2020 – May 2021**
Prof. Sumitava Mukherjee *IIT Delhi, Delhi*

∗ Built scalable ML pipeline processing 50K+ articles: designed **SQL**-based ETL for data ingestion, implemented **feature engineering** (readability scores, TF-IDF, GloVe embeddings) and trained ensemble models for multi-task prediction
∗ Developed NLP preprocessing pipeline using NLTK and spaCy for tokenization, POS tagging, and named entity recognition to extract linguistic features for sentiment analysis
∗ Trained gradient boosting classifiers (XGBoost, LightGBM) with hyperparameter tuning via Optuna, achieving 88% precision in predicting article engagement and reader satisfaction scores

PUBLICATIONS

C1. A. More, **S. Agrawal**, M. Soni and S. Divakr Bhat. *Prior2Posterior: Model Prior Correction for Long-Tailed Learning.* IEEE/CVF Winter Conference on Applications of Computer Vision **(WACV), 2025** [Paper]

J1. D. Baster, P. Mettraux, **S. Agrawal** and H. H. Girault , Chem. Commun., **2019**, 55 , 14633 —14636 [Paper]

SKILLS

Languages & Scripting: Python, C/C++, MATLAB, Bash, TypeScript
NLP & ML: Scikit-learn, spaCy, NLTK, BERT, CLIP, RAG pipelines, LangChain, OpenAI API
Data & Analysis: Pandas, NumPy, OpenCV, Open3D
GPU & Deployment: CUDA, cuDNN, TensorRT, NVIDIA Docker, Cloud GPU Platforms, AWS