JIBEK GUPTA

Washington, DC (202) 961-3129 jibek.gupta@bison.howard.edu linkedin.com/in/jibekgupta

PROFESSIONAL SUMMARY

Undergraduate researcher in Computer Science with hands-on experience in retrieval-augmented generation and natural language processing. Interested in developing reliable RAG systems, robust evaluation frameworks for large language models, and interpretable AI tools that serve underrepresented communities. Proficient in Python, machine learning workflows, and end-to-end experimental design and evaluation. Seeking PhD opportunities to advance research in scalable and trustworthy NLP systems that democratize access to information.

EDUCATION

HOWARD UNIVERSITY Washington, DC
Bachelor of Science in Computer Science Expected May 2026
CGPA: 3.72

RESEARCH EXPERIENCE

University of Chicago, Data Science Institute Chicago, IL
Data Science Research Assistant Jun 2025 - Aug 2025

- Developed AI-powered Q&A chatbot using open-source large language models (Phi-4, Llama-3, Gemma-2B) to provide farmers streamlined access to agricultural seed laws across 78 countries
- Processed and analyzed 183 legal PDF documents across 9 languages and covering legislation from 1981-2023
- Built end-to-end Retrieval-Augmented Generation (RAG) pipeline with ChromaDB vector database, implementing semantic search and document chunking strategies for legal documents ranging from 1-150 pages
- Collaborated with A Growing Culture nonprofit organization to advance global food sovereignty through multilingual document processing system development
- Optimized chatbot response time to 15-20 seconds and evaluated model performance using ROUGE metrics across multiple LLM architectures
- Identified critical gaps in multilingual NLP preprocessing pipelines and developed custom text normalization strategies for non-English legal terminology across 9 languages

Howard University, Quantitative Histories Workshop (REU Program) Washington, DC
Data Analysis Research Assistant Aug 2024 - Present

- Developed comprehensive Census data dashboard to replace complex Census Bureau website navigation, improving researcher access to demographic insights
- Automated processing and analysis of over 100,000 U.S. Census records spanning 2009-2023 with real-time filtering and visualization capabilities
- Created full-stack interactive dashboard using Python, Shiny, and Pandas, resulting in 40% increase in stakeholder engagement and 35% improvement in user experience metrics
- Presented research findings and demographic insights to audiences of 200+ attendees at research symposiums
- Translated complex census data into accessible findings for community organizations, policymakers, and academic researchers

PAPERS IN PREPARATION

- Gray, A., Gupta, J., Jones, A., Rojas, G., Ronemous, P., Bhowmick, S. S., & Sehgal, K. (in preparation). SeedBot: A RAG-Based Conversational AI for Legal Domain Queries.
- Onabajo, B., Alexander, N. N., Gupta, J., Emudianughe, P., & Okonkvo, C. (in preparation). The Impact of Structural Socioeconomic Factors on Hypertension Prevalence in the U.S.: A Case Study of Washington, D.C. Using ARDL and Spatial Metrics.

ABSTRACTS & PRESENTATIONS

- Gray, A., Gupta, J., Jones, A., Rojas, G., Ronemous, P., Bhowmick, S. S., & Sehgal, K. (November 2025). SeedBot: A Q&A Tool for Farmers. Poster presentation at the *Annual Biomedical Research Conference for Minoritized Scientists (ABRCMS)*, San Antonio, TX.
- Gray, A., Gupta, J., Jones, A., Rojas, G., Ronemous, P., Bhowmick, S. S., & Sehgal, K. (August 2025). SeedBot: A Q&A Tool for Farmers. Presentation at the *University of Chicago Data Science Institute*, Chicago, IL.

- Gupta, J. & Alexander, N. N. (April 2025). Parts of a Patchwork Nation: Designing an Open Source Tool to Investigate Demographic Shifts in Urban and Rural Communities. Poster presentation at the *2025 Howard University Research Month Symposium*, Washington, DC.
- Alexander, N. N., Onabajo, B., Gupta, J., Nichols-Smith, K., & Jang, H. (March 2025). Mapping Multidimensionality: Using Census Data to Understand Neighborhood Communities. Presentation at the *2025 DC Open Data Showcase*, Washington, DC.
- Alexander, N. N. & Gupta, J. (December 2024). Ancestral Mathematics. Workshop presented at the *2024 HBCU Alumni Alliance and 100 Black Men STEM Day, University of the District of Columbia*, Washington, DC.

TECHNICAL PROJECTS

MNIST Classification with KNN | Python, Scikit-Learn, Pandas
- Implemented K-Nearest Neighbors classifier achieving 97% accuracy on MNIST digit recognition dataset using systematic grid search hyperparameter optimization
- Applied cross-validation and comprehensive model evaluation techniques including confusion matrices, per-class performance analysis, and error analysis
- Developed complete machine learning workflow with data preprocessing, stratified sampling, and performance visualization using Python, scikit-learn, and statistical libraries

Hybrid Movie Recommender System | Python, Scikit-Learn, Tkinter
- Designed and implemented hybrid recommender system combining TF-IDF content-based filtering with item-item collaborative filtering, achieving 80% accuracy on benchmark datasets
- Implemented TF-IDF feature extraction with cosine similarity and k-nearest neighbors algorithms over sparse Compressed Sparse Row (CSR) user-item matrices for efficient Top-N recommendation retrieval
- Incorporated alpha weighting parameter for runtime optimization between content-based and collaborative filtering approaches, enabling personalized recommendation tuning

Automated Web Scraping Application | Python, BeautifulSoup, Selenium
- Developed comprehensive Python web scraping solution using Selenium WebDriver and BeautifulSoup for automated data extraction from educational technology platforms
- Implemented robust dynamic pagination handling and exception management systems to reliably navigate and extract data from 100+ pages of product listings
- Designed data pipeline to export scraped information to structured CSV format, creating comprehensive dataset of educational technology products for subsequent analysis

TECHNICAL SKILLS

Programming Languages: Python, SQL, HTML, CSS
Machine Learning & AI: Scikit-Learn, TensorFlow, PyTorch, Natural Language Processing (NLP), Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Feature Engineering, Predictive Modeling
Data Science & Analytics: Pandas, NumPy, Matplotlib, Seaborn, ETL Pipeline Development, Statistical Analysis, Data Visualization
MLOps & Cloud: Hugging Face, ChromaDB, Amazon Web Services (AWS), Model Deployment
Tools: Git, GitHub, Jupyter Notebook, Visual Studio Code, Selenium, BeautifulSoup, R Shiny

RELEVANT COURSES

Data Structures & Algorithms | Machine Learning |
Applied Data Science | Database System |
Probability & Statistics | Cloud Computing |
Operating System | Computer Networks |
Software Engineering

AWARDS & SCHOLARSHIPS

HU Leadership Scholarship (HULEA) - Leadership and academic excellence
HU Need (HUNEE) - Merit-based financial support
Mordecai W. Johnson Scholarship - Donor scholarship