JHON G. BOTELLO

+1 (757) 235-7651 | jbote001@odu.edu | Norfolk, VA, USA | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Detail-oriented Data Scientist and Ph.D. candidate with 4 years of experience specializing in Generative AI, Data Analytics, Machine Learning, and LLM training and deployment. Proven ability to deliver actionable insights and improve processes through data-driven strategies and storytelling. Experience collaborating on international research projects with interdisciplinary teams across Colombia, the United States, Greece, Turkey, and Norway.

EDUCATION

Old Dominion University **August 2023 – May 2028 (Expected)**
Ph.D. Computer Science *Norfolk, VA*
Coursework: Machine Learning, Information Retrieval, Data Structure and Algorithms, Quantum Computing, and Data Visualization.

Universidad del Norte **January 2022 - August 2023**
Master's, Industrial Engineering *Barranquilla, Colombia*
Coursework: Machine Learning, Artificial Intelligence, Data Mining, Data Analysis, Operations Management, Experiment Design and Analysis.

Universidad del Norte **July 2017 – September 2022**
Bachelor's, Industrial Engineering *Barranquilla, Colombia*
Coursework: Data Mining, Data Analysis, Simulation, Operations Research, Systems Analysis, Statistical Concepts in Engineering Management.

PROFESSIONAL EXPERIENCE

Virginia Modeling, Analysis, and Simulation Center (VMASC) **Suffolk, VA, USA**
Graduate Research Assistant *August 2023 - Present*

- Designed and implemented a reproducible generative AI workflow to extract, standardize, geolocate, and validate historical mobility records from semi-structured datasets in collaboration with historians at the University of California, Berkeley.
- Benchmarked GPT-4o against API-based tools for geolocating historical place names to advance the state of the art in understanding what LLMs know about the past.
- Fine-tuned open-source LLMs on domain-specific NetLogo code instructions to advance automated and rapid simulation model development.
- Improved GPT-3.5's ability to generate NetLogo procedural code by 50% by implementing retrieval-augmented generation (RAG) techniques to support agent-based modeling.
- Designed and implemented multimodal RAG prototypes to support domain-specific information retrieval and contextual response generation.
- Developed a data-driven tool to automate survey research techniques, including confirmatory factor analysis and structural equation modeling, to accelerate survey validation and insight generation.
- Adapted and validated a survey to contribute to the development of a theory of host community support for migration in Colombia, Greece, and South Africa.
- Applied supervised and unsupervised machine learning techniques to social science datasets to identify and model patterns of public support for migration in Colombia.
- Published and presented research papers at conferences across the U.S., France, and Colombia.

Old Dominion University **Norfolk, VA, USA**
Ph.D Candidacy Independent Research *August 2023 – Present*

- Developed a generative AI–driven tool for analyzing content changes in web archives, enabling automated detection and interpretation of content evolution over time.
- Investigated how web archives can inform modeling and simulation development by extracting information using AI techniques.

Universidad del Norte **Barranquilla, Colombia**
Graduate Research Assistant *November 2021 - August 2023*

- Conducted fieldwork for migration-related data collection across multiple places in Colombia.
- Performed data analysis and survey modeling on migration-related datasets to identify trends and determinants of host communities' support.
- Collaborated with migration experts, government, and stakeholders to contextualize a novel survey instrument measuring host community support of refugees and migrants in Colombia, Greece, and South Africa.

SKILLS

Coding Languages: Python, R.
Simulation Software: Vensim, Netlogo, Arena, Bizagi
Generative AI: Fine-tuning, RAG, n-shot learning, LangChain, Llama Index, Hugging Face, OpenAI, Ollama, VLLM, Unsloth AI.
Machine Learning and Data Analysis: Pandas, NumPy, Matplotlib, Artificial Neural Networks (ANN), Principal Component Analysis (PCA), scikit-learn, regression, classification, clustering.
Database/Management Systems: SQL, Git, GitHub, Docker.
Visualization & Business Intelligence: Tableau, Power BI.
No Coding Software: Orange, SPSS, Jamovi
Languages: English (fluent), Spanish (native).

PUBLICATIONS

Published

Botello, J. G., Firuzan, M., Padilla, J. J., Philliou, C., Frydenlund, E., Anderson, A. Reconstructing Migratory Movements from Ottoman Census Registers: An AI-Driven Approach. *Journal of Digital History*, (In Press).

Botello, J. G., Llinas, B., Padilla,J.J., and Frydenlund, E. (2025). Toward Automating System Dynamic Modeling: Evaluating LLMs in the Transition From Narratives to Formal Structures. *In Proceedings of the 2025 Winter Simulation Conference*. https://doi.org/10.1109/WSC68292.2025.11338856

Botello, J. G., Frew, L., Padilla, J. J., and Weigle, M. C. (2024). Exploring Large Language Models for Analyzing Changes in Web Archive Content: A Retrieval-Augmented Generation Approach. *In Proceedings of the 2024 IEEE Big Data Conference*. https://doi.org/10.1109/BigData62323.2024.10826069.

Martinez, J., Llinas, B., **Botello, J. G.**, Padilla, & Frydenlund, E. (2024). Enhancing GPT-3.5's Proficiency in NetLogo Through Few-Shot Prompting and Retrieval-Augmented Generation. *In Proceedings of the 2024 Winter Simulation Conference*. https://ieeexplore.ieee.org/document/10838967.

Botello, J. G., Palacio, K., Frydenlund, E., Llinás, H., and Padilla, J. J. (2024). Adapting and Validating a Survey to Assess Host Communities' Support for Migration. *Social Indicators Research*, 174(2), 697-720. https://doi.org/10.1007/s11205-024-03397-6.

Pachón-Carrascal, K., **Botello, J. G.**, García-Llinás, G., Viloria-Núñez, C., Romero-Rodríguez, D., Maturana, A. Y., and Salgar, K. S. P. (2023, November). BPM for mutual aid management in industrial emergency response. In *2023 IEEE International Humanitarian Technology Conference (IHTC)* (pp. 1-7). IEEE. https://doi.org/10.1109/IHTC58960.2023.10508837.

On Track

Botello, J. G., Padilla,J.J., and Frydenlund, E. (2025). Fine-tuning a Model for NetLogo's Rapid Simulation Development (Work in Progress).

Botello, J. G., Padilla,J.J., and Frydenlund, E. (2025). ArchiveDiff-LLM: Detecting Important Changes in Archived Web Content (Ready for Submission).

Botello, J. G., Frydenlund, E., Palacio, K., Llinas, H., Padilla, J. J. Assessing Host Community Support for Migration: A Case Study of Barranquilla and Cúcuta, Colombia. (Ready for Submission).

Llinas, B., **Botello, J. G.**, Padilla,J.J., Frydenlund, E., and Nelson, RAG-TEC: Extracting and Classifying Topics in Digital News Collections Using LLMs. (Ready for submission)

OUTREACH

Invited Workshop Speaker – Thinking with evidence, transforming information into knowledge. Workshop on Data Sciences, Online, Bogota, Colombia, February 2026.

Invited Conference Speaker – Panel on the Analysis of Public Opinion and Community Support for Migration and Displacement. International Studies Association Annual Convention 2024, San Francisco, CA, April 2024.

Invited Workshop Speaker – Digital Methods for the Study of Mobilities Workshop, INALCO, Paris, France, June 2024.

Invited Workshop Speaker – Digital Humanities and the Study of Mediterranean Mobilities workshop, University of California, Berkeley, CA, January 2024.

AWARDS AND DISTINCTIONS

Graduate Research Assistantship, Virginia Modeling, Analysis, and Simulation Center – August 2023.
Graduate Distinguished Student, Universidad del Norte – September 2023.
Undergraduate Distinguished Student, Universidad del Norte – September 2022.
Awarded scholarship, Colombian Ministry of Education – November 2016.