Gelson Pagan-Diaz, PhD

Austin, TX ▪ 737-742-8171 ▪ gelsonj.pagan@gmail.com
linkedin.com/in/gjpagandiaz | github.com/gjosuepagan

OBJECTIVE

A seasoned Data Scientist with over 8 years of experience in multidisciplinary research, I am seeking a challenging position where I can leverage my expertise in math, science research, creative and critical thinking, and collaborative teamwork to drive innovation. Transitioning from academia in bioengineering and computational neuroscience, I am eager to contribute my skills to a dynamic industry environment. I am particularly interested in a role at a forward-thinking company that values a growth mindset, where I can substantially impact utilizing my proficiency in statistics, large-scale data handling, machine learning models, and biomimetic reinforcement learning. I am eligible for Interim Secret access to information based on the Advanced NAC investigation completed on 2023/08/15. My tier 3 is being performed by OPM.

PROFICIENCIES

- **Programming Languages:** Python, C++, Bash Scripting, MATLAB, CUDA
- **Data Science Techniques/Methods:** Experimental Design, Regression, Hypothesis Testing, Confidence Intervals, Clustering, Classification, Time Series Analysis, Machine Learning, Deep Learning, Reinforcement Learning
- **Python Packages:** NumPy, Pandas, Scikit-learn, SciPy, matplotlib, Seaborn
- **Deep Learning Frameworks:** PyTorch, TensorFlow
- **Data Visualization:** Matplotlib, PyPlot, Seaborn
- **Tooling:** Git/GitHub, OriginLab, Microsoft Office Suite, AutoCAD, Adobe Illustrator, Microsoft PowerBI
- **Operating Systems:** Windows, Linux

PROFESSIONAL EXPERIENCE

Department of Agronomy and Agriculture, University of Nebraska-Lincoln (remote) **3/2024 – Present**
Machine Learning Consultant

- Lead the exploration and design of Large Language Models (LLMs) to integrate molecular genetic sequences for agricultural genetics consulting services.
- Develop informed hypotheses models customized for test breeding design programs in Africa, prioritizing the prediction of evolutionary fitness.
- Utilize advanced machine learning techniques to drive the enhancement of agricultural practices, fostering sustainable growth and innovation within the field.

Lone Star Analysis, Dallas, TX **5/2023 – 3/2024**
DevOps / Analyst

- Developed ETL processes and managed repository architecture for client data to provide actionable insights, reducing solution preparation time by 50%.
- Streamlined project-wide architecture to automate project updating and enhance user-friendliness by eliminating ~70% of manual input sites across the process pipeline.
- Performed thorough data exploratory analysis and processing of classified data and successfully presented focused conclusions to non-expert clients in preparation for modeling and simulations of their solutions.
- Collaborated with international colleagues on a classified research project to implement our flagship product for Monte Carlo simulations to facilitate medical decision-making and risk assessment.
- Provided lectures on QA testing and test-driven development to our Operations Optimization Solutions department.

Department of Neuroscience, University of Texas, Austin, TX **9/2021 – 4/2023**
Research Engineer

- Collaborated on a CUDA-based simulation project focused on biomimetic reinforcement learning. Implemented advanced neural network modeling of the cerebellum and achieved a 50% reduction in epoch runtime through optimized GPU implementation
- Designed and implemented analytical scripts in C++ to extract ASCII data of neuron-specific firing patterns, and transform them into large insightful datasets.
- Conducted in-depth time series analysis using Jupyter Notebooks, providing valuable insights into the correlation between neural activities and behavior learning

- Championed the migration of analysis and data processing code to Python, enhancing efficiency and collaboration by implementing version control with lab-wide GitHub repositories dedicated to cutting-edge research and simulation development and parallel processing.
- Led the strategic growth of the team through recruitment, training, and mentorship, expanding our expertise in biomimetic simulations and reinforcement learning.
- Explored key neurological behaviors such as forgetting and extinction, aligning with the latest advancements in reinforcement learning and biomimetic neural networks.

Department of Bioengineering, University of Illinois, Urbana-Champaign, IL **1/2020 – 9/2021**
Research Engineer

- Spearheaded the establishment of a iPSC-derived organoid/neuroengineering subgroup, effectively managing personnel and acquiring essential equipment within a starting budget of ~$5,000, contributing to the development of three theses, two publications, and the successful acquisition of two +$1M grants.
- Led the acquisition and processing of biological iPSC-derived neural data from local field potential recordings across hundreds of parallel electrodes, usually involving the processing of ~ terabyte-sized data files.
- Transformed raw signals into structured time series data, employing data science methodologies such as clustering, cross-correlation, and spectral analysis, unveiling insights into behavior modulation in response to controlled stimuli.
- Designed and standardized custom MATLAB scripts for comprehensive data analysis and visualization to be used lab-wide

VOLUNTEER AND PRO BONO WORK

Institute of Active Inference (remote) **1/2024 – Present**
In collaboration: Dr. Sanjeev Namjoshi

- Serving as a volunteer reviewer for a textbook writing project on the fundamentals of active inference and its practical implementation on biomimetic reinforcement learning neural networks.

EXAMPLE MACHINE LEARNING CASE STUDIES

- **COVID-19 Image Classification (cNN/Computer Vision):** Built a convolutional neural network for X-ray classification. (Test Recall: 1.0, F1-score: 0.98)
- **E-Commerce Text Classification (rNN/NLP):** Developed an LSTM-based product categorization model. (Test Accuracy: 0.94)
- **Target Marketing Model (Random Forest/Ensemble Model):** Analyzed customer data and built ensemble classification models tuned with XGBoost for increased purchase rates (Test Accuracy: 0.92)
- **Electrophysiological Data Analysis (Signal Processing/ Clustering):** Collected and analyzed neuronal recordings to modulate engineered neural networks.

EDUCATION AND LANGUAGES

Bioengineering Ph.D., University of Illinois, Urbana-Champaign **2020**
Bioengineering M.S., University of Illinois, Urbana-Champaign **2016**
Mechanical Engineering B.S., University of Puerto Rico, Mayagüez **2014**

CERTIFICATIONS AND CONTINUED EDUCATION

- **MicroBachelors in Programming and Data Structures** (NYUx)
- **Deep Learning Specialization** (DeepLearning.AI)
- **CUDA programming Masterclass with C++** (Udemy)
- **Post-Graduate Program in AI & Machine Learning**: Business Applications (UT Austin)
- **MicroMasters in Data Science** (UCSanDiegoX)

Languages:

- **English -** Native
- **Spanish -** Native
- **French –** Conversational Proficiency