



# Patrick Walsh

## Data Scientist

Data Scientist with over 17 years of experience in data analysis. Background in linguistics with experience in Natural Language Processing (NLP), Large Language Models (LLM), Computer Vision (CV), software development, and cybersecurity. Passionate about Generative AI and all the possibilities that this technology provides to modern business.

## Contact

### Phone

984-301-5062

### Email

patrick.walsh6079@protonmail.com

### Website

<https://www.patrickswalsh.com/>

## Education

2022 - 2025

**M.S. Applied Data Science**

Syracuse University

2019 - 2021 (Magna Cum Laude)

**B.S. Software Development**

University of Maryland Global Campus

2013 - 2014

**A.A. Arabic Language Arts**

Defense Language Institute

## Expertise

- Generative AI
- Machine Learning
- Natural Language Processing (NLP)
- Data Science

## Libraries

Langchain, OpenAI, FAISS, Streamlit, Pandas, NumPy, NLTK, spaCy, Keras, SciKit-Learn, TensorFlow, Matplotlib, Seaborn, OpenCV, BeautifulSoup, Flask

## Experience

### 2024 (February) - Present

ECS Federal

#### Lead Data Scientist

As a Lead Data Scientist working as a contractor with the Department of Homeland Security (DHS)'s Data Services Branch (DSB), I specialize in document classification and text analytics using advanced machine learning techniques. My work focuses on creating and training NLP models, including Naive Bayes, Random Forest, and Support Vector Machines (SVM), implemented through SciKit-Learn. These models are pivotal for classifying multi-page PDF documents, supporting critical DHS operations through accurate multilabel classification.

I also extract key entities from text using Named Entity Recognition (NER) and Regex, enhancing the precision and relevance of data analysis. Beyond model development, I deploy these solutions on AWS platforms like SageMaker and EC2, ensuring scalability and performance. Additionally, I perform UI development with Streamlit, enabling seamless document upload and automated analysis through a user-friendly interface, making actionable insights readily available across DHS.

Additionally, I led development of an agency-wide invoice management platform for senior DHS leadership using Microsoft Power Apps, enabling real-time tracking, approval, cancellation, commenting, and escalation through a centralized UI. I helped design the backend architecture using Microsoft SharePoint and a relational database, supporting secure, scalable data management across multiple business units. I also automated end-to-end invoice workflows with Power Automate, including routing, notifications, audit logging, and exception handling, significantly reducing manual processing. Furthermore, I integrated REST APIs to enable seamless data synchronization between the application UI, backend systems, and external services. Finally, I implemented role-based user interfaces enabling different user groups to securely access and use role-specific features of the application.

### 2023 (October) - 2024 (January)

The Cigna Group

#### Senior Machine Learning Engineer

In my capacity as a Senior Machine Learning Engineer at Cigna, I played a crucial role in advancing Text-to-SQL generation by incorporating Generative AI and OpenAI APIs, with a strong emphasis on integrating LangChain for prompt engineering. Beyond formatting Teradata SQL queries, my responsibilities included making API calls to Large-Language Models (LLM), fine-tuning LLM responses, and utilizing procedural calls to modify SQL responses.

Furthermore, as an R&D AI engineer using LangChain, I actively contributed to groundbreaking projects and explored the forefront of AI technologies. This dual-role experience not only enriched my skill set in text generation, API integration, and SQL manipulation, but also provided me with a deep understanding of prompt engineering, a vital aspect in optimizing language models for specific tasks.



## Languages

- Python
- R
- Java
- SQL
- JavaScript
- C
- HTML/CSS

## Soft Skills

Written and verbal communication, mentorship, adapting to new challenges, resilience to setbacks, interpersonal communication, team building, fostering positive and productive work environments, diligence and follow-through

## Experience (continued)

2023 (January - October)

ECS Federal

### Data Scientist

As a contractor with the Multi-Channel Technologies (MCT) division at the Department of Veterans Affairs (VA), I specialized in developing Generative AI applications. Notably, I created a zero-shot model for automated tagging of case notes based on textual data. My responsibilities encompassed full-stack development of chatbots, utilizing LangChain and Streamlit, alongside Large Language Models (LLM) and semantic search techniques. I integrated FAISS, Chroma DB, and Pinecone for efficient vector database storage, optimizing document processing and storage as vectorized indexes to enhance chatbot functionality.

Through prompt engineering, I enhanced model transparency and user trust in chatbots by designing a system that generated source reports. I also implemented advanced NLP models for analyzing and routing veteran queries, employing modern text classification methods including Retrieval-Augmented Generation (RAG). These enhancements significantly boosted the performance of LLMs, leading to increased user satisfaction and engagement.

Finally, I automated business workflows using Power Apps, Power Automate, and Power BI, developing an integrated ticketing and workforce management system that enabled users to manage employee scheduling, track on-duty status, shifts, vacation, and sick leave, while providing senior leadership with real-time operational visibility through dashboards and analytics.

2017 - 2023

SYSCOM, Inc.

### Data Scientist

As a Data Scientist at SYSCOM, I wrote Python code to conduct data wrangling and configure Deep Learning models, fine-tuning hyperparameters, and executing training jobs within both AWS and Azure environments. One significant achievement was the development and refinement of a computer vision (CV) model adept at accurately identifying and classifying nutrient deficiencies in plants. This breakthrough contributed to improved crop yields and reduced costs for farmers. I took charge of deploying and hosting this model on AWS, ensuring scalability and reliability to cater to clients worldwide.

Additionally, I played a pivotal role in the design and development of an innovative CV model that successfully classifies biofilm pathogens in microscope images, leading to a patented technology. My contributions extended to automating the data preprocessing pipeline for CV model development, significantly enhancing efficiency from a 2-hour process to just 30 seconds.

In the realm of Natural Language Processing (NLP), I created models for Topic Modeling, Keyword Extraction, Sentiment Analysis, and Named Entity Recognition (NER). My expertise also encompassed the development of CV models for image classification and object detection. Engaging directly with customers, I interfaced to understand their business problems, build relationships, and provide effective solutions.

In the realm of data management, I queried, designed, and updated relational databases using SQL and Python. Additionally, I developed graphing and visualization programs, utilizing Python, R, Tableau, and Power BI to analyze medium to large datasets. These efforts underscore my multifaceted contributions in the dynamic field of Data Science at SYSCOM.



## Certifications

- Security + (CompTIA)
- Network + (CompTIA)
- DHS Public Trust (Clearance)
- Microsoft Certified: Azure Data Scientist Associate
- Third Party Risk Management (Security Scorecard)

## Training

- Power Platform Training
- Generative AI / LLM Coursework
- Amazon Cloud Training
- Arabic Immersion (Defense Language Institute)

## Patent

### Method and system for identifying objects in a blood sample

Patent #US-11592657-B2

A system that captures and analyzes images of bodily fluid samples to detect, quantify, and classify biofilm presence, automatically displaying results based on computed biofilm area or volume.

## Experience (continued)

2009 - 2017

U.S. Army

### Data Analyst & Arabic Linguist

In my role as a Data Analyst and Arabic Linguist at the United States Army, I managed a team of 20 data analysts, contributing single-handedly to 25% of the product output—marking the highest volume for the entire division at that time. My responsibilities extended to translating Arabic text and audio into English for further analysis, delivering high-level classified verbal and written reports to customers. Additionally, I played a crucial role in decrypting secure digital communications to exploit organizational weaknesses.

In pursuit of operational efficiency, I developed and implemented streamlined data analysis methodologies, resulting in a notable 15% reduction in analysis time, all while maintaining the highest standards of data accuracy and quality. My leadership skills were further showcased as I led cross-functional teams in the successful completion of complex projects, ensuring on-time delivery of critical milestones and demonstrating adept project management skills.

Recognizing the importance of data accessibility, I spearheaded the development of customized data visualization dashboards. This initiative significantly improved data accessibility and empowered the organization with data-driven decision-making capabilities. Furthermore, I actively coordinated with government agencies, providing high-quality, time-sensitive technical support. My multifaceted contributions, from leadership and project management to linguistic and analytical skills, underscore my dedication to excellence in the dynamic field of Data Analysis and Arabic Linguistics within the United States Army.

## References

### Jeremy Davis

Senior Director, The Cigna Group

Phone: 804-904-5147

Email: Jeremy.Davis@evernorth.com

### Ketan Mane, PhD

Senior Architect, ECS Federal

Phone: 202-509-5274

Email: Ketan.Mane@ecstech.com