Glen Smith

United States Citizen | Secret Security Clearance

Atlanta, GA | (678) 632–1405 | glen.r.smith.jr@gmail.com

https://www.glensmith.pro | linkedin.com/in/glenrsmith

PROFESSIONAL SUMMARY

Data Scientist and Engineer with 7+ years of experience designing and deploying cloud-based data platforms and analytics solutions in academic, commercial, and federal government environments. Experienced in ETL data pipeline development with large-scale datasets, complex business intelligence analyses, developing RESTful backend services, and producing executive-level dashboards to inform decision-making by senior stakeholders.

KEY TECHNICAL SKILLS

Certifications: Google Professional Data Engineer (2021)

Tech Stack

Programming Languages: Python, SQL, JavaScript, ReactJS, R

Platforms: Google Cloud (BigQuery, Dataflow, Vertex AI), AWS, Snowflake, Databricks, Git, Elasticsearch, Hadoop, Spark

Visualization: PowerBI, Tableau, Looker Studio

EDUCATION

Georgia Institute of Technology — Atlanta, Georgia
PhD, Computer Science, Intelligent Systems — Expected: 2026

City, University of London — London, England, UK
MSc, Data Science — Oct 2018

The Johns Hopkins University — Baltimore, Maryland
B.S., Computer Science — May 2017

WORK EXPERIENCE

Data Scientist II - Civilian — Sept 2024 – Sept 2025
AFWERX, United States Air Force — *Remote*

- Developed an in-house, analytics web platform that integrated 25+ datasets (up to Secret classification); this combined GCP's Vertex AI with BigQuery to enable secure, retrieval-augmented generation-based data exploration across the organization.
- Designed and implemented over 20 new return on investment (ROI) measures to align with organization objectives by fusing multiple financial, operational, and portfolio datasets; these measures provided more accurate indicators of portfolio value and mission impact than prior efforts.
- Led the migration of our GCP data ecosystem to AWS gov cloud using Terraform (infrastructure as code) to create an impact-level 5 (highly secure) environment, modernize infrastructure, and improve reliability of internal systems and programs.

Lead Data Scientist — July 2023 – Sept 2024
Virginia Tech Applied Research Corporation — *Arlington, VA*

- Led the design and implementation of several data pipelines using Google Cloud Platform (BigQuery, Dataflow, Cloud Functions, Pub/Sub, Topics), SQL, and Python to enable analytics across 20+ datasets of operational and financial data containing millions of records (10Tb+).
- Created 12 executive-level dashboards that communicated insights such as ROI among the sponsor's portfolio companies, private investment into sponsor portfolio companies, and annual reporting metrics that were presented to the U.S. Congress.
- Led client-facing meetings with sponsor subject matter experts (SMEs), translating mission objectives into technical requirements and analytics roadmaps; built strong partnerships with sponsor analysts and stakeholders to align data capabilities with mission needs.
- Strengthened team execution by establishing SCRUM practices like 2-week sprints, daily standups, and use of Jira as a tracking tool, alongside mentoring junior analysts and engineers on best practices.

Data Engineer Feb 2020 – June 2023

Virginia Tech Applied Research Corporation *Arlington, VA*

- Developed an NLP analytics framework (topic modeling using Word2Vec text embedding, vector-based document retrieval) to enable automated extraction of insight from unstructured text data (publications, patents, news), including detection of emerging technologies, early investment into technologies, and potential threats.
- Delivered multiple sponsor-facing dashboards using Tableau that reduced data reporting time from a month to a few days and enabled leadership to quickly derive insights needed for operational decisions.
- Built ETL pipelines using Azure Data Factory and Python to ingest and transform large, open-source datasets; additionally implemented data lineage tracking using checkpoints to ensure traceability for downstream analytics.
- Developed several backend RESTful API services using Flask, containerized Elasticsearch (Docker), and AWS to support many of the corporation's analytics applications.
- Collaborated with senior leadership to define a comprehensive data management plan, encompassing data security protocols and compliance requirements such as CMMC and NIST 800-171; plan was approved by executive leadership and adopted by all company divisions.
- Worked directly with sponsor to define technical requirements and served as a data engineering subject matter expert (SME) to consult on several programs' data efforts.

Junior Data Scientist Jan 2019 – Jan 2020

Virginia Tech Applied Research Corporation *Arlington, VA*

- Built 3 ad-hoc data search tools using R-shiny and ElasticSearch that automated multi-week data wrangling processes and reduced search and curation time from weeks to a few hours.
- Built and evaluated a Random Forest model in Python to assess operational risk and training readiness for U.S. Air Force pilots; insights were briefed to an Air Force Wing Commander and pilot instructors to help reduce overtraining costs and improve readiness outcomes.
- Created a Python-based topic-modeling and text analytics tool to analyze several military forum datasets; findings were briefed to sponsor leadership and used to shape policy and technology investment decisions.
- Delivered technical briefings and demonstrations of developed tools to company executives and U.S. Air Force leadership.
- Worked closely with other corporate departments (Legal, Accounting, etc.) to support data acquisition and management workflows.

Note: Resume has been abbreviated. Please email me or visit my website if you would like additional information such as PhD research, published papers, and awards and honors.