

Glen Smith

United States Citizen | Secret Security Clearance

Atlanta, GA | (678) 632-1405 | glen.r.smith.jr@gmail.com
glensmith.pro | linkedin.com/in/glenrsmith

PROFESSIONAL SUMMARY

Data Scientist with 7+ years of experience designing and deploying cloud-based data platforms and analytics solutions in secure research and defense environments. Experienced in large-scale integration of complex datasets, developing scalable backend services, and producing executive-level dashboards and ROI metrics that inform decision-making by senior stakeholders.

KEY TECHNICAL SKILLS

Certifications: Google Professional Data Engineer (2021)

Tech Stack

Programming Languages: Python, SQL, JavaScript, R

Platforms: GCP (BigQuery, Dataflow, Vertex AI), AWS, Flask, Git, Elasticsearch, Hadoop, Spark

Visualization: Looker, Tableau, PowerBI

EDUCATION

Georgia Institute of Technology

PhD, Computer Science, Intelligent Systems

Atlanta, Georgia

Expected: Fall 2026

City, University of London

MSc, Data Science

London, England, UK

Oct 2018

The Johns Hopkins University

B.S., Computer Science

Baltimore, Maryland

May 2017

WORK EXPERIENCE

Data Scientist II - Civilian

AFWERX, United States Air Force

Sept 2024 – Sept 2025

Remote

- Developed an in-house analytics platform integrating 25+ heterogeneous datasets (up to Secret classification) and combining BigQuery search with Vertex AI generative intelligence to enable secure 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 performance datasets; these measures provided more accurate indicators of portfolio value and mission impact than prior efforts.
- Led the migration of our GCP data ecosystem using Terraform to impact-level 5 (highly secure) environments to modernize infrastructure and improve reliability of internal systems and programs.

Lead Data Scientist

Virginia Tech Applied Research Corporation

July 2023 – Sept 2024

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 dashboards and data exploration tools that reduced reporting time from a month to a few days and enabled leadership to quickly derive insights needed for operational decisions.
- Built strong partnerships with sponsor leadership, analysts, and stakeholders to align data capabilities with mission needs and ensure analytic products directly supported DoD operational objectives.
- Strengthened team execution by establishing structured agile practices like sprint planning and daily standups, mentoring junior analysts and engineers, and aligning technical execution to sponsor needs.

Data Engineer*Virginia Tech Applied Research Corporation*

Feb 2020 – June 2023

Arlington, VA

- Led implementation of enterprise-scale data architectures using Hadoop, Spark, and Elasticsearch to centralize storage and analysis of corporate- and program-level data.
- Built multiple ETL pipelines to scrape websites for large, publicly-available, and heterogeneous datasets, which significantly improved data availability and reliability for research teams.
- Developed reusable NLP and analytics frameworks (topic modeling using text embedding, search analytics) 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 (Tableau and Python tools) that communicated insights such as ROI among the sponsor's portfolio companies, private investment into sponsor portfolio companies, and annual reporting metrics.
- Developed several backend 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; plan was approved and adopted by all company divisions.
- Worked directly with sponsor leadership to define technical requirements and served as a data engineering subject matter expert (SME) to consult on several programs' data efforts.

Junior Data Scientist*Virginia Tech Applied Research Corporation*

Jan 2019 – Jan 2020

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 predictive machine learning models 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 reduce overtraining costs and improve readiness outcomes.
- Created Python and R-based topic-modeling and text analytics tools to analyze large military forum datasets; findings were briefed to sponsor leadership and used to shape policy and technology investment decisions.
- Delivered technical briefings and demonstrations of engineering tools to company executives and U.S. Air Force leadership.
- Worked closely with other corporate departments (Legal, Accounting, HR) to support data acquisition and management workflows.

RESEARCH & SCHOLARSHIP**I. Refereed Publications**

- **Smith, G. R.**, & MacLellan, C. J. (2025) LEARN: A Hybrid Architecture For Language-Guided Induction of Hierarchical Task Networks. In Twelfth Annual Conference on Advances in Cognitive Systems.
- Zhang, Q., **Smith, G.**, Li, Z., Dong, Y., Harpstead, E., & MacLellan, C. (2025). Dice Adventure: An Asymmetrical Collaborative Game for Exploring the Hybrid Teaming Effects. In Proceedings of the 20th International Conference on the Foundations of Digital Games (pp. 1-9).
- Hannan, D., Nesbit, S. C., Wen, X., **Smith, G.**, Zhang, Q., Goffi, A., ... & MacLellan, C. J. (2024). Interpretable Models for Detecting and Monitoring Elevated Intracranial Pressure. arXiv preprint arXiv:2403.02236.
- Hannan, D., S. C. Nesbit, X. Wen, **G. Smith**, Q. Zhang, A. Goffi, V. Chan, M. J. Morris, J. C. Hunninghake, N. E. Villalobos, E. Kim, R. O. Weber, and C. J. MacLellan. "MobilePTX: Sparse Coding for Pneumothorax Detection Given Limited Training Examples". Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 13, Sept. 2023.
- **Smith, G.**, Zhang, Q., MacLellan, C.J. (2022). Do it Like the Doctor: How We Can Design a Model That Uses Domain Knowledge to Diagnose Pneumothorax. In Proceedings of the AAAI 2022 Spring Symposium on Machine Learning and Knowledge Engineering for Hybrid Intelligence (AAAI-MAKE 2022).

II. Non-Refereed Publications

- **Smith, G.**, Gupta, A., & MacLellan, C.J (2024). Apprentice tutor builder: A platform for users to create and personalize intelligent tutors. arXiv preprint arXiv:2404.07883.
- **Smith, G.**, Shah, R., Adjaye, R. (2017). Approaches to analysing social media data to improve marketing. Unpublished literature review, City University of London, England, U.K.

III. Posters

- **Smith, G.**, & MacLellan, C.J (2025). Bridging the Expertise Gap: A Hybrid AI Architecture for Teacher-Centered Tutor Authoring. Presented at NSF Summit for AI Institutes Leadership (AIVO).
- **Smith, G.**, Gupta, A., & MacLellan, C.J (2024). Apprentice Tutor Builder: A User-Friendly Platform for Building Personalized and Inclusive AI Tutors. Presented at NSF Annual Evaluation Meeting, NSF AI ALOE.

HONORS & AWARDS

CSSE@GT Ph.D. Fellowship: From Prototypes to Products [Software Engineering Fellowship]

2025-2026

NSF Vital Prize Challenge: Mathematical Literacy to Promote a Future STEM Workforce (Semi-Finalists, \$20,000 prize)	2023-2024
The Baltimore Scholars Scholarship [Full-Tuition Grant]	2013-2017
Comcast Leaders and Achievers Award	2013