United States Citizen | Secret Security Clearance

Glen Smith

871 3rd St NW, Atlanta, GA, 30318

Phone: (678)-632-1405 | E-mail: glen.r.smith.jr@gmail.com | Website: https://www.glensmith.pro

PROFESSIONAL SUMMARY

Data Scientist with 7+ years of experience in data engineering and analysis, machine learning, cloud computing, and generative/ knowledge-based AI. Industry background includes the design of data pipelines and analytics dashboards, complex analyses using statistical and machine learning methods to inform business decisions, and full-stack web application development. These efforts were supported by expertise in cloud environments such as Google Cloud (GCP) and AWS, and visualization tools such as Tableau, Looker, PowerBI, and Python/R. Currently in the final year of a PhD in Intelligent Systems at Georgia Tech, conducting research at the intersection of machine learning, human-computer interaction, and agentic AI design. Research applications span intelligent tutoring systems and human-machine teaming.

Certifications/Licenses: Google Professional Data Engineer, 2021 | Secret Security Clearance, exp. 2030

Key Competencies: Data modeling and analysis | Python | machine learning | agentic AI | data visualization | cloud computing | fundamental and applied AI research

EDUCATION

Georgia Institute of Technology | Atlanta, Georgia
PhD, Computer Science, Intelligent Systems | Expected: Fall 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

PROFESSIONAL EXPERIENCE

Data Scientist II - Civilian | 9/9/2024 – 9/30/2025
United States Air Force | Remote
Supervisor: Christopher Drye

Summary: Led the development of cloud-based AI capabilities by integrating 25+ sensitive datasets, modernizing our GCP infrastructure, and designing ROI and mission-impact metrics to improve decision-making across Air Force programs.

Key Accomplishments:
- Led the development of an in-house analytics platform integrating 25+ mission-critical datasets (up to Secret classification), enabling secure data exploration through a hybrid capability combining BigQuery search with Vertex AI generative intelligence.
- Designed and implemented new return on investment (ROI) measures by fusing multiple financial, operational, and performance datasets to provide more accurate indicators of portfolio value and mission impact.
- Led the migration of our GCP data ecosystem using Terraform to modernize infrastructure and improve reliability and security of internal systems and programs.

Lead Data Scientist | 7/1/2023 – 9/6/2024
Virginia Tech Applied Research Corporation | Arlington, VA
Supervisor: Salvatore Pascale

Summary: Led data strategy and analytics initiatives for federal sponsors by building scalable cloud data ecosystems, automating multi-source pipelines, delivering executive dashboards, and aligning technical execution with mission, operational, and congressional reporting needs.

Key Accomplishments:
- Led the design and implementation of several data pipelines using Google Cloud Platform (BigQuery, Cloud Functions, Pub/Sub, Topics), SQL, and Python to enable analytics across 20+ datasets of operational and financial data containing millions of records (1Tb+).
- Delivered 10+ executive dashboards and advanced data exploration tools that reduced manual reporting time by several weeks and enabled leadership to rapidly derive insights needed for strategic and 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, mentoring junior analysts and engineers, and aligning technical execution to sponsor needs.

Data Engineer   2/1/2020 – 6/30/2023
Virginia Tech Applied Research Corporation   Arlington, VA
Supervisor: Salvatore Pascale

Summary: Designed and implemented enterprise Big Data architectures, secure ETL pipelines, NLP and analytics frameworks, dashboards, and backend services across AWS and GCP to centralize corporate data and support sponsor programs.

Key Accomplishments:
- Designed and led implementation of enterprise-scale Big Data architectures across AWS and GCP using Hadoop, Spark, and Elasticsearch to centralize storage and analysis of corporate- and program-level data.
- Built multiple ETL pipelines for large, heterogeneous mission datasets, which significantly improved data availability, performance, and reliability for research teams.
- Developed reusable NLP and analytics frameworks (topic modeling, text embedding, search analytics) to enable automated extraction of insight from unstructured text data (publications, patents, news), including detection of emerging technologies, early investment, 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 cloud services to support multiple of the corporation's analytics applications.
- Worked closely with company executives to draft and refine a corporation-wide data management plan that unifies program- and corporate-level data into a single, efficient framework.
- Worked directly with sponsor leadership to define technical requirements, co-led sprint planning, and consulted on several programs to provide data engineering insight.

Junior Data Scientist   1/2/2019 – 1/31/2020
Virginia Tech Applied Research Corporation   Arlington, VA
Supervisor: Matt Wolfe

Summary: Built internal analytics applications, Python tools, and machine-learning–driven systems that brought automated previously manual workflows, accelerated analytics timelines, and supported operational decision-making. Delivered high-impact briefings to sponsor leadership and adopted engineering solutions that demonstrated value and informed sponsor technology direction.

Key Accomplishments:
- Built 3 internal data search tools using R-shiny and ElasticSearch that automated multi-week data wrangling processes, reducing analysis 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 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 sponsor 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