Jonathan Philip Schlosser

Lead Data Scientist | Generative AI & LLM Specialist | Educator & Speaker

Email: JonathanPhilipSchlosser@gmail.com **Website:** JonathanSchlosser.com
LinkedIn: www.linkedin.com/in/jonathanphilipschlosser **GitHub:** www.github.com/JPSchloss

KEY EXPERIENCE

AlignAI
Lead Data Scientist – Generative AI
October 2024 – Present
- Developed and integrated core generative AI features into AlignAI's platform, including a chatbot for automated use-case documentation and an assistive agent for generating technical, client-specific educational playbooks.
- Consulted on machine learning and AI best practices for a major automotive manufacturer, contributing to model development, deployment planning, and process documentation for a vehicle quality forecasting system.
- Collaborated cross-functionally with internal and external teams to ensure modeling and AI efforts aligned with business goals, infrastructure constraints, and user needs.

Senior Data Scientist – Generative AI
April 2024 – October 2024
- Led multiple ML consulting engagements, including the design and implementation of predictive pipelines for a major auto manufacturer to estimate both parts availability and vehicle delivery timelines to dealerships.
- Worked directly with engineering and operations teams to define model requirements, validate data pipelines, and support deployment across production environments.
- Guided external companies on implementing end-to-end ML technology stacks and AI strategies while also supporting their internal upskilling and the creation of technical enablement resources.
- Developed AI strategy frameworks and recommendations to help client organizations establish long-term AI capabilities and MLOps best practices.

UNC Chapel Hill – School of Data Science and Society
Adjunct Faculty – Master of Applied Data Science
August 2024 – Present
- Taught *DATA 780: Introduction to Machine Learning*, covering supervised and unsupervised learning techniques, regression, classification, decision trees, and introductory neural networks with a dual emphasis on applied implementation and underlying mathematical theory.
- Delivered two semesters of *DATA 785: Deep Learning*, guiding Master's-level students through advanced neural network architectures including CNNs, RNNs, attention mechanisms, Transformers (BERT, T5), and modern LLMs (ChatGPT, Llama, Claude, Gemini).
- Created and delivered hands-on coding labs, case-based assignments, and lecture content aimed at bridging theoretical foundations with real-world deployment challenges, while emphasizing a solid foundation of key concepts, reproducible ML workflows, responsible AI principles, and practical model evaluation techniques.

Stealth Start-Up
Lead AI Developer and Co-Founder
March 2023 – April 2025
- Led the end-to-end development of a generative AI–driven sales enablement assistant designed to guide sales professionals in strategy, execution, and goal attainment using LLM-based automation and workflows.
- Built a fully operational Slack-integrated product leveraging ChatGPT, custom query classification models, and a tailored prompt orchestration layer to deliver contextualized and role-specific sales insights.
- Architected and implemented the full backend stack including prompt delivery logic, query routing, data ingestion pipelines, and infrastructure for model execution and testing.
- Owned the complete technical lifecycle—from initial prototyping to production deployment—while aligning product vision with market needs and user feedback.

Nielsen
Senior Data Scientist
September 2021 – April 2024
- Led the design and implementation of Nielsen's first ML-driven meter fault classification system which modeled raw minute-level audio signal data and meter metadata to detect hardware anomalies, reducing error resolution time from 14 days to under 3 days and replacing the legacy threshold-based logic with a scalable and explainable ML solution.
- Engineered a fully automated pipeline monitoring tool using Python, Spark, and Airflow that ingested and analyzed 10+ GB of data daily across 10+ environments, automated over 50 critical analyses, and transformed a previously

manual multi-day diagnostic process into a streamlined tool with runtimes measured in minutes.
- Achieved a 60x speedup in data processing by reengineering Python-based analysis workflows across both local CLI and Databricks environments—reducing runtime from several hours to just 83 seconds on a 200+ GB dataset from a 10,000+ household nationwide pilot, enabling near real-time evaluation of personal meter device performance.
- Partnered with the Audio Measurement team to develop evaluation metrics and A/B testing infrastructure for pipeline data shifts, including second-level granularity enhancements across key audience measurement streams—gaining executive-level visibility and receiving a cross-departmental performance award.
- Assessed and validated a new high-volume data stream (3+ GB/day) for integration into core crediting processes by designing in-house visualization tools, building quality assurance frameworks, and proactively identifying data gaps, inconsistencies, and optimization paths.

Correlation One
Data Science Lead Instructor & Technical Coach | DS4A: Empowerment
April 2021 – April 2024 | 5 Programs
- Guided 50+ teams (300+ fellows) across five cohorts through full-cycle data science and machine learning projects—including data collection, data wrangling, model development, evaluation, and deployment.
- Supported real-world and socially impactful projects such as measuring redlining's effect on health outcomes, analyzing disparities in COVID vaccine disbursement, and applying deep learning computer vision techniques to detect melanoma in BIPOC individuals.
- Provided consistent technical leadership and instruction across diverse domains, becoming one of the longest-serving and most recognized instructors in the program.

Data Engineering Lead Instructor & Technical Coach | DS4A: Data Engineering
Feb. 2022 – July 2023 | 2 Programs
- Advised 16+ project teams (80+ fellows) in building industry-level data pipelines and data products from raw data ingestion and ETL to ML model integration and deployment using tools like S3, EC2, Lambda, and Redshift.
- Led the instruction in applied data engineering topics including SQL, PostgreSQL, MySQL, Presto, Spark, and PySpark and enabled learners to move from foundational concepts to full end-to-end data engineering solutions using modern tools and cloud infrastructure.

Machine Learning Lead Instructor | DS4A: Colombia Data Science
March 2022 – August 2022
- Delivered 6 lecture sessions to an audience of 1800+ data professionals, with topics focusing on data cleaning, univariate analysis, interaction effects, normalization, regularization, and advanced recommender systems.
- Recognized for high instructional quality and technical depth in a large-scale international program.

Data Analytics Lead Instructor & Curriculum Consultant | Amazon Analytics & Data Science
June 2021 – December 2021
- Designed and taught real-world case studies in analytics and data science using Excel, SQL, Python, and Tableau.
- Helped shape the 14-week curriculum that had such success that it expanded from 70+ fellows to over 600 and ultimately became one of Amazon's most successful internal upskilling programs.

ADDITIONAL EXPERIENCE

Analytics Vidhya – Instructor	Jan 2024
KaggleX – BIPOC Mentorship Program – ML, NLP & Generative AI Mentor	Aug 2023 – Aug 2024
Measure of Music – ML & DS Mentor	Feb 2023, Feb 2024
Google CompSci Research Mentorship Program (CSRMP) – Mentee & Alumni Presenter	Jan 2021 – Mar 2023
Carolina Association for Data Science – Instructor & Mentor	Jun 2021 – Jun 2022
UNC Odum Institute for Research in Social Science – DS, ML & Statistics Consultant	Aug 2020 – Oct 2021
UNC Hussman School of Journalism and Media – Undergraduate Adjunct Instructor	Aug 2019 – Oct 2021
National Consortium for Data Science – DS, ML & Deep Learning Teaching Assistant	Aug 2019 – Aug 2021
State University of New York at Sullivan – Environmental Engineering Adjunct Instructor	May 2017 – Jun 2018
Microsoft TEALS (Technology Education and Literacy in Schools) – Teaching Assistant	Jun 2017 – May 2018
Sullivan County Visitors Association – Director of Social Media & Digital Outreach	Apr 2016 – Dec 2016
Lancaster University Environment Centre – Computational Modeling Research Assistant	Jan 2015 – Sept 2015
Binghamton University – GIS Analyst & Research Assistant	Jan 2014 – Jun 2014
Binghamton University – Environmental Data Analyst & Research Assistant	Sept 2013 – Jun 2014

EDUCATION

PhD Media and Communication	UNC Chapel Hill	Expected Dec 2025
MSc Environment and Development	Lancaster University	*Merit Honors*
BS Environmental Science	Binghamton University	*Cum Laude Honors*

TECHNICAL SKILLS

Core Expertise: SQL (12+ years), Applied Machine Learning (10+ years), Natural Language Processing (10+ years), Python (10+ years), Deep Learning (8+ years), Data Engineering (8+ years), MLOps & Evaluation (6+ years), End-to-End AI Product Development & Deployment (5+ years), Transformer Architectures & Attention Mechanisms (5+ years)

Generative AI & LLM Ecosystem: Generative AI & LLM Systems (4+ years), OpenAI API (3+ years), ChatGPT (3+ years), LangChain (2+ years), Prompt Engineering (2+ years), Chain-of-Thought Prompting (2+ years), RLHF (2+ years), Multi-Modal GenAI Systems (2+ years), Llama (2+ years), Agentic AI (2+ years), Pydantic (2+ years), Claude (1+ year), Gemini (1+ year), Small Language Models (1+ year), smolagents (1+ year), Model Context Protocol – MCP (1+ year)

LLM Training, Tuning & Orchestration: TensorFlow/Keras (7+ years), PyTorch (6+ years), HuggingFace (4+ years), LLM Fine-Tuning & Adaptation – LoRA, QLoRA, PEFT (2+ years), Weights & Biases (2+ years), Vector Databases – Chroma (2+ years), RAG Architectures (2+ years), Custom LLM Evaluation Tools (2+ years), Token-Level Analysis & Hallucination Detection (2+ years), Synthetic Data Generation (2+ years), Multi-Agent Orchestration (1+ year)

Programming & Developer Tooling: Git/GitHub/GitLab (10+ years), Linux (10+ years), HTML/CSS (10+ years), VSCode (10+ years), CLI (10+ years), Tableau (9+ years), JupyterLab/Notebooks (8+ years), Pip (8+ years), MySQL (7+ years), SQLAlchemy (5+ years), PostgreSQL (5+ years), Flask (5+ years), FastAPI (3+ years), Streamlit (2+ years), Cursor (2+ years)

Cloud Platforms & Infrastructure: HTTP (10+ years), AWS – SageMaker, Lambda, EC2, S3, RDS, Redshift (6+ years), Docker (6+ years), Google Cloud & Colab (5+ years), Spark/PySpark (5+ years), Databricks (5+ years), Presto/Trino (2+ years), Apache Airflow (2+ years), Vercel (2+ years)

Strategic Leadership & Cross-Functional Delivery: Academic Teaching & Mentoring (5+ years), Cross-Functional Team Collaboration – Product, Legal, Infra, Design (5+ years), Decision Architecture – GenAI vs Classical ML vs Rule-Based (5+ years), Team Mentorship & Upskilling (4+ years), Explainability & AI Safety (4+ years), Roadmapping & AI Product Strategy (2+ years), B2B AI Enablement (2+ years)

GENERATIVE AI & APPLIED ML PROJECTS

Enterprise Documentation Assistant & Playbook Generator (LLM-Powered) | *AlignAI* | *2024*
Led the design and deployment of two generative AI agents for AlignAI's enterprise platform: a documentation assistant that dynamically generates use-case narratives from structured client input, and an adaptive playbook builder that produces instructional content customized by audience, tone, and technical depth. Built using ChatGPT, Pydantic, FastAPI, and internal prompt logic frameworks to serve customer onboarding and solution engineering teams at scale.

Automotive Parts Availability & Vehicle Delivery Forecasting Models | *AlignAI Consulting Engagement* | *2024*
Developed and productionized two ML pipelines for a major automotive manufacturer: one for forecasting vehicle arrivals to dealerships, and another for predicting supply chain delays in parts delivery. Provided model architecture, validation strategy, and deployment support while also advising internal engineering teams on MLOps infrastructure and scalable model handoff patterns. Contributed to broader AI strategy development and internal technical enablement.

Slack-Based Sales Enablement Assistant (LLM-Oriented) | *Stealth Startup* | *2023 – 2025*
Built a complete LLM-powered assistant for sales professionals using ChatGPT, custom intent classification models, and structured prompting pipelines. Engineered all layers of the stack—from query understanding and prompt routing to interface and backend execution—resulting in a real-time, context-aware Slack application used in sales coaching and performance improvement.

Audience Measurement Fault Classification (Real-Time ML Deployment) | *Nielsen* | *2022 – 2023*
Pioneered the first ML-driven system within Nielsen's Meter Technology team to classify hardware faulting behavior using minute-level time-series audio signal data. Replaced legacy thresholds with an explainable classification model, reducing issue resolution time from 14 days to under 3 days. Embedded directly into operational analysis pipelines across 10+ environments and handling 5GB+ of data daily.

FELLOWSHIPS & CERTIFICATIONS

CompTIA DataX Certification	07/2023 – 07/2026	UNC CITAP Research Fellow	08/2020 – 05/2022
CompTIA Data+ Certification	08/2021 – 08/2024	Summer Institute for Computational	
Google CSRMP Fellowship	01/2021 – 05/2021	Social Science (SICSS)	08/2019
Correlation One DS4A Fellowship	10/2020 – 03/2021	UNC Chapel Hill Park PhD Fellowship	08/2018

For more information about my skills and past experiences, please check out my LinkedIn!