NASA's Glenn Research Center 21000 Brookpark Road Cleveland, OH 44135	Jonathan Cachat, PhD jc@jcachat.com 440-654-1960	Ref:MSG98896917 HRC1404620

ATTN: Onsite Training Purchases Team, NASA Shared Services Center

LLMs & GenAI for Scientific R&D - from data to insight delivery

instructor-led training; flipped learning model

course repo: https://github.com/cachatj/LLM_GenAI_for_Sci_course_2024

RE: provide a comprehensive instructor-led-training program on generative AI (Gen AI) for. This program will leverage a flipped classroom approach to equip participants with the knowledge and skills necessary to utilize Gen AI for tasks such as:

- document summarization,
- content creation,
- code assistance, and
- research support.

TARGET AUDIENCE

intended to instruct 25+ NASA researchers and engineers interested in using LLM & GenAI workflows in their projects. Students are not expected to have previous experience with LLMs, or Gen AI. However, familiarity with python, SQL & cloud data analytics platforms (GCP, AWS, Azure) is a head start.

This course lays out a roadmap, and provides students with learning resources they can get hands-on with immediately. As a flipped learning model, all participants are expected to have done the reading & attempted the exercise *BEFORE* coming together for an in-person session.

INSTRUCTORS

Jonathan Cachat, PhD (Lead) - linkedin.com/in/jcachat
John Cachat, MS - linkedin.com/in/johncachat
Heather Skinner - llinkedin.com/in/heathercacha

Kyle Robinson MS (backup)
Kerim Tericic (backup)
CCV Research, LLC

COURSE OVERVIEW

Open hours, weekly - Mon, Tues, Thursday - 11am - 3pm EST

In-person sessions, bi-weekly - Thursday, Friday 11am - 6pm EST - Computer Lab @ NASA

	Focus / Topic
wk_0	
wk_1	
wk_2	
wk_3	
wk_4	
wk_5	
wk_6	
wk_7	
wk_8	
wk_9	
wk_10	
wk_11	
wk_12	
wk_X	

QUESTIONNAIRE

What experience does the instructor have using Gen AI for science or engineering applications?
Jonathan Cachat, PhD has over a decade of course-work development experience covering data science, analytical chemistry & PROVE THAT: Public Sector Training: history of successfully training engineers, scientists, and business personnel within the public sector.

Subject Matter Mastery: Deep understanding of the topics covered in the statement of work.

Course Content Authority: In-depth knowledge and ability to convey the course material effectively.

What experience does the vendor have conducting Gen AI training courses?

What is the process for selecting qualified, alternative teachers if a substitute instructor is needed?

If a substitute instructor is needed, we have a rigorous selection process to ensure continuity and quality of instruction:

-We maintain a network of qualified professionals with expertise in generative AI, data science, and related fields. This includes colleagues from my work at CCV Research, Cardinal Health, and other organizations.

-Potential substitutes are evaluated based on their technical knowledge, teaching experience, and familiarity with the course material.

-I would personally brief the substitute on the course structure, learning objectives, and any ongoing projects to ensure a smooth transition.

-In the event of a planned absence, I would introduce the substitute to students in advance and remain available for consultation to maintain course continuity.

What experience does the instructor have with facilitating training in a non- academic setting?

As the Director of Data Science & MLops at Torchlight AI and in my various roles as a consultant and technical director, I have extensive experience facilitating training in non-academic settings:

- Led DS, DnA & Engineering teams through complete cloud infrastructure transformation, which involved significant on-the-job training.
- Conducted workshops and training sessions for clients on data science methodologies, cloud technologies, and generative AI applications.
- Developed and delivered training programs for cannabis industry professionals on analytical techniques and regulatory compliance.
- Facilitated knowledge transfer sessions for cross-functional teams in healthcare, finance, and energy sectors.

My approach focuses on practical, hands-on learning experiences that directly apply to real-world business challenges. See links below for previous teaching, seminar & collaborative work of the instructor.

Drugs & the Human Brain - Dr. Jonathan Cachat

Three-dimensional neurophenotyping of adult zebrafish behavior

What platforms and other tools are recommended for assignments and coursework?

Based on industry relevance and ease of use, I recommend the following platforms and tools:

- Google Colab or Jupyter Notebooks for interactive coding exercises
- GitHub for version control and collaborative project work
- Hugging Face for accessing pre-trained models and datasets
- TensorFlow and PyTorch for deep learning assignments
- LangChain and LlamaIndex for building generative AI applications
- Google Cloud Platform (specifically VertexAI) for cloud-based machine learning projects
- Streamlit for creating interactive web applications to showcase projects

These tools will provide students with hands-on experience using industry-standard technologies.

What datasets are recommended for assignments and coursework?

To ensure relevance and diversity, I recommend the students use their own datasets before publicly available options.

“Chat with your PDF Library”, “Ask questions directly to your data”, “co-develop on the projects codebase with agents & tools”

PDFs, Documents, Data, Databases, Multimedia - whatever the learner is interested in interacting with and utilizing in the weeks exercises, notebooks, readings. This also allows students to learn about data pre-processing, standardizing and loading. When the use-case is “me” the effort needed to overcome & learn by doing is needed.

I will encourage everyone to spend more time looking for Github Repos or Cloud Engineering notebooks or cookbooks and attempting to push their data of interest through the analysis more so that starting with public, faceless generic data.

There are many public datasets available:

- Common Crawl dataset for large-scale text analysis
- Wikipedia dumps for knowledge-based tasks
- ImageNet for computer vision projects
- LibriSpeech for speech recognition tasks

- UCI Machine Learning Repository for a variety of structured datasets
- Kaggle datasets for real-world problem-solving scenarios
- OpenAI's GPT-3 API for generative text tasks (with appropriate usage limits)
- Public sector datasets from data.gov for government and policy-related projects

Provide references for previous Gen AI courses that have been taught.

While I have not previously taught a dedicated Generative AI course, my experience in related fields positions me well to deliver high-quality instruction in this area:

1. As Director of Data Science & MLops at Torchlight AI, I led teams in implementing advanced AI technologies, including several production-deployed LLM chatbots and agents using LangChain, LlamaIndex, and Vector databases. This hands-on experience with cutting-edge AI tools directly informs the practical aspects of the proposed course.
2. In my role at Cardinal Health, I developed and delivered training on AI/ML pipelines, which included elements of generative AI for internal dev tools and user experience improvements. This experience in corporate training demonstrates my ability to communicate complex AI concepts effectively.
3. As a consultant with CCV Research, I've guided numerous clients in implementing generative AI solutions across various industries. This broad exposure to real-world applications of generative AI will enrich the course content with relevant case studies and practical examples.
4. My academic background, including a Ph.D. in Neuroscience with a focus on psychopharmacology, provides a strong foundation in the underlying principles of machine learning and neural networks that are crucial to understanding generative AI.
5. I've authored two books and over 50 peer-reviewed articles, demonstrating my ability to articulate complex technical concepts clearly and effectively – a crucial skill for teaching advanced AI topics.

While these experiences don't constitute formal Gen AI course instruction, they collectively represent a depth of knowledge and practical application in AI and machine learning that will translate directly into effective teaching of Generative AI concepts and applications.

Describe experience with teaching in a non-academic setting.

My experience teaching in non-academic settings is extensive and varied, drawing from my professional roles across multiple industries:

1. Corporate Training at Cardinal Health (2021-2024): As a Senior Data Engineer in AI/ML pipelines, I regularly conducted training sessions for cross-functional teams. These sessions covered topics such as data engineering best practices, machine learning fundamentals, and the integration of AI solutions into existing business processes. I developed and delivered workshops tailored to both technical and non-technical audiences, ensuring that complex concepts were accessible to all participants.
2. Data Science Leadership at Torchlight AI (2022-2024): In my role as Director of Data Science & MLops, I led comprehensive training initiatives during our cloud infrastructure transformation. This involved designing and delivering a series of hands-on workshops covering:
    - Modern data architecture principles
    - Cloud-based data processing using Google Cloud Platform
    - Implementation of MLops practices These sessions were crucial in upskilling our team and ensuring a smooth transition to new technologies.
3. Cannabis Industry Training (2017-2021): As Scientific Director at Midway Labs and Hocking College Cannabis Lab, I spearheaded educational programs for industry professionals. This included:

    - Developing and teaching courses for the nation's first accredited Associate's Degree in Cannabis Lab Technician
    - Conducting workshops on analytical techniques and regulatory compliance for cannabis industry stakeholders
    - Training lab personnel on state-of-the-art equipment and methodologies
4. Consultancy Work with CCV Research (2015-Present): Through my consultancy, I've designed and delivered numerous training programs for clients across various sectors including finance, AdTech, energy, and healthcare. These programs often focus on:
    - Data-driven decision making
    - Implementation of AI and machine learning solutions
    - Best practices in data science and engineering
5. Startup Environment (2014-2018): As co-founder and CTO of multiple cannabis industry startups, I developed and led training sessions for team members on topics such as:
    - Data-driven cultivation techniques
    - Use of IoT devices for grow room intelligence
    - Analytical methods for product quality assessment

Throughout these experiences, I've honed my ability to adapt teaching methods to diverse audiences, from executives to technical professionals. I emphasize hands-on learning, real-world application, and interactive problem-solving to ensure engagement and knowledge retention. My approach focuses on bridging the gap between theoretical concepts and practical implementation, which is crucial in non-academic settings where immediate application of knowledge is often required.

What platforms and tools would students use for assignments and course work?

Local - VSCode, JetBrains, conda/pip/poetry/git

Cloud - VertexAI workbench, BigQuery Studio - jupyter notebooks

Tools will be selected based on their relevance in the industry and their ability to support a wide range of generative AI applications. Students will gain experience with:

- Cloud-based development and deployment
- Large-scale data processing and analysis
- Building and fine-tuning generative models
- Creating end-to-end AI applications
- Collaborative development practices

By using these platforms and tools, students will develop a robust skill set that aligns with current industry practices in generative AI. The hands-on experience with these tools will prepare students for real-world scenarios they're likely to encounter in their professional careers.

Syllabus and course curriculum for previous courses is optional but welcome.

PROPOSAL DELIVERABLES

- ☑ ~~Vendor will deliver course syllabus 2 weeks before training start date. (Syllabus in Git Repo)~~
- ☑ ~~Vendor will deliver course lesson plans before the training start date.~~
- ☑ ~~Vendors will deliver course materials before the training start date.~~
- ☑ ~~Vendor will deliver a list of training tools 4 weeks before training start date.~~
- ☑ ~~Vendor will deliver a list of instructors 2 weeks before training start date.~~
- ☑ ~~Vendors will use only course materials that are Section 508 compliant.~~
- ☑ ~~Vendor will deliver all training sessions before October 30, 2024.~~

COSTS & COMPENSATION

Services Provided

- Compute & storage environments & GCP instances during length of course.
- Custom, on-demand LLM & GenAI Coursework w/ notebook library & starter code repos.
- Preparation, modification and validation of student assignments in real-time
- 3 Live Instructors
- 24hr in-person instruction, 108 open chat hours (remote office) from lead instructor
- Bi-weekly Travel to CLE

Total Cost: $23,500

Payable to: CCV Research, LLC / Jonathan Cachat, PhD

OPEN-SOURCE RESOURCES

Course material & assignments will be build from, or based on the following open-source, authoritative resources.

- https://github.com/GoogleCloudPlatform/generative-ai
- https://github.com/microsoft/generative-ai-for-beginners
- https://github.com/GURPREETKAURJETHRA/Advanced_RAG

Google Cloud SKills Boost

- **01 - Introduction to Generative AI**
  - https://www.cloudskillsboost.google/paths/118/course_templates/536
- **02 - Introduction to Large Language Models**
  - https://www.cloudskillsboost.google/paths/118/course_templates/539
- **08 - Inspect Rich Documents with Gemini Multimodality and Multimodal RAG**
  - https://www.cloudskillsboost.google/paths/183/course_templates/981