Thomas Jaensch

Machine Learning Engineer
Mount Rainier, MD 20712
thomasjaensch@gmail.com
+1202-2538183

Versatile software developer, currently mostly interested in all things ML/MLOps, LLMs/LLMOps, Cloud (GCP, Databricks, AWS), TensorFlow, Spark, etc. I have programmed in Bash, Go, Java, JavaScript, Python, Ruby, Scala, TypeScript and others in the past 10+ years, without being an expert in any one of them. Just whatever needed to get done, extended, or made the most sense at the time.

Authorized to work in the US for any employer

Work Experience

Machine Learning Engineer

Booz Allen Hamilton - Washington, DC
March 2022 to Present

Enterprise Data Analytics and Visualization (EDAV), CDC, 2022-Present
ML and Data Analytics in Databricks/Azure environment.
• Provide Databricks platform support for internal and external customers with a focus on ML capabilities
• Create and improve EDAV Databricks documentation
• Create and demo Machine Learning proof of concept workflows to customers and stakeholders
• Deployed Deep Learning chest x-rays image ML application https://github.com/scotthlee/hamlet in Azure Databricks environment that was previously siloed and inaccessible except for original dev team
• Developed Spark NLP topic modeling application workflows in Databricks to easily analyze public CDC user comments
• Built a Streamlit app using the Azure OpenAI edav-openai-pilot workspace to create a Q&A-on-custom-documents application and deployed on EDAV posit Connect environment
• Created various Databricks workflows to use Meta's Llama 2 models inside CDC environment without the need for external dependencies

AI Ops Genomics Pipeline, CDC, 2022
• Deployed Deepmind's AlphaFold model (https://github.com/deepmind/alphafold/) for the prediction of protein structures to Google Cloud Platform's Compute Engine leveraging a Nvidia machine image with 12 vCPUs, 85 GB of RAM, a 100 GB boot disk, the model weights on an additional 3 TB persistent disk, and an A100 GPU
• Created Jupyter notebooks on terra.bio as well as GCP's AI Platform to visualize the application's output PDB files as 3D images

Data Engineer

Booz Allen Hamilton - Washington, DC
May 2019 to August 2022

Data Engineer, National Cancer Institute CTRP, NIH, 2021-2022

Engineered & monitored data workflows in AWS environment. Software tools included Amazon Aurora RDBMS, AWS DynamoDB, AWS ECS, AWS Fargate, AWS Lambda, AWS OpenSearch/ElastiCache, AWS Step Functions, AWS SQS, Docker, Docker Compose, GitLab, Python, Travis CI/CD.

• Rewrote ctrp-strap-cleanup-lambda repo to run as a serverless, scheduled AWS Fargate Docker deployment to avoid AWS Lambda time constraints
• Added SEC Prior Therapy to CTS API ELT pipeline and trial returns
• Extended nci-ctsapi-etl Python repo with CTS API Enrichment from external APIs for Structured Eligibility Criteria data
• Wrote proof of concept Python/SQL script to populate data in PA DB in order to see how these fields display/function in Scientific Trials Analytics Program (STRAP) UI
• Wrote Python/SQL script to clean up biomarker display names in CTRP source AWS Aurora RDBMS
• Wrote Python script to add new users to the PA/PO/Registry/Accrual applications on the database side
• Wrote Python script to load Structured Eligibility Criteria data into PA AWS Aurora DB in reusable format from Excel spreadsheets
• Wrote Python script to automate STRAP enrichment testing after EVS API upgrade

Full Stack Developer, NCHS Hospital Survey, CDC, 2020-2021
Provided full-stack application development support as a Data Engineer for the National Center for Health Statistics (NCHS), a branch of the Centers for Disease Control and Prevention (CDC), for the Registration Portal as well as the Data Processing and Collection workstreams. Software tools included ActiveMQ, Angular 2, Cypress, GitLab, Java, Spring Boot, SQL Server, and TypeScript.

Data Engineer, Opioid Data Warehouse, FDA, 2019-2020
Worked as a member of the Data Engineering team on data ingest tasks leveraging tools like Docker, ElasticSearch, GitLab, Kafka, Kibana, NiFi, and PostgreSQL.

Software Developer
ERT, Inc - Silver Spring, MD
March 2016 to March 2019

NOAA Affiliate assigned to the Software Engineering Support Branch for the https://data.noaa.gov/onestop project initiated to improve discovery and access to NOAA data. Assisted in the implementation of a NOAA federal enterprise system by improving and connecting the existing metadata systems and building the next generation of metadata tools. Agile development, Bash, ElasticSearch, D3, Golang, Java (Spring Boot framework), JSON, Linux, Python, XML, XPath, XSLT. Data processing, conversions, and transforms.

Video Content Implementer/Developer
AT&T - Bothell, WA
May 2015 to September 2015

Online Content Specialist
National Center for Biotechnology - Bethesda, MD
December 2012 to December 2014

Senior Production Specialist
American Association for the Advancement of Science - Washington, DC
November 2011 to November 2012

Web Production Coordinator
American Association for Clinical Chemistry - Washington, DC
July 2010 to November 2011

Indexer
Alexander Street Press - Alexandria, VA
June 2008 to June 2010

Education

MA in English And History
University of Stuttgart - Stuttgart
1997 to 2003

Links

http://tjaensch.github.io/

https://github.com/tjaensch

https://www.linkedin.com/pub/thomas-jaensch/21/526/3b8

Certifications and Licenses

AWS Certified Developer Associate
June 2021 to June 2024

https://www.credly.com/badges/ad4d63d4-0242-4235-81a5-b6ae6231c2e2/public_url

Google Cloud Professional Machine Learning Engineer
September 2021 to September 2025

https://www.credential.net/94d2d4d9-d435-497e-9f1a-88b9a0ff8d37#gs.cprcdt

AWS Certified Solutions Architect – Associate
November 2021 to November 2024

https://www.credly.com/badges/085e8d50-3a77-433f-8bc3-95f98485efa7

AWS Certified Machine Learning - Specialty
February 2022 to February 2025

https://www.credly.com/badges/5c714d4c-4927-4d68-bd80-6282d8dd7642

TensorFlow Certified Developer
July 2022 to July 2025

https://www.credential.net/97736706-d083-4ce1-81b4-aa42b4895381#gs.54z1i2

Machine Learning Engineering for Production (MLOps)
July 2022 to Present

https://www.coursera.org/account/accomplishments/specialization/certificate/2HXXZWF5GUWJ

Databricks Certified Associate Developer for Apache Spark 3.0
August 2022 to August 2024

https://credentials.databricks.com/892eb3c3-eff5-43a8-b1e4-eca3e2dec48e

Databricks Certified Data Engineer Associate
May 2023 to May 2025

https://credentials.databricks.com/3b188e2f-6f65-491c-afea-e6422bdfe805#gs.xcl0dw

Databricks Certified Machine Learning Professional
May 2023 to May 2025

https://credentials.databricks.com/ddac5ad4-3de1-4f03-98fd-2514b1aeabc8

Generative AI with Large Language Models
July 2023 to Present

https://www.coursera.org/account/accomplishments/certificate/VARDBTF998YU

Google Cloud Professional Cloud Architect
December 2021 to December 2025

https://www.credential.net/24cbb772-b538-41b0-9410-e2542f450ec7

Publications

Diving into GCP’s BigQuery ML & AutoML Tables with Covid19 Big Data
https://medium.com/@tebugging/diving-into-gcps-bigquery-ml-automl-tables-with-covid19-big-data-3208b2a0d702
March 2022

Detailed walkthrough how to build Machine Learning models with GCP's BigQuery ML and AutoML on New York Times' public Covid19 dataset

Databricks End-To-End Machine Learning - Create An Ingest-To-Serving MLOps Pipeline
https://medium.com/@tebugging/databricks-end-to-end-machine-learning-create-an-ingest-to-serving-mlops-pipeline-2f739f90e65f
October 2022

Hands on workflow how to build a Machine Learning XGBoost multi-class classification model all the way from real world data ingest (public CDC dataset) to model deployment and making real-time predictions. And once done, automate the whole enchilada MLOps-style.

Train and deploy a text classification model with Spark NLP, BERT transfer learning, MLflow, and Databricks
https://towardsdev.com/train-and-deploy-a-text-classification-model-with-spark-nlp-bert-transfer-learning-mlflow-and-6541e5c61e24
February 2023

Step by step instructions how to train a binary text classification model via transfer learning on a pretrained BERT (Bidirectional Encoder Representations from Transformers) model and make batch predictions with the fine-grained model on new data

Quick Topic Modeling with Spark NLP and Colab

https://medium.com/@tebugging/quick-topic-modeling-with-spark-nlp-and-colab-c74918e02c5

March 2023

How to quickly get a sense of topics discussed in a large corpus of New York Times article comments with a few lines of Spark NLP code in a Colab notebook

Summarize any text based document with 8 lines of Python and LlamaIndex

https://medium.com/@tebugging/summarize-any-text-based-document-with-8-lines-of-python-and-llamaindex-17335986877b

September 2023

Use the power of LLMs to summarize or ask questions about any custom document without having to write model plumbing code, LangChain or vector database stuff whatsoever.