

Jinen Setpal

+1 (765)-490-1435 | jinen@setpal.net | github.com/jinensetpal | jinen.setpal.net

EDUCATION

Purdue University

Bachelor of Science in Data Science

Aug. 2021 – May 2024

West Lafayette, IN, USA

Relevant Coursework:

- **Graduate Level:** Deep Learning, Advanced Topics in Reasoning with Large Language Models, Applied Regression Analysis, System Security
- **Undergraduate Level:** Data Mining & Machine Learning, Large Scale Data Analysis, Intro to AR/VR, Embedded Systems, Data Structures & Algorithms, Linear Algebra, Statistical Theory, Multivariate Calculus, Probability, Statistical Programming

EMPLOYMENT

Machine Learning Engineer

DagsHub

Jun. 2022 – Present

Tel Aviv, Israel

- Developed PyTorch and TensorFlow dataloaders leveraging intelligent prefetching, automatic path-column and datatype detection, data streaming and automated tensorization towards the [Data Engine](#).
- Developed and deployed [DPT](#): a conversational agent that enables users to interact with DagsHub documentation, and debug machine learning projects incorporating tools integrated within the DagsHub stack.
- Developed a data streaming client by monkeypatching Python's `open()` and extending FUSE to lazily pull files from a specified remote using DagsHub's web APIs.
- Built trainer integrations (automatic data, model, experiment and artifact logging) with HuggingFace's Transformers library, the PyCaret framework, and YOLOv8.
- Implemented and deployed open-source data science projects reproducing and extending past research.

PUBLICATIONS

BoilerBot: A Reliable Task-Oriented Chatbot Enhanced with Large Language Models

Oct. 2023

2nd Proceedings of Alexa Prize TaskBot (Alexa Prize 2023). Hu, Setpal, et al.

Purdue University, USA

- Fine-tuned 8-bit quantized large language models using LoRA for downstream tasks such as task title augmentation and patching failures within speech recognition.
- First-author on the grant proposal; earned \$250,000 in funding and an AWS account with unrestricted compute.
- Extended Amazon's COBOT (Conversational Bot) Toolkit, integrating custom APLs and logic modules for constraint-based state management.
- Developed custom CI/CD pipelines for monolithic server and lambda deployment with containerized WSGI for versioned data updates based on user annotations.

CutLang V2: Advances in a runtime-interpreted analysis description language for HEP data

Jul. 2021

Frontiers in Big Data, 4, 27. Ünel, Sekmen et al.

CERN, Switzerland

- Developed Interpreter Functions through lexical analysis using Flex & Bison (.cpp).
- Setup CI/CD Scripts w/ Automated Email Delivery using GitHub Actions & SendGrid.

ArchiMeDe @ DankMemes: A New Model Architecture for Meme Detection

Dec. 2020

7th Evaluation Campaign, Final Workshop, EVALITA 2020. Setpal, Sarti

Turin, Italy

- Achieved .7664 F1-Score on test dataset (+.2466 baseline) w/ Video Presentation during final workshop.
- Built a multimodal ensemble using transfer learning by fine-tuning AlexNet, DenseNet & ResNet.

INDEPENDENT RESEARCH

Black-Box Multigroup Generalization

Nov. 2022 – Present

<https://dagshub.com/jinensetpal/lint.git>

Purdue University

- Improved Worst-Group Accuracy by formalizing inductive loss functions that leverage implicit biases and (approximate) translational equivariance in CNNs to prevent shortcut learning.
- Parameterized mask reliability using a 2-way 5-shot siamese model minimizing triplet loss, used as the secondary cost function setting up a bi-leveled optimization task.
- Developing an interpretable basis for parameter optimization to reduce the task to a convex optimization problem, guaranteeing converge to the global minima **without utilizing group information**.