Updated: June 17, 2024

Jinen Setpal

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

Education

Purdue University Aug. 2021 – May 2024
Bachelor of Science in Data Science *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 Jun. 2022 – Present
DagsHub *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. Examples: CheXNet, Panoptic Deeplab, YOLOv6.

Systems Developer Sep. 2020 – Jul. 2021
Teachiq AB / exam.net *Stockholm, Sweden*

- Packaged custom security implementations by forking open source xmodmap(.c) utility to a node.js module for exam.net's linux-based exam-delivery kiosk application.
- Reproducibly exploited the assessment kiosk on exam.net's web client and recommended mitigations.

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
7^{th} *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**.

Funded Research

Drone Video Object Recognition Jan. 2022 – May 2023
NSF Award 21204301 – PI: Prof. Yung-Hsiang Lu *Purdue University*

- Team lead over the Spring 2023 Semester. Our team attempted to leverage Gazebo, ROS2 & the previous year's scoring function to develop a multi-agent reinforcement learning approach to the sample solution. Won 2nd Place for the **Undergraduate Research Expo Award** under Purdue College of Science.
- Developed an architecture for split-confidence resolution, achieving .9937 test accuracy as part of the reference solution made for the IEEE international autonomous UAV competition. Bootloader patching and setup for linux-based drones with OpenVINO accelerated IoT.

A Systematic Study of Cryptographic Function Identification Approaches in Binaries Aug. 2021 – Dec. 2022
NSF Award 2047991 – PI: Prof. Christina Garman *Purdue University*

- Employed rudimentary techniques within NLP to establish a baseline approach for reconstructing cryptographic functions from disassembler code used to generate corresponding binaries.
- Evaluated current state-of-art classification tools against rigorous benchmark scripts. Currently under peer-review.

Trade Secrets

Semi-Supervised Class Activation Mappings for Target Localization & Super-Resolution Sep. 2021 – Apr. 2022
Final Presentation, TE AI Cup 2022. ***Setpal****, et al.* *TE Corporate, UK*

- Won the **Best Innovation Award**, developing subclassed TensorFlow layers for accurate, efficient prediction over classes with minute differences.
- By evaluating feature vectors from the model's penultimate convolutional layer over a dynamic weight threshold, we generate a bounding box to localize the region of the image critical to the final classification.

Leveraging Latent Features for Modular Multiclass Classification Sep. 2021 – Apr. 2022
Final Presentation, TE AI Cup 2022. ***Setpal****, et al.* *TE Corporate, UK*

- Designed & developed a novel modular, scalable architecture for classification achieving .99846 real-data classification accuracy over a +.2466 baseline.
- Implemented a latent feature aggregator network to enable minimal re-training for appending and removing target connectors from the multi-class classifer.

Conference Presentations

The Machine Learning Angle for Open Source Science 25^{th} Oct. 2023
The Linux Foundation Member Summit (LFMS) 2023 *Monterey, CA, USA*

Interpretability Tools as Feedback Loops 30^{th} Nov. 2022
Toronto Machine Learning Summit (TMLS) 2022 *Toronto, Canada*

Technical Skills

Languages: Python, C, C++, x86 Assembly, Java, Kotlin, Bash, JavaScript, MATLAB, R, SQL, ROS2
Frameworks: PyTorch, JAX, TensorFlow, Keras, NumPy, Pandas, Pillow, ROOT, Matplotlib, FUSE, Node.js, Express.js
Tools: Git, MLFlow, DVC, Docker, Radare2, Ghidra, TravisCI, GitGuardian, Kubernetes, Gazebo
Cloud Utilities: Google Cloud Console (Compute, Networking, Storage), Amazon Web Services (Redshift, ECR, ECS, S3, Sagemaker, CodePipeline, CodeCommit, CloudWatch, CloudFormation, Lambda), Azure Pipeline, GitHub Actions

Teaching

Course Instructor Aug. 2022/2023 – Dec. 2022/2023
CS 39000 – Web Application Development @ Purdue University *West Lafayette, IN, USA*

- Curriculum design and course instructor for a two-credit course. Net enrollment: 100 students.
- Covered HTML/CSS, JavaScript, React, Node.js, Express.js, MongoDB, Web Security & Cloud Hosting.

Undergraduate Teaching Assistant Feb. 2022 – May 2022
STAT 190 – Topics in Statistics for Undergraduates @ Purdue University *West Lafayette, IN, USA*

- Lab instructor for Purdue's Corporate Partner MISO, developing industry solutions using Data Science.
- Graded assignments, held office hours, conducted code review. Taught classes on git, CI/CD & Data Mining.

Projects

[Re] Graph R-CNN for Scene Graph Generation Sep. 2023 – Nov. 2023
DagsHub × ML@Purdue Hackathon Fall 2023 *West Lafayette, IN, USA*

- Reproduced the Graph R-CNN for Scene Graph Generation paper as the template repository for the Scene Graph Generation Challenge in the beginner section of the DagsHub × ML@Purdue Hackathon.
- Included functions for model training, inference, registration and data processing for the VisualGenome dataset.

Time-Series Modelling for Outbreak Prediction Oct. 2021
CERN's The Port Hackathon *West Lafayette, IN, USA*

- Predicted *oidium* outbreaks within vineyards in Germany.
- Achieved test accuracy of 0.995 (± 0.0025) when predicting outbreak risk, trained on daily data from 2013 - 2020.

Embedded Realtime Semantic Segmentation Feb. 2021 – Apr. 2021
Independent *Mumbai, India*

- Embedded DeeplabV3+ with a MobileNetsv3 backbone to an android application.
- Established a data conversion pipeline (NV21 $\rightarrow$ YUV_420_888 $\rightarrow$ JPEG $\rightarrow$ Bitmap $\rightarrow$ TensorImage), with an inference framerate of $\approx 25 fps$ on a Qualcomm SM8150 Snapdragon 855 (7 nm) processor.

Technical Presentations

Omnipredictors 25^{th} Apr. 2024
ECE ML Reading Group *West Lafayette, IN, USA*

Direct Preference Optimization 28^{th} Mar. 2024
ML@Purdue Reading Group *West Lafayette, IN, USA*

Towards Monosemanticity (SAEs) 8^{th} Feb. 2024
ML@Purdue Reading Group *West Lafayette, IN, USA*

A Mathematical Framework for Transformer Circuits 1^{st} Feb. 2024
ML@Purdue Reading Group *West Lafayette, IN, USA*

Deduplicating Training Data Makes Language Models Better 2^{nd} Nov. 2023
CS 592-LLM — Advanced Topics in Reasoning with Large Language Models *West Lafayette, IN, USA*

Groups, Rings & Fields 25^{th} Oct. 2023
ML@Purdue Reading Group *West Lafayette, IN, USA*

OOD Generalization via Risk Extrapolation 28^{th} Mar. 2023
ML@Purdue Reading Group *West Lafayette, IN, USA*

Enforcing Group-Transformation Invariance in MLPs 24^{th} Feb. 2023
ML@Purdue Reading Group *West Lafayette, IN, USA*

Interpretability Tools as Feedback Loops 21^{st} Jan. 2023
Boilermake X *West Lafayette, IN, USA*

AlphaTensor 15^{th} Nov. 2022
SIGAI Reading Group *West Lafayette, IN, USA*

Workshops

Utilizing HPC with SLURM — 31^{st} Mar. 2024
Catapult Hacks — *West Lafayette, IN, USA*

Parameter-Efficient LLM Fine-Tuning — 9^{th} Nov. 2023
Purdue Hackers × ML@Purdue — *West Lafayette, IN, USA*

Intro to Data Engine — 21^{st} Oct. 2023
DagsHub × ML@Purdue Hackathon Fall 2023 — *West Lafayette, IN, USA*

Intro to DagsHub — 14^{th} Oct. 2023
DagsHub × ML@Purdue Hackathon Fall 2023 — *West Lafayette, IN, USA*

Training Word Embeddings — 14^{th} Sep. 2023
Purdue Hackers × ML@Purdue — *West Lafayette, IN, USA*

Model Registry and Deployment with MLFlow — 6^{th} Apr. 2023
DagsHub — *West Lafayette, IN, USA*

Experiment Tracking for Machine Learning with MLFlow — 24^{th} Mar. 2023
DagsHub — *West Lafayette, IN, USA*

Linear Regression — 8^{th} Mar. 2023
ML@Purdue Spring 2023 Workshops — *West Lafayette, IN, USA*

MLOps for Research Reproducibility — 6^{th} Oct. 2022
SIGAI Fall 2022 Workshops — *West Lafayette, IN, USA*

Outreach

Program Chair — Oct. 2023 – Nov. 2023
DagsHub × Purdue Hackathon Fall 2023 — *West Lafayette, IN, USA*

Technical Advisor, Officer — Jan. 2022 – May 2024
ML@Purdue / SIGAI — *West Lafayette, IN, USA*