Jinen Setpal

+1 (765)-490-1435 | jsetpal@cs.purdue.edu | github.com/jinensetpal | jinensetpal.github.io

EDUCATION

Purdue University

Bachelor of Science in Data Science **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

- 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 <u>YOLOv8</u>.
- Implemented and deployed open-source data science projects reproducing and extending past research. Examples: CheXNet, Panoptic Deeplab, YOLOv6.

PUBLICATIONS

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

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.
- 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	J.,1 2021	
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 through AlexNet, DenseNet & ResNet pre-trained networks.

INDEPENDENT RESEARCH

Interpretable Risk Minimization https://dagshub.com/jinensetpal/lint.git

Nov. 2022 – Present 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.

Aug. 2021 – May 2024 West Lafayette, IN, USA

Tel Aviv, Israel

Jun. 2022 – Dec. 2023

Oct. 2023

TRADE SECRETS

Semi-Supervised Class Activation Mappings for Target Localization \mathfrak{G} Super-Resolution

Final Presentation, TE AI Cup 2022. Setpal, et al.

- 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

Final Presentation, TE AI Cup 2022. Setpal, et al.

- 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.

Funded Research

Drone Video Object Recognition

NSF Award 21204301 – PI: Prof. Yung-Hsiang Lu

- 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

NSF Award 2047991 - PI: Prof. Christina Garman

- 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.

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
Teaching	

Course Instructor CS 39000 – Web Application Development @ Purdue University

• 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

STAT 190 – Topics in Statistics for Undergraduates @ Purdue University

- 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.

Technical Skills

Languages: Python, C, C++, x86 Assembly, Java, Kotlin, Bash, JavaScript, MATLAB, R, SQL, ROS2 Frameworks: PyTorch, 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

TE Corporate, UK

Sep. 2021 – Apr. 2022

Sep. 2021 – Apr. 2022 TE Corporate, UK

Jan. 2022 – May 2023

Purdue University

Purdue University

Aug. 2021 – Dec. 2022

Aug. 2022/2023 - Dec. 2022/2023 West Lafayette, IN, USA

> Feb. 2022 – May 2022 West Lafayette, IN, USA