NURBEK TASTAN

nurbek.tastan@mbzuai.ac.ae ◇ +971 (54) 346-9180
website: tnurbek.github.io ◇ linkedin: nurbek-tastan ◇ github: tnurbek

PROFILE

Doctoral researcher with over four years of professional experience in machine learning with an understanding of the modern machine learning pipeline and deep mathematical/probabilistic thinking. Expert in modeling complex machine learning, computer vision, deep learning, and distributed optimization algorithms. My research primarily focuses on the areas of **federated learning** and ensuring the **trustworthiness** of artificial intelligence systems.

EDUCATION

PhD in Machine Learning, Mohamed bin Zayed University of Artificial Intelligence 2023 - Present
Abu-Dhabi, United Arab Emirates
Relevant Coursework: Foundations and Advanced Topics in Machine Learning, Advanced Probabilistic and Statistical Inference, Advanced Topics in Continuous Optimization, Federated Learning, Safe and Robust Computer Vision, Advanced Machine Learning

MSc in Machine Learning, Mohamed bin Zayed University of Artificial Intelligence 2021 - 2023
Abu-Dhabi, United Arab Emirates
Relevant Coursework: Mathematical Foundations for AI, Machine Learning, Advanced ML, Trustworthy AI, Probabilistic and Statistical Inference, Causality, Reinforcement Learning, Optimization

BSc in Systems of Information Security, International Information Technology University 2017 - 2021
Almaty, Kazakhstan
Relevant Coursework: Programming (Python, C++, Java), ML, Cybersecurity, a lot of Maths
GPA **3.94**/4.00 (Top 1st of 2021 graduating class (out of 900 students))

EXPERIENCE

Graduate / Teaching Assistant Jan 2022 - Present
MBZUAI *Abu-Dhabi, United Arab Emirates*

- Creating mathematical problems and lab materials, guiding students on their projects
- Teaching concepts of probability theory, statistics, and evaluating the goodness of estimators
- Big Data and Parallel Computing, ML in Spark, Data Mining, Link analysis (PageRank) and LSH
- More information on this website

Data Scientist May 2020 - Sep 2021
InCyberService (Healthcare) *Almaty, Kazakhstan*

- Created a model that can classify highly imbalanced-class data accurately
- Utilized explainability methods to evaluate the obtained outcomes of a classification model
- Developed object detection models to extract meaningful information from paper-based documents
- Created a new pipeline: controlling data processes using Airflow, Apache Spark (PySpark), Hadoop

Python Developer and Teacher Jan 2018 - Mar 2019
Bolashak School, IITU *Almaty, Kazakhstan*

- Developed a school management system to control internal processes using Python/Django
- Trained students Python and C++, including OOP, functional programming, and web parsing

SKILLS

Languages	Python, C++, Scala
Frameworks	PyTorch, Tensorflow, ML and CV frameworks/libraries, Spark / PySpark
Experienced areas	Federated Learning, Privacy-Preserving ML, Safety and Robustness of AI, Fairness, Anomaly Detection, Causal Learning
Other skills	Big Data, Apache Spark, Airflow, Hadoop, Spark Streaming, Apache Kafka, SQL DBMSs, Postgres, Cassandra, Clickhouse, Docker

PUBLICATIONS

CYCle: Choosing Your Collaborators Wisely to Enhance Collaborative Fairness in Private Decentralized Learning.
***Nurbek Tastan**, Samuel Horvath, Karthik Nandakumar.* NeurIPS, 2024. *Under Review.*

FedPeWS: Personalized Warmup via Subnetworks for Enhanced Heterogeneous Federated Learning.
***Nurbek Tastan**, Samuel Horvath, Martin Takac, Karthik Nandakumar.* NeurIPS, 2024. *Under Review.*

Redefining Contributions: Shapley-Driven Federated Learning.
***Nurbek Tastan**, Samar Fares, Toluwani Aremu, Samuel Horvath, Karthik Nandakumar.* IJCAI, 2024.

Collaborative Learning of Anomalies with Privacy (CLAP) for Unsupervised Video Anomaly Detection: A New Baseline.
*Anas Al-lahham, Muhammad Zaigham Zaheer, **Nubrek Tastan**, Karthik Nandakumar.* CVPR, 2024.

A Coarse-to-Fine Pseudo-Labeling (C2FPL) Framework for Unsupervised Video Anomaly Detection.
*Anas Al-lahham, **Nubrek Tastan**, Muhammad Zaigham Zaheer, Karthik Nandakumar.* WACV, 2024.

CaPriDe Learning: Confidential and Private Decentralized Learning based on Encryption-friendly Distillation Loss.
***Nurbek Tastan**, Karthik Nandakumar.* CVPR, 2023.

Valid and Invalid Bitcoin Transactions.
*Saule Amanzholova, **Nurbek Tastan**, Kamila Kalkamanova, Amina Yessenalina.* ACM ICEMIS, 2020.

Burglary Detection Framework for House Crime Control.
***Nurbek Tastan**, Abdul Razaque, Mohamed Frej, Saule Amanzholova, R. Ganda, F. Amsaad.* IEEE ICCSA, 2019.

PROJECTS

Collaborative Learning with Robustness to Poisoning Attacks. Implemented several poisoning attacks (label flip, random, noise, backdoor) and came up with a defense mechanism against them. Used anomaly detection on the information being shared between participants and eliminated malicious parties. Implemented using PyTorch.

Registration Plate Number Detection. Real-time video processing by extracting frames and detecting vehicle registration plates. Then, fed it into OCR to extract the plate number. Used my own data and Google Dataset. Built an application that opens the barrier and controls cars entering the building. Used Yolo and Tesseract OCR.

Causality Problem: Cyber-Security Attacks on Graph Data. Built an application that helps the cyber security community in analyzing attacks in an efficient way. The input to this problem is security vulnerabilities (e.g., buffer overflow, gateway attacks, etc.) represented as DAGs. Used two well-known methods: greedy equivalence search (GES) (score-based: BIC and BDeu) and Peter Clark (PC) (constraint-based) algorithm.

Density Estimation in Continuous Bayesian Network. Reasoning under uncertainty. Built models to learn parametric probability density functions (pdf) of a continuous Bayesian network. Used linear Gaussian models with an assumption that all pdfs are Gaussian. It outperformed the current methods by showing smaller TVD values.

Weather Forecasting. Implemented an LSTM model that predicts combined conditions of the atmosphere in Almaty. Used Google's Earth Engine to get the historical data.

CERTIFICATES

Machine Learning Specialization by DeepLearning.AI — Coursera, Dec, 2022
- Supervised Machine Learning: Regression and Classification
- Advanced Learning Algorithms
- Unsupervised Learning, Recommenders, Reinforcement Learning

Cybersecurity by IBM — Coursera, Jun, 2020

HONORS & AWARDS

"Hack The Space" Hackathon, Dubai, United Arab Emirates, "The Best Math Team"	2022
Republican scientific competition among students, Kazakhstan, II place (Silver)	2021
Scientific competition among students in IITU, Kazakhstan, II place	2021
Republican scientific competition among students, Kazakhstan, II place (Silver)	2020
Scientific competition among students in IITU, Kazakhstan, III place	2020
Presidential Scholarship, Ministry of Education, Kazakhstan	2020
Republican mathematics competition among bachelor students, III place (Bronze)	2019
ICPC regional, II place	2019
Mathematics competition, IITU, I place	2019
Mathematics competition, IITU, II place	2018
Award "Altyn Belgi" (Golden Badge), Ministry of Education, Kazakhstan	2017
National Presidential Olympiad, Ministry of Education, Kazakhstan (I - regional, II - final)	2016
Republican Applied Mathematics scientific competition, III place	2016
Regional Applied Mathematics scientific competition, I place	2015
Mathematics Olympiad, I place	2013

REFERENCES

Dr. Karthik Nandakumar
Primary Supervisor
Associate Professor at MBZUAI
✉ karthik.nandakumar@mbzuai.ac.ae

Dr. Samuel Horvath
Secondary Supervisor
Assistant Professor at MBZUAI
✉ samuel.horvath@mbzuai.ac.ae