FourthBrain

Machine Learning Engineer

Program Syllabus

Program Outcomes

FourthBrain trains aspiring Machine Learning engineers in the technical and practical skills necessary to contribute immediately to an AI team. At the end of the Machine Learning Engineer program, students will be able to:

- Build, validate and test data models via supervised, unsupervised, and semi-supervised methods
- Apply Deep Learning methods to computer vision, perception and other scenarios
- Utilize Generative Adversarial Networks for data and model scaling.
- Apply time series models for contextual learning and inference
- Benchmark, optimally parameterize and modularize system frameworks for extendability
- Build explainable models by leveraging visualization techniques
- Deploy scalable, optimized end-to-end models in production environments
- Implement Continuous Integration and Deployment (CI/CD) Pipelines for continuous deployed applications.
- Integrate themselves into industrial teams as key contributors for production-ready environments.
- Run Deployed ML demonstrations of group projects with polished Github repositories

Our program emphasizes more than just technical skills. In addition to the outcomes listed above, our students will also be able to:

- View ML and AI projects through an ethical lens, with an awareness of common issues and knowledge of how to access further resources
- Learn effective communication skills to contribute as part of a team.
- Approach their career goals with the skills and knowledge of how to apply their newly gained skills in their chosen field

Program Outline

The Machine Learning Engineer program is divided into two parts.

Part 1: Individual learning and concept application

You'll build skills in core ML concepts, and complete several individual projects. Each week, you'll complete standalone mini-projects. Some projects can be combined to create larger, more comprehensive projects.

Individual Project Topics:

1. Predictive Data Analytics (Weeks 1-4)

   Given: Historic data regarding customer online purchase behavior.

   Goal: Predict quarterly sales and customer purchases, analyze categories of customer behavior, and gauge customer response for new products or automated product categorization.

2. Deep Learning and Computer Vision (Weeks 5-7)

   Given: Images and videos annotated for object detection for cars, pedestrians, bicyclists, lane markers, street signs.

   Goal: Design optimized end-to-end and explainable models for a) object detection, b) tracking, and c) semantic segmentation.

3. Generative Adversarial Networks (Week 9)

   Given: A finite image/video annotated data for semantic segmentation.

   Goal: Automatically increase training image data volumes and training image dimensionality, to reuse annotations and enable robust object detection models.

4. Time Series Modeling (Recurrent Neural Networks) (Week 10)

   Given: Volumes of web analytics data, with session-based information regarding consumer purchase behavior for viewing, adding and removing products from the cart, and purchasing.

   Goal: Create a stochastic prediction of customer purchase probability per session that will enable the creation of marketing nudge models.

Part 2: Peer Collaboration and Concept Scaling

In Part 2, you will build a capstone project in a group of 2-3 students, and also gain experience with scaling and deploying your models.

You'll develop a project in a group of two or three students. Your task is to develop a scalable final solution that works on multiple data sets, is optimized and benchmarked for performance (run time and accuracy) and is deployable.

> ***Sample Team Project:*** Design a Recommendation Engine for consumers based on past purchases and user preferences. Possible use cases include showing the right product thumbnail to enhance impressions.
>
> ***Sample Team Project****:* Design a Digital Twin model that will find the closest match to the consumer interests using the configured user choice. Possible use cases include finding street view maps for satellite imagery, similar homes for a realtor business, or finding matching facial traits.

Ethical and Responsible AI

Responsible and ethical AI is a rapidly growing field, and addressing all areas of this field is outside the scope of this course. We address ethical concerns throughout the course, with a focus on two topics: bias in data and how to mitigate unintended outcomes, and data privacy compliance with statutes like GDPR/CCPA. We encourage students to continue learning and researching outside these specific areas.

> **Bias in data:** This is a complicated issue with no easy answers. You will gain an awareness of common issues, what questions to ask, and understand the limitations of certain datasets.
>
> **Data privacy and security**: Data privacy and the security of systems is integral to maintaining trust with your users or the public. You will gain an awareness of some common errors that compromise privacy and how to prevent them.

Communication

Communication with technical and non-technical colleagues is a crucial skill for all engineers. We emphasize the importance of regular verbal and written communication throughout the program.

In Part 1, you'll regularly collaborate with peers in breakout sessions to reinforce engineering team settings. This emphasizes the importance of speaking up, troubleshooting together, and being open to other's thoughts and approaches, especially with peers you may be working with for the first time.

In Part 2, you'll collaborate for an extended time with your project team. Your group is expected to plan the work, divide tasks, and provide each other regular updates. Additionally, you'll provide updates to the entire cohort on your project, in both written and verbal form, and receive feedback from your classmates accordingly. Giving and receiving direct and constructive feedback while maintaining personal and collegial relationships will receive specific attention. A verbal presentation during a demonstration day is part of the final project.

Career Growth

The Machine Learning Engineer program is designed for you to acquire the skills and knowledge required for a Machine learning Engineer. Whatever your reason for taking the program - to get a new role at a new company, to gain skills for your current job, or just for fun - we will support your career growth by helping you connect to professionals and employers, via guest speaking events and inviting employers to the final project presentation day.

The following is a list of sample roles that candidates who successfully complete the Machine Learning Engineer program are expected to qualify for:

- ML Engineer
- AI Engineer
- Deep Learning Engineer
- Cloud Architect
- Solutions Engineer
- Computer Vision Engineer
- Autonomous Driving
- Data Scientist
- AI Product Manager
- Machine Learning Research Scientist
- AI Research Scientist
- Applied AI Scientist
- Applied ML Scientist

Career services assistance is available after graduation to help ensure that all candidates achieve their career goals.

Weekly Schedule

January 2022

Week	Topics	Project
Week 0 / Orientation 1/29 (Sat) 2/1 (Tues)	• Intro to AI, Machine Learning, Data Science, Cognitive Systems. • Weekly assignments, optional courses • Review Syllabus, • Tools, Deliverables, Projects • Expectation Management (Individual + team projects) • Introduction to polynomial Regression	
Week 1 2/5 (Sat) 2/8 (Tues)	• Supervised Learning for Regression Applications • Feature Engineering (one-hot encoding) • Generalized Linear Models (GLM) • Decision Trees, Classification and Regression Trees	Given: Historic data regarding online sales of products for Q1 and 2. Data: Highly pre-processed sales data. Goal: To predict sales for Q3 and Q4.
Week 2 2/12 (Sat) 2/15 (Tues)	• Supervised Learning for Classification Applications • Logistic Regression • Support Vector Machines • Random Forest • XGBoost, Bagging, Boosting • Data Imbalance Problem	Given: Historic data regarding customer online shopping journeys over Month 1 and Month 2. Data: Publicly available raw google analytics data that is somewhat pre-processed. Goal: To predict online customer conversion from viewing to purchasing.

Week 3 2/19 (Sat) 2/22 (Tues)	• Unsupervised Learning • Clustering (k-means) • Dimensionality Reduction • PCA, t-SNE • DBSCAN • Data Visualization using *Tensorboard*	Given: Historic data regarding customer online shopping journeys over Month 1 and Month 2. Data: Publicly available raw google analytics data that is slightly pre-processed. Goal: To analyze categories of customer behavior based on server data only.
Week 4 2/26 (Sat) 3/1 (Tues)	• Semi-supervised Learning • Transductive SVM (label propagation) • Co-training algorithms • Zero-shot learning • Label Propagation/Label spreading • AutoML Libraries (TPOT, autosklearn)	Given: Historic data regarding customer online shopping journeys over Month 1 and Month 2. Data: Publicly available raw google analytics data (no-pre processing). Goal 1: If a new cosmetic or electronic product is to be launched next month, predict customer response. Goal 2: Metadata tagging. To group products as “new favourites”, “fall special”, “winter collection” etc. Introduction to ML System Design.
Week 5 3/5 (Sat) 3/8 (Tues)	• Introduction to deep Learning for Images and Video • Convolutional Neural Network (CNN) • Encoder-Decoder Networks • Recurrent Neural Networks (RNN) • Generative Adversarial Networks (GANs)	Given: A large data set of medical images and their labels for pathology severity. Goal: To train a classifier that predicts pathology severity that will enable automated patient prioritization.

Week 6 3/12 (Sat) 3/15 (Tues)	• Object Detectors (CNN) • Methods and Loss functions for 2D detectors (One shot, two shot detectors) • Models: SSD, YOLO, Fast RCNN, FasterRCNN, MaskedRCNN	Given:Images with objects to be detected and explained. Goal: To design end-to-end and explainable models for 2D object detection Project Pitching Day (Think Tank Forum) Project Feedback (written) and Capstone Project Team formation.
Week 7 3/19 (Sat) 3/22 (Tues)	• Semantic segmentation (encoder-decoder) • U-net • Squeezenet • Depth-wise separable filter design for parameter optimization	Given: Annotated images for semantic segmentation on Medical Images. Goal: To design end-to-end and explainable models for semantic segmentation of pathology using a minimal set of training images.
Week 8 3/26 (Sat) 3/29 (Tues)	Midterm Project and Capstone Project proposals	Project Teams submit Proposals.
Week 9 4/2 (Sat)	• Introduction to Generative Adversarial Networks (GAN) • Applications for Data Augmentation (cGANS) • Paired model (pix2pix) • Unpaired model (CycleGAN)	Given: images and annotations for image segmentation and specific domain set up. Goal 1: To generate an augmented image stack to transfer images to

4/5 (Tues)		another domain (day to night conversion). Goal 2: To increase dimensionality of data (black and whiite to color images).
Week 10 4/9 (Sat) 4/12 (Tues)	• Time Series Models • Introduction to Recurrent Neural Networks (RNN) • Seq2seq Models (applications to Natural Language Processing) • Simple RNN, GRU, Long Short Term Memory (LSTM) • Introduction to ML Model deployment using Flask	Given: Historic data regarding customer online shopping journeys over Month 1 and Month 2. Data: Publicly available raw google analytics data (no-pre processing). Goal: To predict customer conversion probability per session with the intention of creating nudge models.
4/16	Break	
Week 11 4/23 (Sat) 4/26 (Tues)	• Advanced ML and Deployment • Introduction to Tensorflow(TF) Lite • TF Lite for Model conversion, Optimization and Quantization • TF-Lite Transfer Learning • TF Lite Model Deployment on AWS, Google Cloud • ML Deployment using FastApi	Team Project: Deploy an ML webapp on Elastic Beanstalk in AWS
Week 12 4/30 (Sat) 5/3	• Big Data Processing for ML • AWS Demo for Ec2 Instances and Sagemaker • Transfer Learning for ML • Big Data Tools (Hadoop, Mapreduce)	Individual Project: Implement Dask and Pyspark frameworks for classification tasks. Midpoint Capstone Project Review (Verbal and Written Presentations)

(Tues)	• 2Big Data Processing (PySpark, Databricks) • Multi-processing (Dask, joblib)	
Week 13 5/7 (Sat) 5/10 (Tues)	• Cloud2-based AutoML and Data Visualization • AutoML with Google Cloud • AutoML with Azure (classification, regression, time series models) • AutoML on AWS • Data visualization using LIME and SHAP	Individual Project 1: Data visualization LIME vs SHAP Individual Project 2: Deploy an AutoML trained model on Google AI platform and run inference on it. Midpoint Capstone Project Review (Verbal and Written Presentations)
Week 14 5/14 (Sat) 5/17 (Tues)	• MLOps and Special Topics • ML and DevOps on Azure • MLOps on Google Cloud • Kubeflow and AWS • MLflow and LoudML • SSH Tunneling with Ngrok • CI/CD pipelines and triggers for continuous deployment • Tools, Platforms and Systems (cont'd) and Digital portfolio creation	Individual Project: Implement a Continuous Delivery trigger for cloud runs for an ML webapp on Google Cloud.
Week 15 5/21 (Sat) 5/24 (Tues)	• ML Pipelines and Final project presentations • Kubeflow exercises	Graduation and Project Presentations will occur during the last week of class.

Platforms and Tools

- Python, Tensorflow, Keras, pandas, scikit-learn, Pillow, OpenCV, Matplotlib
- Tensorboard for modeling and visualization
- ML Model optimization and deployment using TensorflowLite
- Hadoop, Mapreduce, Databricks
- Dask, joblib, Pyspark (MlLib)
- AutoML on AWS, Google Cloud, Azure, Libraries (TPOT, Autosklearn)
- ML deployment RestAPIs: Flask, FastApi
- Detailed learning of ML deployment friendly libraries: TFLite and TFlite serving using Flask and FastApi
- MLOps, CI/CD Pipelines on Microsoft Azure, Google cloud, AWS
- ML model containerizing (Docker, Kubernetes)
- Pipeline Solutions: Kubeflow, MLflow, LoudML
- Introduction to to AWS Sagemaker and Ec2 instances
- ML model and app deployment on AWS (Elastic Beanstalk) and Google Cloud
- SSH tunneling and webapp deployment using “ngrok”