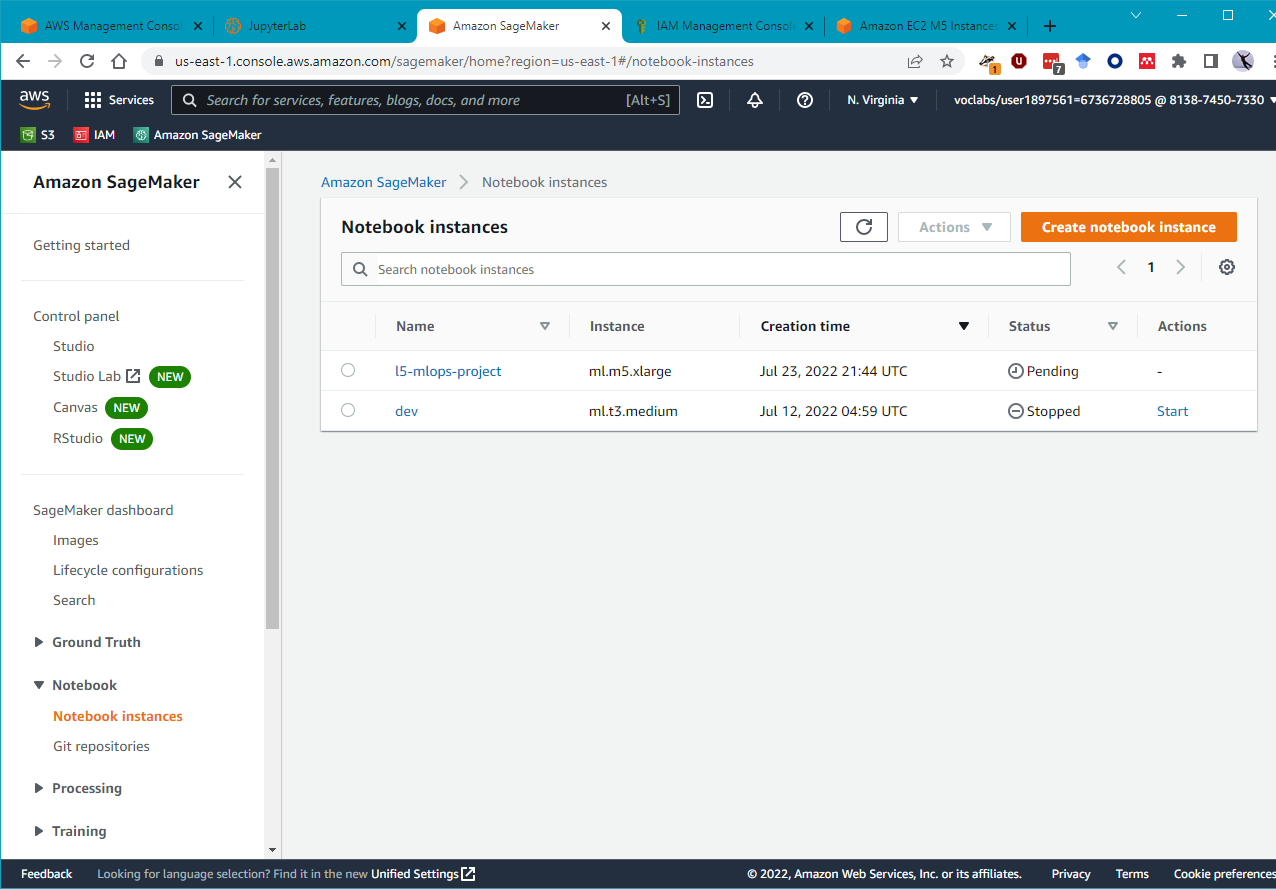
# Initial setup, training and deployment

Computer vision workload and HPO (Hyperparameter optimization) requires a larger instance to run.

* Chosen Instance Type: ml.m5.xlarge with JupyterLab platform

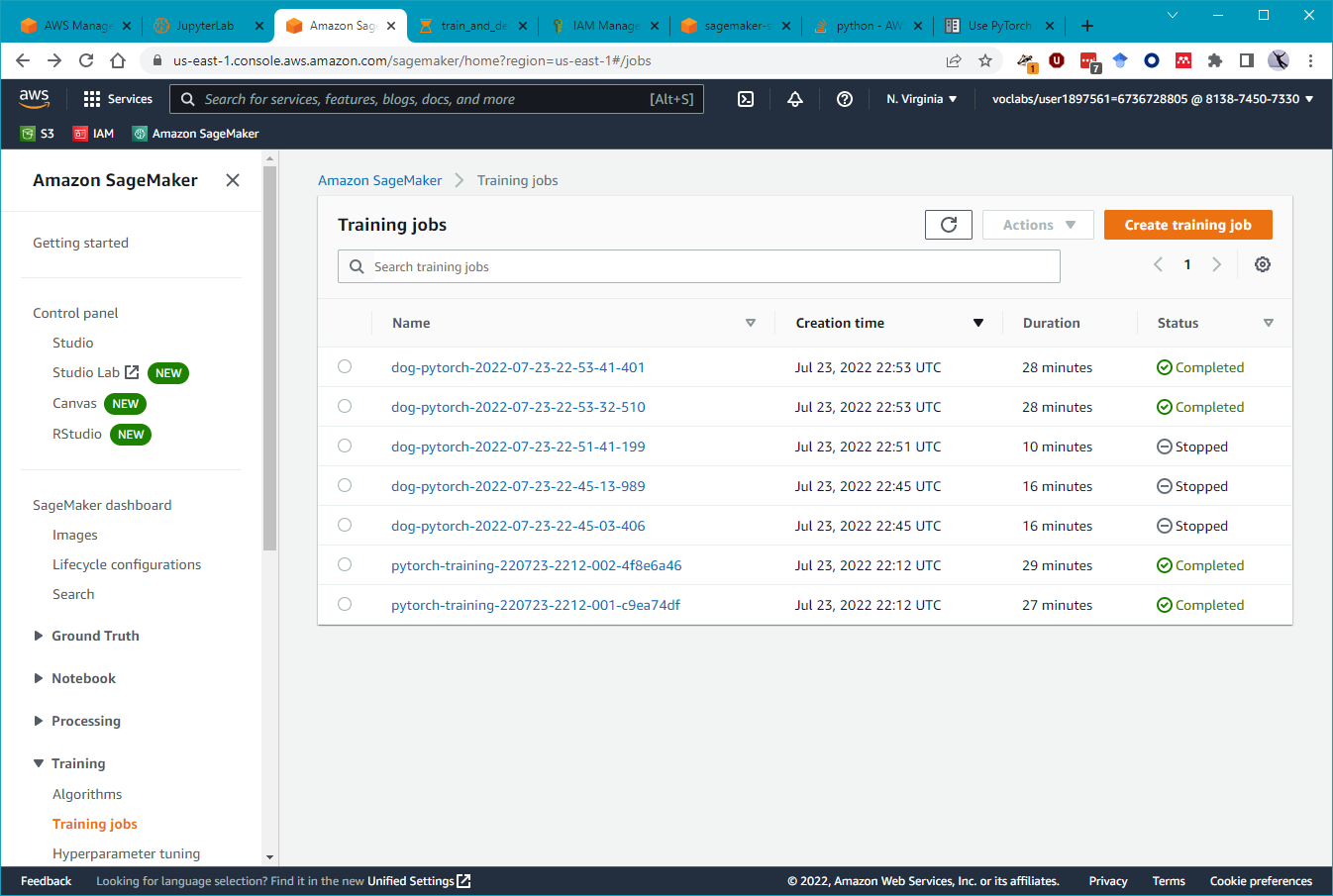
A ml.m5.large instance type is for general purpose compute workload with 4 vCPUs, 16GB RAM and high network bandwidth.

Even though there is no GPU on these instance, M5 instances provide support for the Intel Advanced Vector Extensions 512 (AVX-512) instruction set, offering up to 2x the FLOPS per core compared to the previous generation M4 instances.



* Download data to “S3://sagemaker-studio-qm1g8vhuin/dataset”

Training and Deployment (Takes more than 20 minutes)



Deployment Endpoint

