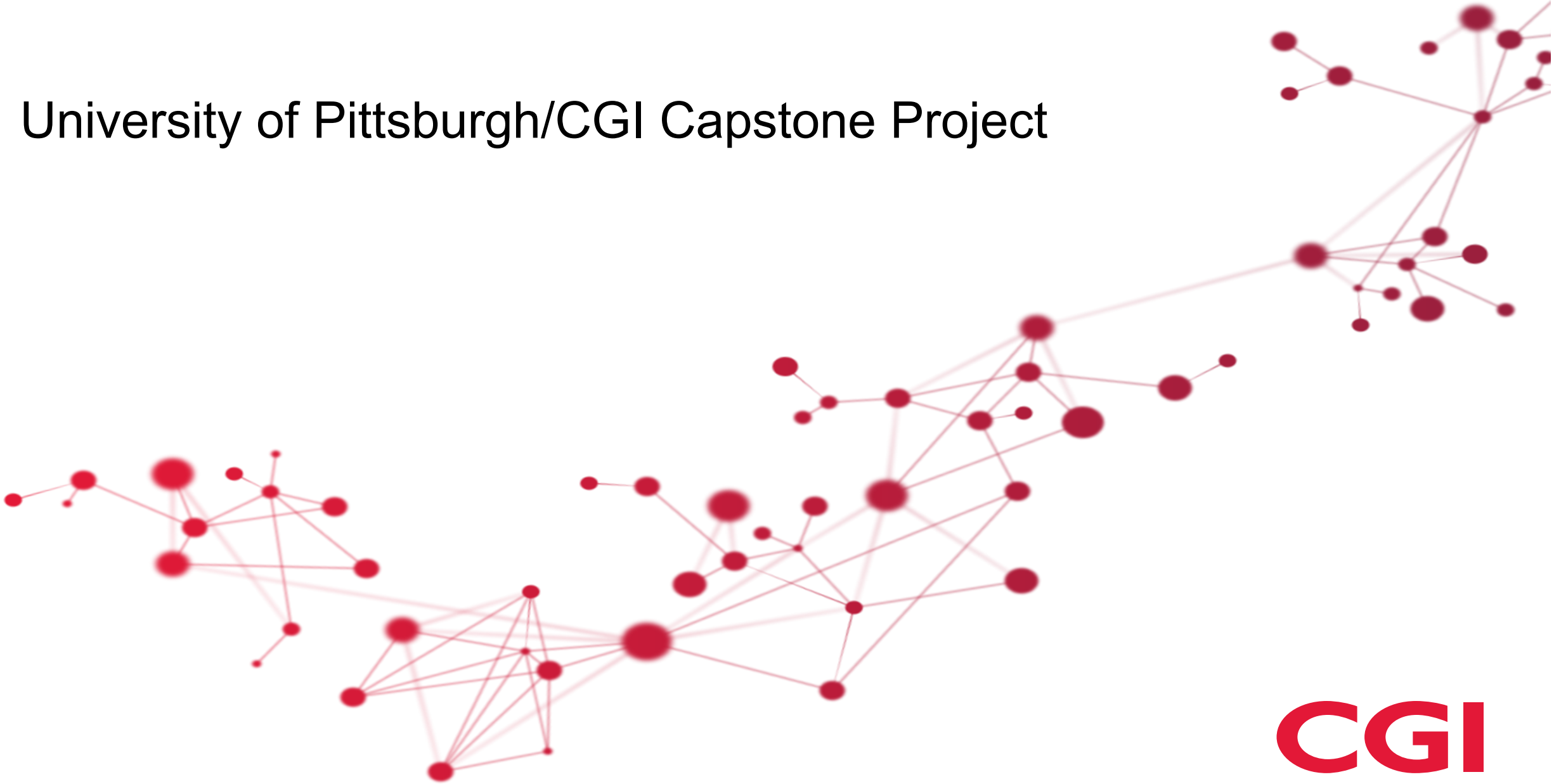
University of Pittsburgh/CGI Capstone Project



CGI

Experience the commitment®

Agenda

- Introductions
- Project Overview
- Project Plan
- Technology
- Questions

Introductions

Project Lead:

- Surya Patchipala

Technology Leads:

- Sandeep Pamarthi
- Babul Jha

Product Sponsor

- Tony Deluca

Campus Recruiter

- Heather Fusko

CGI Sponsor

- Bernie Mongilio

Project Overview

Use Case:

Our goal is to categorize Banking Ticketing support text information automatically which is extracted from historical support tickets and matching a given category with the correct recipient using Distilled BERT Text Classification.

For our implementation, we choose the BERT model, due to its popularity, performance and availability of open-source implementations.

Scope:

- Scope of this exercise is to train the Distil-Bert Model using 3 billion words on a single GPU.
- Save the model output and deploy the model as REST end point with curated environment.
- Call the end point for any future prediction.

Project Plan

Spring Semester 2022

- Understand Text Classification Algorithms
- Implement few Text Classification Algorithms in local CPU
- Analyze the accuracy and performance metrics
- Train the high accuracy Text Classification model in Cloud
- Crete pipelines for the end-end deployment using TensorFlow and GPU's.
- Create end points on the Text Classification model
- Create explainability algorithm on top the model
- Automate/End point creation for the explainability algorithm

Technology

For Development

- Python
- Cloud Technology (Azure)
- Basic Machine Learning
- Explainability Algorithms
- Git
- SQL Database (Any cloud persistence DB)
- IntelliJ
- Trello

Azure Services

- Databricks
- Block Blob Storage
- Pytorch curated environment for GPU utilization
- Azure Endpoints, Models, Pipelines

For Communication

- Microsoft Teams

Questions?

- Comments?
- Notes:
  - There will be a NDA for our project.

Thank you!