# AI Across Attain

The purpose of this document is to provide a high-level overview of how your project is incorporating AI/ML/Data Science / RPA, or places where you see these areas potentially fitting into your project. Please feel free to use the following template when filling out the document:

**Name**: Jennetta George

**Project/ Branch**: Edgar / SEC

**AI/ML/DS/RPA Use Case (existing / potential?)**: We are currently building out a PoC to serve as an entity matching service for applicants coming into the edgar system against existing applicants in Hub and Watchlist. The goal is to produce a model which, when given an applicants name and some other biographic data, to return a potential match in the Hub system along with a confidence score.

**Name:** Jeremy Harris

**Project / Branch:** Watt (ML Computing System) **/** Department of Energy

**AI/ML/DS/ RPA Use Case (existing / potential?):** The National Energy Transportation Laboratory (NETL) at the Department of Energy (DOE) has a very powerful AI/ML computing system. My goal with this system is to set it up so that users of various technical skills can utilize the system to take their research and develop meaningful results much faster than they currently are about to do. In addition, I am taking projects that individual scientists have and helping them utilize AI/ML to complete their work. Currently, I’m working on Image Segmentation using Neural Networks with U-net using Keras and Tensorflow in Python.

**Name: Braden Catlett**

**Project / Branch: CoupleCommunication. Defense Health Agency Web and Mobile Technology**

**AI/ML/DS/ RPA Use Case (existing / potential?): DHA WMT has tasked us with building an application that can analyze text messages on a user’s device to determine if the language used is defensive or critical and then suggest other phrases or wording to soften the text message. This will hopefully improve communication between couples. I don’t have many details beyond that since I was taken off this project back in March to work on a high priority Covid-19 project. The idea will be to train a NLP model on a set of defensive and critical statements we have procured through using Amazon mTurk.**

**Name: Jason May / Joe Webb**

**Project / Branch: USCIS Biometrics**

**AI/ML/DS/ RPA Use Case (existing / potential?):** More DS than AI/ML. Leveraging Databricks and Pyspark, we created summary statistics and visualizations on almost 20,000 columns across 5 systems and around 1000 tables. Several of these tables had over 1 billion rows and most had over 10 million, so we were dealing with massive amounts of data. Yet, the automation and parallelization enabled by Pyspark/Databricks made this doable in a matter of hours.

Visualizing the massive amounts of disparate tables allowed us to compare columns across different tables and assess their similarity (ie how does the distribution of the length of first names differ across systems and tables).

**Name:**

**Project / Branch:**

**AI/ML/DS/ RPA Use Case (existing / potential?):**

**Name:**

**Project / Branch:**

**AI/ML/DS/ RPA Use Case (existing / potential?):**

**Name:**

**Project / Branch:**

**AI/ML/DS/ RPA Use Case (existing / potential?):**

**Name:**

**Project / Branch:**

**AI/ML/DS/ RPA Use Case (existing / potential?):**

**Name:**

**Project / Branch:**

**AI/ML/DS/ RPA Use Case (existing / potential?):**