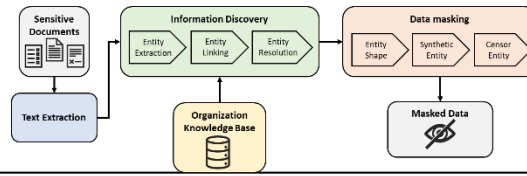# AI-Powered Sensitive Data Identification & Masking

Project Summary:

This research will build upon previous DOD Natural Language Processing (NLP) efforts to demonstrate techniques for efficiently identifying and subsequently cleansing sensitive information from operational data within structured and unstructured data sets. It will focus on first identifying sensitive data and how it relates across a data set, and then providing options to either censor or mask sensitive data with synthetic data. USTRANSCOM's global mission creates the need to engage with a diverse set of commercial and foreign partners. The primary use case for

Visual Representation of the Unstructured Data

this effort would be to facilitate a safe and agile information pipeline with trusted commercial and foreign partners. It is imperative as the need to share operational information grows, that methods to ensure this is done safely are developed.

Benefit:

USTRANSCOM's global mission scope requires safe and agile information across trusted commercial and foreign partners. This research would help facilitate, ensuring that those information pipelines improve and meet the increasing demand for data rich collaboration while maintaining operational security. The capability to identify and employ protective measures on sensitive data protects one of the command's most valuable assets, its operational data. The need to facilitate information sharing across trusted partners and supply emerging technical solutions with the training data they need will continue to grow with operational data requirements. It is imperative that mechanisms for doing this safely develop with those needs.

Duration of project: FY24 – FY25

Participants: PEO-T, all PORs, DevSecOps, AI/ML Dev Efforts

Project advocacy (funding or otherwise): TCJ-6