DATA+AI SUMMIT
BY databricks

AUTOMATING GOVERNMENT TRANSPARENCY IN DECLASSIFICATION AND FOIA TASKS

Samuel Stehle
June 13, 2024

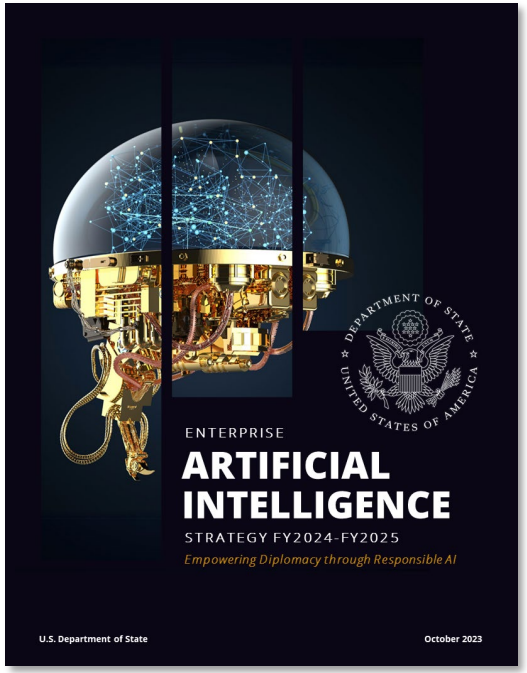
CENTER FOR ANALYTICS

Who we are, our mission

- The Department of State's enterprise data management and analytics capability
- Led by the Chief Data and AI Officer
- Transform data into insights to make better management and foreign policy decisions
- Expand data access and analytic expertise across the Department through our data and tech platform; Data.State

ENTERPRISE AI STRATEGY

- DoS will responsibly and securely harness the full capabilities of trustworthy artificial intelligence to advance United States diplomacy and shape the future of statecraft
- Guided by the CfA, the EAIS is the product of DoS's AI leaders and policy experts from over 25 bureaus and offices across the enterprise
    1. Leverage Secure AI Infrastructure
    2. Foster a Culture that Embraces AI Tech
    3. Ensure AI is Applied Responsibly
    4. Innovate



EXECUTIVE ORDER 13526

Classification and Declassification

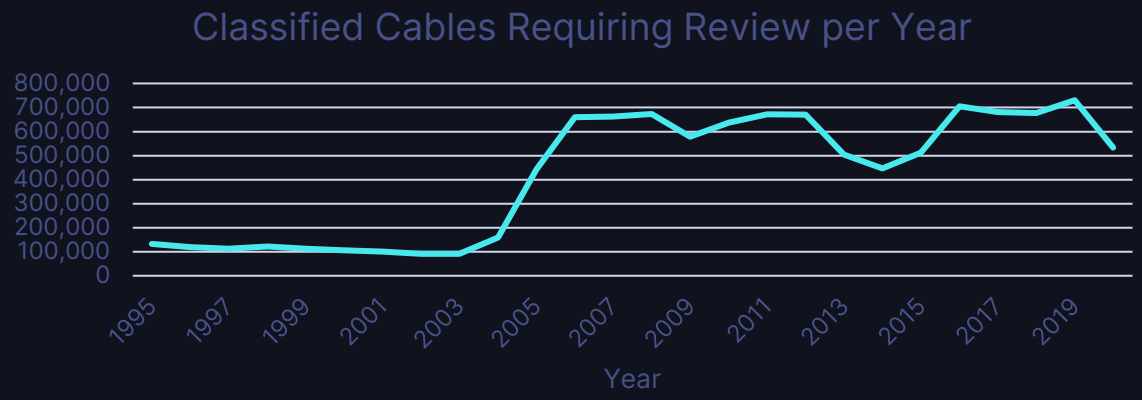
All classified records will be automatically declassified on the final day of the last year of the period of protection,

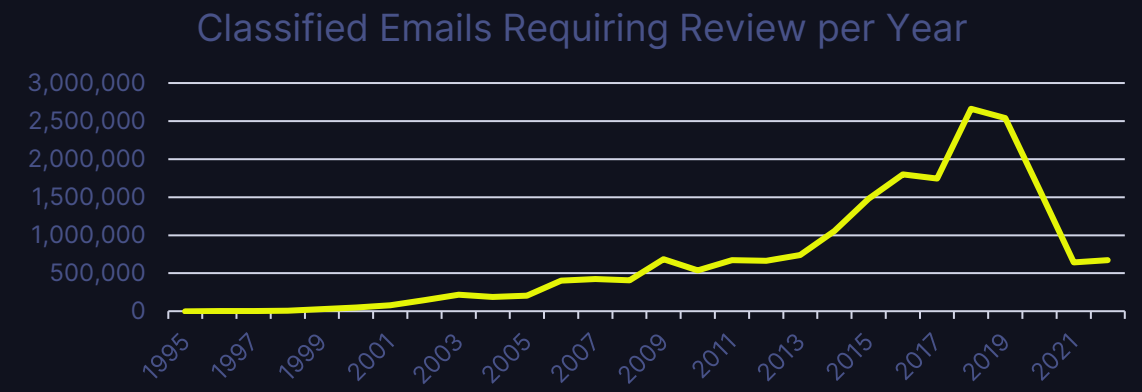
unless an explicit reason to *exempt* that record from declassification is provided.

Establishes a review process to find exemptions.

THE IMMINENT CHALLENGE

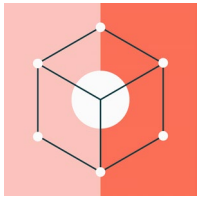
- Cables are the authoritative reporting by U.S. diplomatic and consular posts overseas
- The volume of cables requiring review will render manual review unsustainable
- Inability to review cables by year end poses a national security risk to the Department
- Transfer of records from State to NARA takes additional time which delays public access.





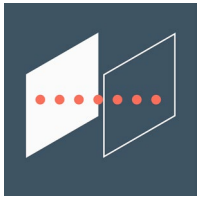
PROJECT APPROACH

1) Pilot 2) Use previous decisions 3) Quality control



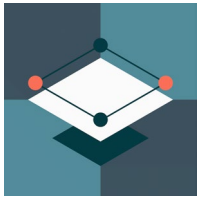
Start small, limit scope

- Chose one electronic record type: cables
- Cables are uniformly structured, readily available in eRecords
- Used 1995-1997 cables, already human reviewed (labeled data)



Train models on human decisions

- Trained ML models on past decisions by human reviewers (whether to “declassify” or “exempt from declassification”

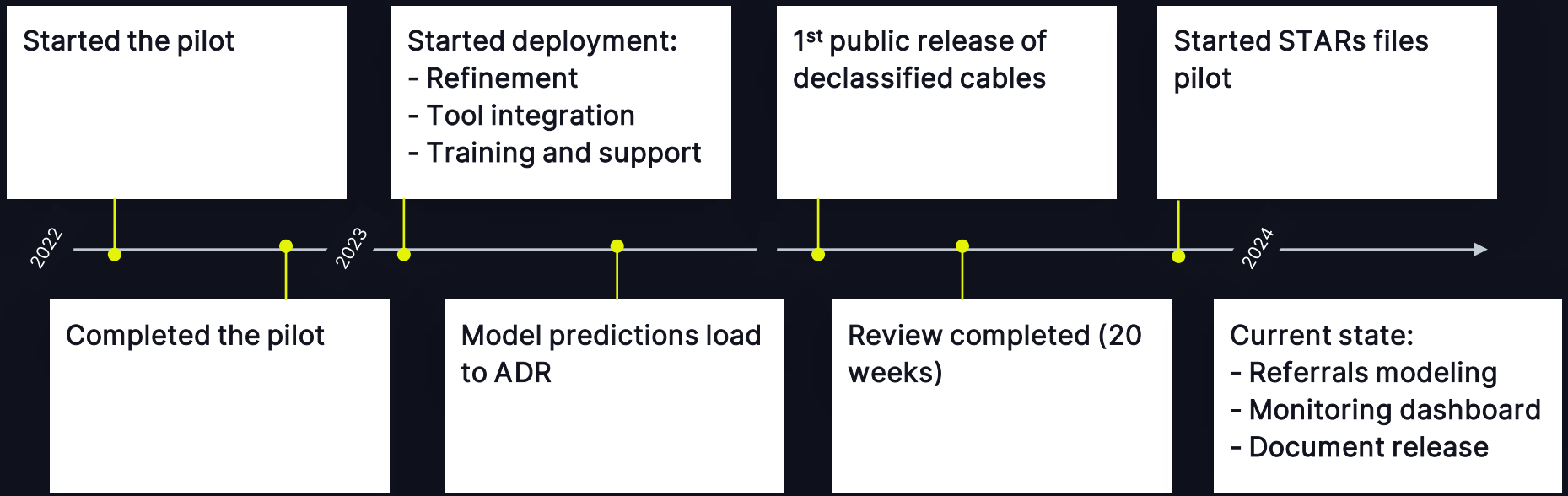


Retain human review by design

- Review/label training data as necessary
- Perform Quality Control (QC) checks
- Review cables the model is unsure of
- Pick up on topic drift over time

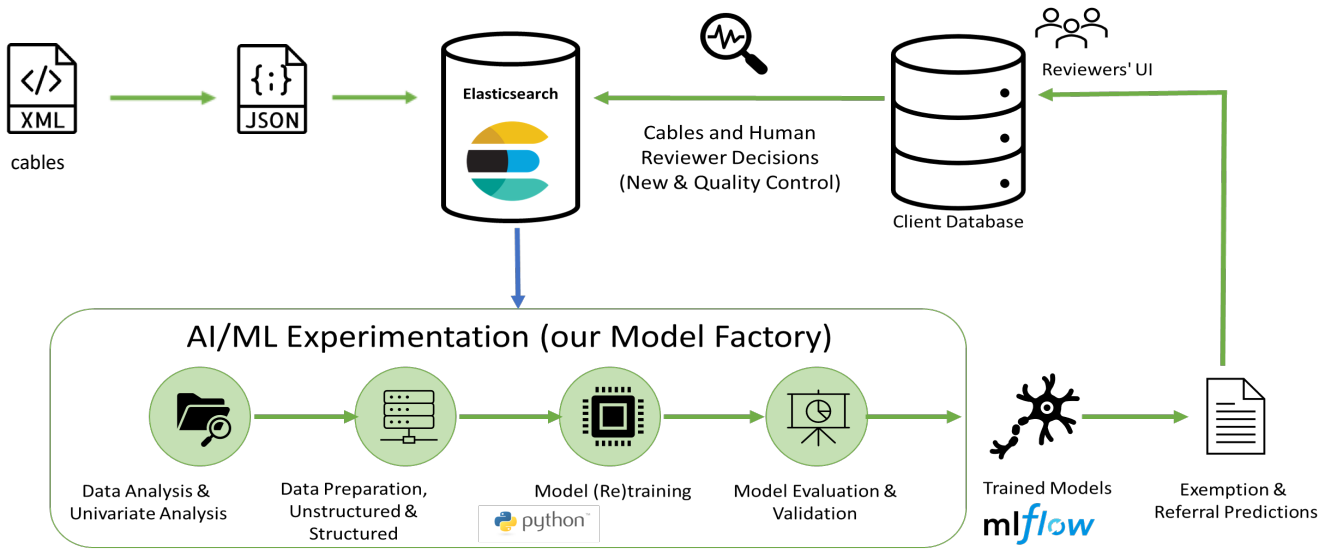
PROJECT APPROACH

Deployment timeline



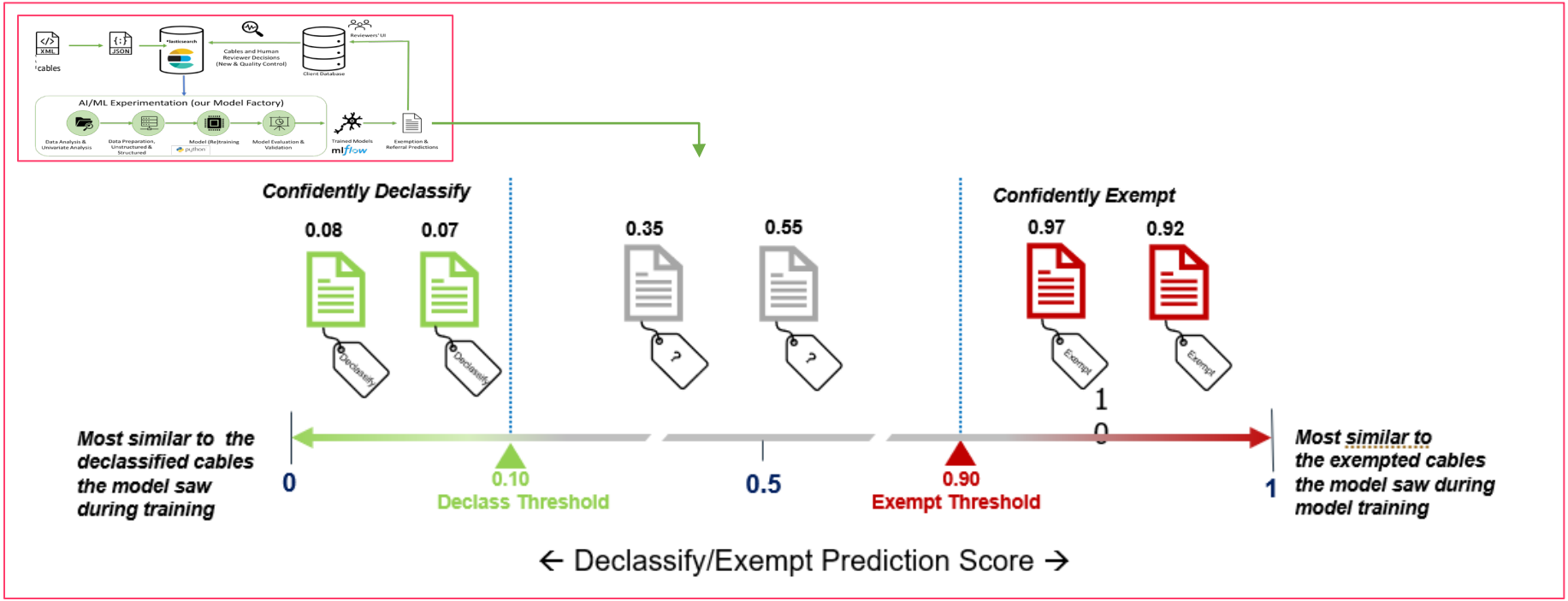
DATA ARCHITECTURE

Open source by necessity



DOUBLE DECISION THRESHHOLDING

Human in the loop
Risk tolerance vs F1 score



TECH STACK

Not one single model

Data Storage, Exploration

- ElasticSearch
- Python
- Kibana



Data Analysis, Modeling

- Spyder, Jupyter Notebook
- scikit-learn, mlflow, imbalanced-learn, xgboost, gensim, small-text
- nltk, regex, spacy, sBERT
- shap, matplotlib



LLMs – bonus!

- Mpnet, bart

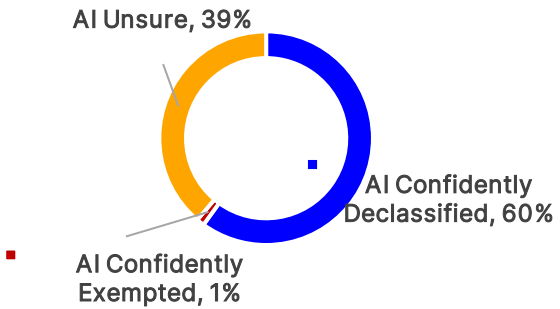

DEPLOYMENT - REVIEW OF 1998 CABLES

Human in the loop in action

1. INITIAL MODEL PREDICTIONS

Thresholds were determined a test **error rate of 1%**

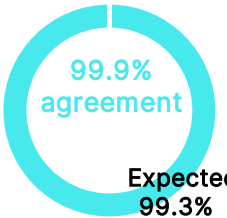
Model made confident decisions on **61%** of cables



2. FIRST HUMAN QC CHECK

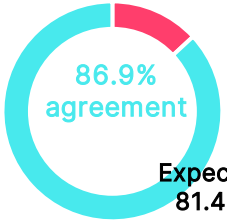
Power analysis suggested **2.5% sample** of predicted cables for human QC

AI-Declassified Cables Human QC Check



Expected agreement: 99.3%

AI-Exempted Cable Human QC Check Result



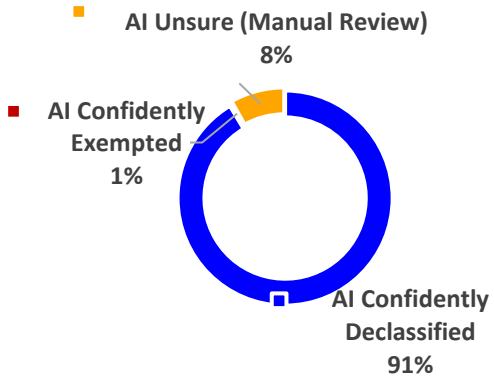
Expected agreement: 81.4%

DEPLOYMENT - REVIEW OF 1998 CABLES

Human in the loop in action

3. RETRAINING, RE-PREDICTIONS

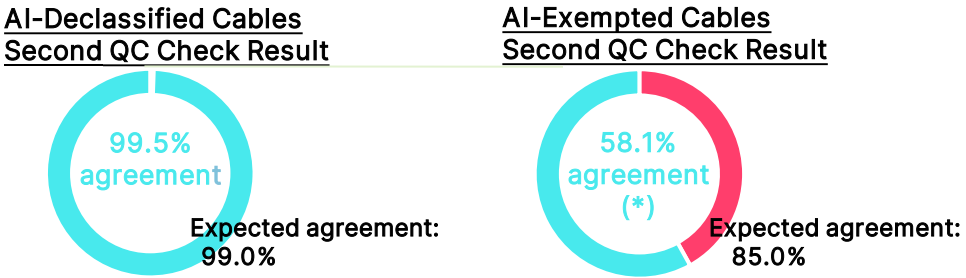
Manual review vastly reduced the remaining Unsure cables



4. SECOND HUMAN QC CHECK

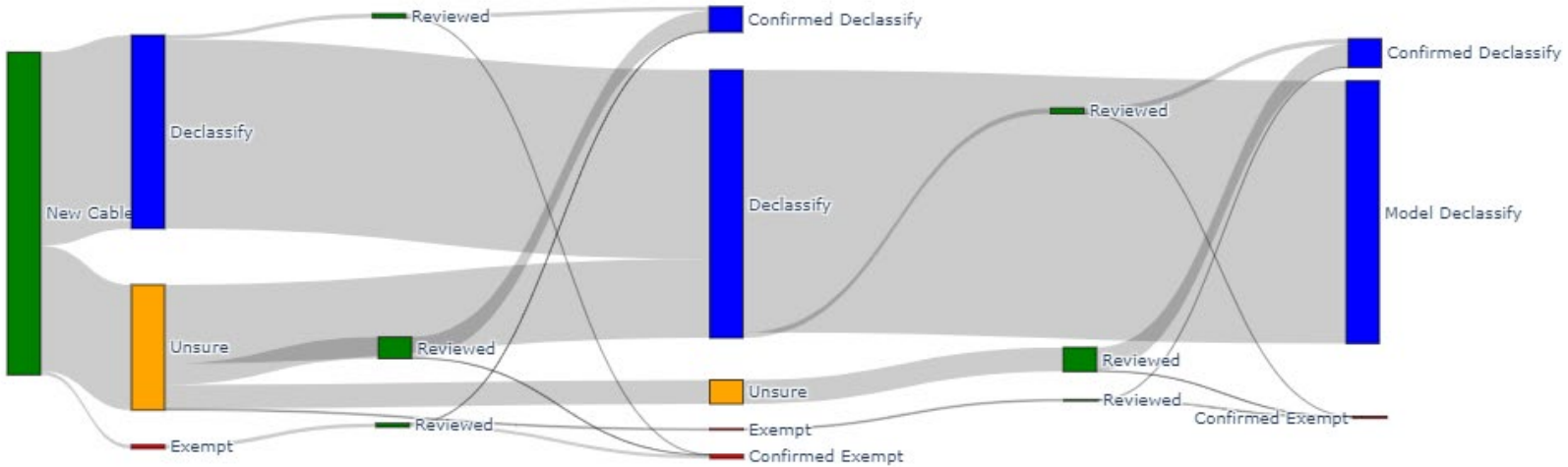
Required human review on only **20%** of cables

Misinterpretation of declassification guidance was rectified during review



DEPLOYMENT - REVIEW OF 1998 CABLES

	Declassified for DOS Equities 98%	Exempted 2%
Decided by human reviewers	18%	2%
Accepted AI predictions	80% 95% confidence of accuracy is 99% +/- 1%	0%



IMPACT OF AI-AUGMENTED REVIEW

Reduced human review volume (80%)

- Focus on other initiatives
- Lower risk of unintended disclosure

Adaptable data-agnostic approach

- Expand to other file types

Repeatable process

- More consistent decisions
- Enhanced policies and training

PROACTIVE DISCLOSURE

Click HERE for access to proactively disclosed cables or visit FOIA.State.gov and search case number:
S-2023-00002

Additional cables will be released regularly

FOIA

Freedom of Information Act

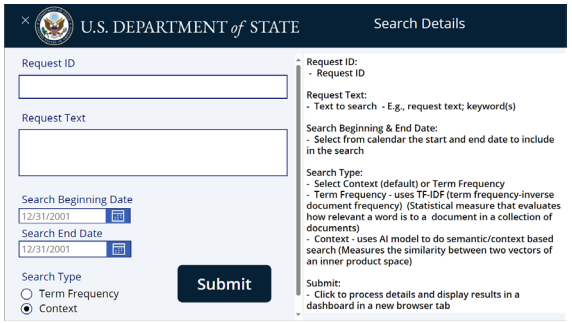
Customer Experience

Search Adequacy

Minimize Redundancy

AI-ASSISTED FOIA REVIEW

Request Input



PowerApps

User Input

(1) New Request submission creates list item which triggers Databricks notebook

List

Data Processing

NER

Sentence Vectors

Matching

SBERT.net

Data

Elasticsearch + Databricks

Microsoft Azure

Reading Room

FOIAXpress

(2) Databricks gets request and executes backend data processing/matching w/ES

Visualization

Kibana

(3) Results are displayed and filters in Kibana Dashboard for analysis of similar docs

UNIQUE CHALLENGES

Data Drift

- Ever-changing political & historical themes deteriorate model performance
- Left unchecked, the model will remain rigid & lose predictive power

- Frequent model retraining, tuning and human quality control to combat historical data drift

Imbalanced Data

- Few exemptions (<5%) compared to declassifications
- Most cables don't get referred to given agency

- Over/under sampling, outlier models to identify outliers

Model Explainability

- Model outputs decisions & confidence with little explanation
- Model makes pass/fail decision on entire cables, not subsections

- More detailed feature importance plus analysis of cable sections improve buy-in & transparency

LESSONS ON AI IMPLEMENTATION

At the State Department

Data Management

Data quality and **accessibility are critical to any AI effort;** the Department of State eRecords platform empowered the team to train, test, and deploy with high quality data

Continuously improve tools and processes with AI/ML features

Start Small

Start small, with a pilot. The experimental approach with well defined **performance metrics** helped the team measure improvement with each iteration.

One **well-suited document type** with manageable volume has helped the team **scale their approach**

Process Transformation

Consider early how AI/ML **transformation** will be incorporated into **process and tools**

Explore concurrent improvements to process and tools, even if not related to ML

Sam Stehle
stehlesk@state.gov