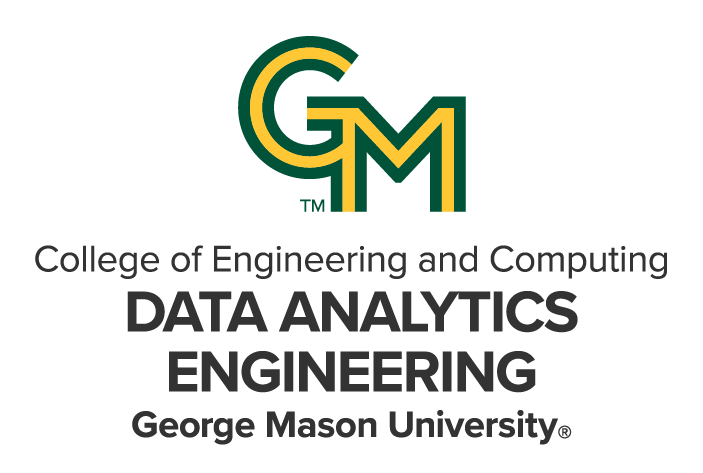
A black rectangle with a black background

Description automatically generated with low confidence

DAEN 690

Project Report

Bhargav Dasari

Sagarika Komati Reddy

Venkata Krishna Kothapalli

Kamani Madasu

Uday Madivada

Suparna Mannava

Spring 2025

Project Title

**About the Cover**

This semester, the DAEN program is proud to spotlight one of our esteemed capstone partners, Daniel Erasmus—a visionary whose groundbreaking work influences leaders worldwide. As the founder and CEO of Erasmus.AI, Daniel is a renowned futurist and a pioneer in scenario planning, artificial intelligence, and strategic foresight. His innovative approach to blending AI with human-centric decision-making has profoundly shaped global conversations on technology, sustainability, and future-readiness. Through his thought leadership, Daniel continues to inspire organizations across the globe to embrace change and build resilient futures.

At Erasmus.AI, Daniel conceived and led the development of ClimateGPT—the world’s first foundational AI model family focused on climate change. Built on over a decade of collecting and processing planetary-scale datasets, this groundbreaking innovation leverages AI to uncover hidden connections in global news, from Human-Centered Extreme Weather Dashboards to maps of global innovations, risks, and breakthroughs. The Erasmus.AI platform exemplifies his commitment to using technology to inform and address some of the world’s most pressing challenges.

As co-founder of The Digital Thinking Network (DTN), Daniel has spent over 25 years leading large-scale scenario planning and transformation processes. His work has driven notable actions, such as initiating a response to food security challenges during COVID-19 that delivered 1 million meals within three months and has since provided over 60 million meals in Sub-Saharan Africa. His scenario processes have also anticipated major global events, including the Global Financial Crisis in 2006 and the Oil Price Collapse in 2012—each resulting in multi-billion-dollar benefits for his clients. In the public sector, DTN's transformative initiatives include the Rotterdam Advisory Board, which spearheaded the Rotterdam Climate Initiative in 2005 with the ambitious goal of halving CO2 emissions by 2025, and the creation of the 30-year global future scenarios Ci’Num.

An accomplished author, Daniel has written three books on innovation and the networked society, as well as numerous columns, including the Information Society column for the Financial Times Review. He has also held various prominent board positions and fellowships, including serving on the University of Stellenbosch’s Faculty of Science Advisory Board, Cambridge-based Titan Advanced Energy Solutions, and the supervisory board of the Quad9 Foundation. Through his visionary leadership, Daniel continues to shape the future across disciplines and industries.

Contents

Table of Contents

[Abstract 1](#_Toc176093315)

[Section 1: Introduction 3](#_Toc176093316)

[1.1 Report Purpose 3](#_Toc176093317)

[1.2 Report Readership 3](#_Toc176093318)

[1.3 Report Structure 4](#_Toc176093319)

[Section 2: Problem Definition 4](#_Toc176093320)

[2.1 Problem Space 4](#_Toc176093321)

[2.2 Research 6](#_Toc176093322)

[2.3 Solution Space 7](#_Toc176093323)

[2.4 Project Objectives 7](#_Toc176093324)

[2.5 Primary User Stories 7](#_Toc176093325)

[2.6 Product Vision 8](#_Toc176093326)

[2.6.1 Scenario #1 8](#_Toc176093327)

[2.6.2 Scenario #2 8](#_Toc176093328)

[Section 3: Datasets 8](#_Toc176093329)

[3.1 Overview 8](#_Toc176093330)

[3.2 Field Descriptions 9](#_Toc176093331)

[3.3 Data Context 10](#_Toc176093332)

[3.4 Data Conditioning 10](#_Toc176093333)

[3.5 Data Quality Assessment 10](#_Toc176093334)

[3.6 Other Data Sources 11](#_Toc176093335)

[3.7 Storage Medium 11](#_Toc176093336)

[3.8 Storage Security 11](#_Toc176093337)

[3.9 Storage Costs 11](#_Toc176093338)

[Section 4: Select Section Title Based on Project Type 12](#_Toc176093339)

[4.1 Algorithms and Analysis 12](#_Toc176093340)

[4.1.1 Algorithms 12](#_Toc176093341)

[4.1.2 Analysis 12](#_Toc176093342)

[4.2 Machine Learning Model Exploration and Selection 13](#_Toc176093343)

[4.2.1 Model Exploration 13](#_Toc176093344)

[4.2.2 Model Selection 13](#_Toc176093345)

[4.3 Solution Approach 13](#_Toc176093346)

[4.3.1 Systems Architecture 13](#_Toc176093347)

[4.3.2 Systems Security 13](#_Toc176093348)

[4.3.3 Systems Data Flows 13](#_Toc176093349)

[Section 5: Select Section Title Based on Project Type 13](#_Toc176093350)

[5.1 Visualizations 14](#_Toc176093351)

[5.1.1 Visualizations and Interpretations 14](#_Toc176093352)

[5.2 Machine Learning Model Training, Evaluation, and Validation 14](#_Toc176093353)

[5.2.1 Model Training 14](#_Toc176093354)

[5.2.2 Model Evaluation 14](#_Toc176093355)

[5.2.3 Model Validation 14](#_Toc176093356)

[5.3 Testing and Validation 15](#_Toc176093357)

[5.3.1 Testing 15](#_Toc176093358)

[5.3.2 Validation 15](#_Toc176093359)

[Section 6: Findings 15](#_Toc176093360)

[Section 7: Summary 15](#_Toc176093361)

[Section 8: Future Work 15](#_Toc176093362)

[Appendix A: Domain Background 17](#_Toc176093363)

[Appendix B: Glossary 18](#_Toc176093364)

[Appendix C: GitHub Repository 19](#_Toc176093365)

[Overview 19](#_Toc176093366)

[GitHub Repository Link 19](#_Toc176093367)

[GitHub Repository Contents 19](#_Toc176093368)

[Appendix D: Risks 20](#_Toc176093369)

[Sprint 1 Risks 20](#_Toc176093370)

[Sprint 2 Risks 20](#_Toc176093371)

[Sprint 3 Risks 21](#_Toc176093372)

[Sprint 4 Risks 21](#_Toc176093373)

[Sprint 5 Risks 22](#_Toc176093374)

[Appendix E: Agile Development 24](#_Toc176093375)

[Scrum Framework Team Approach 24](#_Toc176093376)

[Sprint 1 Lessons Learned 24](#_Toc176093377)

[Sprint 2 Lessons Learned 25](#_Toc176093378)

[Sprint 3 Lessons Learned 25](#_Toc176093379)

[Sprint 4 Lessons Learned 25](#_Toc176093380)

[Sprint 5 Lessons Learned 26](#_Toc176093381)

[Works Cited 27](#_Toc176093382)

Table of Figures

[Figure 1: ChatGPT is the latest Natural Language Processing tool for Data Analytics. 5](#_Toc176093306)

[Figure 2: Data Analytics Engineering has the power to be transformative for an organization. 6](#_Toc176093307)

[Figure 3: Sprint project dates. 24](#_Toc176093308)

Table of Tables

[Table 1: Glossary Table 18](#_Toc176093309)

[Table 2: Sprint 1 Risks 20](#_Toc176093310)

[Table 3: Sprint 2 Risks 21](#_Toc176093311)

[Table 4: Sprint 3 Risks 21](#_Toc176093312)

[Table 5: Sprint 4 Risks 22](#_Toc176093313)

[Table 6: Sprint 5 Risks 22](#_Toc176093314)

**This page intentionally left blank**

Abstract

Abstract

The proliferation of synthetic opioids, especially fentanyl analogues, has escalated into a global crisis, exacerbated by the illicit distribution of precursor chemicals. This project focuses on identifying and prioritizing chemicals listed in the DEA Special Surveillance List (SSL) that are most associated with illicit synthetic opioid production and trafficking. Our central research question explores how a semi-automated risk-ranking model can help prioritize chemicals of concern by analyzing risk factors associated with company behaviors and government complicity. Previous studies and regulatory frameworks have primarily focused on tracking the chemicals themselves rather than the actors and systemic factors facilitating their misuse. Our research fills this gap by shifting the analytical focus from substances to supplier behavior and structural risk contributors. This approach enables a more strategic allocation of enforcement and investigative resources.

To achieve this, we developed a weighted scoring system based on three main risk dimensions: sanctions and indictments, government subsidies, and complicity indicators. Each category is divided into sub-factors, which are assessed and weighted to generate an overall risk score for each chemical. Data sources include government documents, legal case files, and international trade datasets. Additionally, we built an intuitive interface that allows users to enter chemical or company names and receive a visual risk ranking along with key risk insights.

Our findings reveal that a limited number of chemicals, when linked with high-risk entities and lax regulatory environments, disproportionately contribute to global trafficking routes. For example, we identified that complicity indicators, such as weak enforcement or government subsidy of suspect industries, are stronger predictors of illicit chemical use than chemical properties alone.

This model offers law enforcement, policymakers, and researchers a robust tool to prioritize interventions based on behaviorally and systemically grounded evidence. It also demonstrates the value of integrating data science techniques with policy analysis to combat complex, transnational threats like synthetic opioid trafficking.

**This page intentionally left blank**

Report

# Introduction

## Report Purpose

This report documents the development of an **Illicit Synthetic Opioid (ISO) Determination Engine** to automate and standardize the identification of illicit activity. Addressing a critical data analytics challenge, the project leverages algorithms, data structures, and rule-based conditions to improve detection accuracy. The report details the implementation of logging and event tracing, threshold-based classification, and visualization tools for better decision-making. By measuring the impact before and after automation, this project aims to enhance transparency, efficiency, and maintainability in illicit activity detection.

## Report Readership

This report is intended for key stakeholders, including TraCCC, MITRE, and the Select Committee staff, who are involved in research, policymaking, and enforcement related to illicit synthetic opioid (ISO) activity. These partners will use the report to understand the automation and standardization efforts in determining illicit activity. The insights presented will help refine methodologies for tracking illicit supply chains, identifying key indicators, and improving intervention strategies. Additionally, the report provides a foundation for future enhancements, ensuring that the automated determination engine remains effective and aligned with evolving regulatory and enforcement needs.

Furthermore, this report serves as a valuable resource for future teams working on the project. It documents the algorithms, data structures, and user interface elements developed during this phase, providing a roadmap for continued improvements. Future contributors can build upon the existing work to enhance the accuracy, scalability, and usability of the ISO determination engine. The detailed insights into event logging, parameter weighting, and visualization tools will support ongoing development and maintenance, ensuring the system remains adaptable to new challenges in combating illicit synthetic opioid activities.

## Report Structure

This report is structured to provide a clear and logical flow of information. It begins with a Problem Definition, outlining the challenges associated with detecting illicit activities in controlled substance trade. The Methodology section details the data collection and analysis techniques, including heuristic modelling, web scraping, and network analysis. The modelling and validation section presents key insights derived from the analysis, followed by a discussion of Implications for law enforcement and policy. The report concludes with future work, offering strategic actions to enhance regulatory enforcement and disrupt illicit supply chains.

# Problem Definition

## Problem Space

This project represents a crucial step in addressing the fentanyl crisis by providing law enforcement and regulatory agencies with the tools to efficiently track and analyze illicit synthetic opioid activities. By focusing on specific aspects of the crisis, such as sanctions, indictments, government subsidies, and complicity indicators, the system ensures that enforcement bodies have a clear historical record of illicit actors. This will enable more targeted interventions, reducing the time spent on irrelevant cases and improving decision-making efficiency.

The implementation of a rule-based weighting system adds another layer of sophistication to the project. By assessing risk based on the chemical compositions of fentanyl precursor substances, this system can flag potential illicit activity early. It ranks chemicals and substances by their likelihood of being used in fentanyl production, allowing investigators to focus on high-risk compounds that pose the most significant threat. This level of detail will also help to anticipate trends in illegal distribution and prevent new, emerging threats from gaining ground.

Finally, the system's ability to measure the impact of interventions both before and after actions will drastically improve enforcement efficiency. By tracking the outcomes of operations, the system can highlight which tactics are most effective, allowing investigators to refine their approaches over time. This feedback loop will help to continually improve efforts to combat the opioid crisis by focusing on strategies that work while abandoning less effective methods. Through all these features, this project contributes to a more strategic and data-driven approach to tackling the fentanyl crisis.

## Research

**Introduction**

The research process for this project was designed to understand the **illicit synthetic opioid (ISO) trade,** particularly fentanyl, and develop a **data-driven approach** to enhance its detection and tracking. This section summarizes the research activities undertaken by the team, including **literature reviews, online searches, and collaborations** with external partners.

**Research Methodology**

The research was conducted using three primary approaches:

**Literature Review:**

* Studying existing reports from government agencies such as the U.S. Drug Enforcement Administration (DEA) and Centers for Disease Control and Prevention (CDC) on fentanyl trends [8][9].
* Reviewing case studies and past research on synthetic opioid supply chains, including insights from the U.S. Government Accountability Office (GAO) and Customs and Border Protection (CBP) [10][11].

**Online Data Collection and Analysis:**

* Extracting and analyzing data from federal seizure reports, surveillance lists, and regulatory frameworks related to fentanyl precursors.
* Using data analytics methodologies to correlate known precursor chemical trade routes with illicit drug networks.

**Collaborations:**

* Engaging with The MITRE Corporation and TraCCC, who provided insights into law enforcement strategies.
* Using DEA Special Surveillance Lists (DEA SLLs) to develop a framework for identifying fentanyl-related precursor chemicals.

The research was conducted in a **chronological manner**, starting with foundational readings, followed by data collection, and finally, integration of findings into a solution framework.

### Analysis of Evidence Provenance and Consolidated Evidence Files

* **Chemical Analysis of Fentanyl Variants:** Identification of different fentanyl analogs and their potency levels.
* **Supply Chain Tracking:** Documentation of precursor chemical origins, transport routes, and manufacturing locations.
* **Law Enforcement Case Studies:** Examination of major drug busts, seizures, and interdiction patterns.

### **Thematic Analysis:**

### a) Evolution of Fentanyl Trafficking

* Shift from legal pharmaceutical fentanyl to illegal, modified versions.
* Chinese suppliers provide raw materials, while Mexican labs produce the drug.
* Online markets and anonymous payments make distribution easier.

### b) Law Enforcement Challenges

* Constantly changing fentanyl formulas to bypass laws.
* Weak coordination between countries in enforcing regulations.
* Hard to track online drug sales and payments.

### c) Application of ISO Determiners

* AI models classify fentanyl variants based on their structure.
* Advanced forensic tools improve drug identification.
* Mapping supply chains help detect illegal activity.

### **Key Findings and Insights**

Our research revealed several key findings and insights, including:

* **Majority companies listed are based in China, with a smaller number located in Mexico and other countries.** This finding is consistent with the known supply chain for fentanyl, which involves the sourcing of precursor chemicals from China and the production of fentanyl in Mexico.
* **Several companies have been sanctioned or indicted for their involvement in the illicit fentanyl trade.** This highlights the importance of ongoing enforcement efforts to disrupt the supply chain and hold companies accountable for their actions.
* **The weight scoring model effectively identified several high-risk substances that warrant further investigation.** These substances may be precursors or analogs of fentanyl, or they may be used in the production of other illicit drugs.
* **Network analysis revealed potential clusters or networks of activity that could be indicative of illicit behavior.** These networks may involve companies working together to source, produce, or distribute fentanyl or its precursors.

Based on our research findings, we recommend the following next steps:

* Continue to monitor the activities of the companies listed: This monitoring should include open-source investigations, as well as collaboration with law enforcement and intelligence agencies.
* Conduct further research on the high-risk substances identified by the weighted scoring model. This research should focus on their chemical structures, legitimate uses, and potential roles as precursors or analogs of fentanyl.
* Investigate networks of activity identified through network analysis**.** This investigation should aim to determine whether these networks are involved in illicit behavior.

This information sharing can help to inform and support ongoing efforts to address the fentanyl crisis. Our research has provided valuable insights into the fentanyl supply chain and the companies and substances involved. By continuing to monitor these activities and conducting further research, we can contribute.

## Solution Space

The Illicit Synthetic Opioid (ISO) Determination Engine will be developed to semi automate and standardize the detection of illicit activity. This system will:

* Log and trace key events such as sanctions, indictments, and financial crimes.
* Implement a weighting system with predefined thresholds to assess:
  + Chemical compositions of substances
  + Criminal events (e.g., sanctioned or indicted companies, fraudulent transactions)
  + Other indicators of illicit activity that will be continuously discovered and integrated.
* Provide interactive visualizations of illicit activity determination for enforcement agencies.
* Allow for easy configuration and maintenance of detection parameters.
* Measure the impact of enforcement actions before and after implementation, evaluating accuracy, consistency, and scalability.

By leveraging data-driven analysis and automation, the ISO Determination Engine will enable law enforcement to prioritize the most significant threats, disrupt fentanyl supply chains, and enhance interdiction efforts. This systematic and scalable approach will support both law enforcement and policy-making efforts, strengthening the global fight against fentanyl trafficking.

## Project Objectives

* **What does the team assume it will learn after finishing this project?**

We assume we will be learning the following after finishing the project

Key Indicators in Fentanyl Distribution - Understanding the significant factors that contribute to fentanyl movement and its associated risks.

Effectiveness of a Semi-Automated Weighing System - Evaluating how well data-driven ranking improves prioritization and decision-making.

Challenges in Data Integration - Identifying the difficulties in merging structured and unstructured reports, surveillance data, and financial transactions.

* **What does the team assume they will achieve as a solution when they finish this project?**

The team assumes the project will deliver a semi-automated weighing system that can: Assign weighted risk scores to fentanyl precursor chemicals, Provide structured prioritization insights to support decision-makers in focusing on the most impactful cases. Improve resource allocation by reducing manual workload and enabling targeted responses.

* **What does the team assume it will achieve in terms of understanding about the problem after they finish this project?**

Expected Understanding of the Problem Space Up on project completion, the team assumes it will have a clearer perspective on:

The role of precursor substances and how their movement influences supply availability. The advantages and limitations of semi-automation in prioritization, including areas where further refinement is necessary. The impact of different weighing factors on ranking and how adjustments affect prioritization accuracy.

* **What does the team assume it will provide in value as a product of this project work to the world, targeted group, etc.?**

The semi-automated weighing system aims to deliver practical benefits to multiple stakeholders, including:

For Monitoring and Enforcement Agencies - A data-backed prioritization model to enhance decision-making efficiency.

For Policymakers & Regulators - Insights into supply chain vulnerabilities to aid in developing targeted policies.

For Public Health & Safety Organizations - A proactive tool to identify areas of concern and mitigate risks before they escalate.

For Future Research & Development - A scalable framework that can integrate machine learning and AI for enhanced predictive capabilities.

## Primary User Stories

**As a regulatory analyst or law enforcement officer, I need a way to easily input data on opioid transactions, sanctions, and indictments into the ISO Determination Engine.** I want the system to quickly analyze this information and provide an **automated classification of potential illicit activity** along with a **confidence score** that helps me assess the risk level. With these insights, I can **spot high-risk cases faster, track patterns of illegal distribution, and make informed decisions** to take the right action at the right time.

## Product Vision

### Scenario #1

The ISO Determination Engine is a regulatory analysis and risk assessment system designed for regulatory agencies and law enforcement task forces combating illicit fentanyl production. These organizations need a way to efficiently track and analyze opioid precursor transactions to identify illegal supply chains. Unlike traditional investigative methods that rely on manual data collection and slow case-by-case analysis, the system uses a rule-based weighting mechanism to detect complicity indicators and flag high-risk activities. Our product automates risk assessment, highlights patterns of illicit distribution, and accelerates enforcement actions through data-driven insights. However, its effectiveness depends on the accuracy and timeliness of data sources, as well as seamless integration with enforcement agencies existing databases and policies.

### Scenario #2

The ISO Determination Engine is a powerful platform that uses natural language programming and rule-based techniques to flag suspicious opioid-related transactions. Designed specifically for data science teams and analysts in government and compliance departments, it helps data analysts enhance the accuracy of models used to detect illicit activity. Unlike older, static systems that need constant manual adjustments and struggle with new threats, our system combines data from various sources, helps identify patterns, and provides data-driven risk assessments. With the help of dynamic learning algorithms and data-driven analysis, the engine improves detection accuracy and reduces false positives. The system needs ongoing training and validation to stay accurate, and integrating external data sources may bring challenges.

# Datasets

## Overview

The following are the data sets we are going to use in our project and a small overview of the datasets:

**Dataset-1: USDOJ fentanyl precursor chemical manufacturing indictments**

This dataset contains U.S. Department of Justice (USDOJ) indictments related to fentanyl precursor chemical manufacturing. It provides a detailed record of companies and individuals involved in producing and distributing fentanyl precursors, which are the raw materials used to manufacture illicit fentanyl. The dataset tracks these entities, their geographic locations, aliases, legal actions, and law enforcement findings, offering valuable intelligence on the global fentanyl supply chain.

**Dataset-2: Fentanyl precursors**

This dataset contains information on chemical substances used in fentanyl production. It lists CAS Registry Numbers (unique chemical identifiers), substance names, synonyms, and weight classifications for various fentanyl precursors. The dataset helps track key chemicals involved in illicit fentanyl manufacturing and their alternative names, making it useful for law enforcement and regulatory agencies.

**Dataset-3: Integrated Data and Evidence Collection(Company Reference Sheet)**

This dataset contains a list of companies involved in the production and trade of chemical substances, many of which are related to the fentanyl supply chain. It documents company names, aliases, operational status (active/inactive), team assignments, and associated risk indicators. Some companies are flagged as "PRC-Funded Hits", suggesting potential government connections. Others are categorized as suspected fentanyl precursor suppliers based on previous investigations.

## Field Descriptions

The following are the field descriptions for the datasets we have chosen

**Dataset-1: USDOJ fentanyl precursor chemical manufacturing indictments**

This section describes the fields in the dataset related to **USDOJ fentanyl precursor chemical manufacturing indictments**. Each field is crucial for tracking indicted companies and individuals involved in illicit activities.

1. **Press Release Date (Type: DateTime)** - The official date when the U.S. Department of Justice (USDOJ) publicly announced the indictment. This field follows the **DD MMM YYYY** format (e.g., "23 Jun 2023"). It helps establish the timeline of enforcement actions.
2. **Press Release URL (Type: string)** - The web address linking to the official USDOJ press release. This URL allows users to access the full details of the indictment. Each URL corresponds to a unique indictment and should not be duplicated in the database.
3. **USDOJ Action (Type: string)** - The legal action taken by the U.S. Department of Justice against the entity. The most common value in this field is **"indictment"**, indicating criminal charges filed.
4. **U.S. District (Type: string)** - The judicial district handling the case (e.g., **EDNY** for the Eastern District of New York, **MDFL** for the Middle District of Florida). This helps identify where the legal proceedings are taking place.
5. **Criminal Case Number (Type: string)** - The official case reference number assigned by the court (e.g., **23-CR-263**). This unique identifier ensures the traceability of court proceedings and related documents.
6. **Action Date (Type: Date Time)** - The specific date on which legal action was taken. This is distinct from the **Press Release Date** and indicates when the indictment was formally issued.
7. **Company Role (Type: string)** - Describes the involvement of the indicted entity. The primary role in this dataset is **"chemical manufacturer"**, which refers to companies engaged in producing fentanyl precursor chemicals.
8. **Company or Individual (Type: string)** - Indicates whether the indicted entity is a **company** or an **individual**. Most records in this dataset pertain to companies.
9. **Primary Name (Type: string)** - The official name of the indicted company or individual (e.g., **Anhui Rencheng Technology Co. Ltd.**). This is the main identifier used in legal documentation.
10. **In Evidence Database(Type: Boolean)** - Indicates whether the entity is included in an official law enforcement evidence database. Values are either **"yes"** or **"no"**.
11. **Top 25 List Inclusion (Type: string)** - Specifies whether the entity is among the **"Top 25"** most critical targets for enforcement. If included, this field displays its ranking (e.g., **“8"**). Otherwise, it is marked as **"no"**.
12. **Unique ISOTAC ID (Type: string)** - A tracking identifier assigned to each indicted entity (e.g., **DOJ-0001**). This ID ensures consistency in referencing companies across multiple records and legal documents.
13. **Aliases / DBA (Doing Business As) (Type: string)** - Alternative names used by the indicted company, including English nicknames. For example, **"Anhui Rencheng Technology Co. Ltd."** is also known as **"Rencheng"**. Multiple aliases are concatenated with a comma.

**Dataset-2: Fentanyl precursors**

This section describes the dataset fields related to chemical substances and their registry information. Each field is essential for tracking chemical compounds, their classifications, and alternative names.

1. **CAS Registry Number (Type: string)** - The unique identifier assigned to chemical substances by the Chemical Abstracts Service (CAS). Each chemical has a distinct CAS number (e.g., 19099-93-5) that ensures consistency across regulatory and scientific records.
2. **Weight (Type: integer)** - A numerical value representing the assigned significance or classification level of the chemical substance. In this dataset, all entries have a weight of 8.
3. **Category (Type: string)** - Indicates the classification system used for the substance. In this dataset, the category is CAS\_Reg\_No, signifying that the classification is based on CAS Registry Numbers.
4. **Substance (Type: string)** - The official chemical name of the substance. In this dataset, all entries correspond to 1-Z-4-Piperidone, 4-Piperidone, which is a key chemical compound used in various applications.
5. **Synonyms (Type: string)** - Alternative names or chemical derivatives of the substance. This field lists various synonyms (e.g., 1-Piperidinecarboxylic acid, 4-oxo-, phenylmethyl ester) that may be used in different regulatory, scientific, or commercial contexts. Multiple synonyms are stored as separate entries.

**Dataset-3: Integrated Data and Evidence Collection(Company Reference Sheet)**

This section describes the fields in the dataset related to **Integrated Data and Evidence Collection(Company Reference Sheet).** Based on columns **L to Q** in your dataset, here is a structured description of these fields:

**Dataset Fields Description (Columns L to Q)**

1. **Special Alert** **(Type: string)** - Flags specific alerts related to a company. It includes identifiers like "SA: PRC Funded Hit" and "SA: T3," indicating different categories of alerts based on predefined criteria.
2. **Company Evidence Count** **(Type: integer)** - Represents the number of evidence points associated with a company. A higher count suggests a stronger case or more supporting data regarding the company’s status.
3. **Company Evidence Weight (Type: integer)** - A weighted score assigned to the collected evidence for a company. This weight helps in assessing the significance of the evidence.
4. **Flagged by Matt/Andy** (Type: integer) - A binary indicator (0 or 1) showing whether a company has been flagged by specific individuals (Matt/Andy) for review or further analysis.
5. **Sanctioned Company** **(Type: integer)** - Indicates whether the company is under sanctions. A value of 1 means the company is sanctioned, while 0 means it is not.
6. **On Leads List** **(Type: integer)** - A binary indicator (0 or 1) showing if the company appears on a specific leads list, potentially for further monitoring or investigation.
7. **Special Alert (Duplicate Column)** **(Type: integer)** - This column appears to be a duplicate of "Special Alert," possibly used for different filtering or categorization purposes.
8. **Total Company Weight** **(Type: integer)** - Represents the final computed weight assigned to a company based on various factors such as evidence weight, special alerts, and flagged status.

## Data Context

**USDOJ fentanyl precursor**

• The USDOJ fentanyl precursor chemical manufacturing indictments dataset captures indictments and legal actions taken against companies and individuals suspected of producing and distributing fentanyl precursor chemicals.

• It includes press release dates, U.S. district court case numbers, company aliases, and geographic locations of these entities.

• The presence of “PRC Associations” and government-backed entities suggests that some companies operate with state involvement or indirect government support.

• this dataset is significant because it highlights the legal response to fentanyl precursor distribution and serves as a primary source for identifying high-risk entities.

**Fentanyl Precursors Expanded**

• The Fentanyl Precursors Expanded dataset catalogs chemical substances, their CAS numbers, synonyms, and classifications, providing insight into the specific ingredients used in illicit fentanyl synthesis.

• This data is essential in tracking supply chain movements, detecting illicit shipments, and monitoring precursor sales in international markets.

• Understanding the chemical context allows law enforcement and regulatory agencies to flag suspicious transactions, regulate exports, and disrupt illicit manufacturing.

• This dataset is particularly relevant when cross-referenced with company data, as it helps identify businesses that stockpile or distribute these precursor chemicals.

**Integrated Data and Evidence Collection(Company Reference Sheet)**

• The Integrated Data and Evidence Collection dataset provides a structured database of companies linked to fentanyl production.

• It integrates findings from multiple investigative teams, categorizing companies by:

- Active vs. inactive status (e.g., revoked licenses, name changes, or companies under investigation).

- Corporate aliases and rebranding tactics, which companies use to evade detection.

- Direct links to PRC funding or government-backed initiatives.

• Contextually, this dataset is crucial for understanding business operations, tracking entity relationships, and identifying new corporate structures formed after regulatory crackdowns.

## Data Conditioning

As part of our project, we have not performed any data conditioning or cleaning on the data. The datasets USDOJ fentanyl precursor chemical manufacturing indictments, Fentanyl precursors, and Integrated Data and Evidence Collection (Company Reference Sheet) have been used as they were provided, without any modifications or preprocessing. This means the data may still contain inconsistencies, missing values, or duplications that could impact analysis accuracy.

As we collect evidence, we may encounter empty values in the Integrated Data and Evidence Collection dataset, particularly in company records and operational status details. The remaining datasets, USDOJ fentanyl precursor chemical manufacturing indictments and Fentanyl precursors are clean as of our knowledge, with structured chemical identifiers and legal case references appearing consistent.

We may implement validation checks, standardization techniques, and entity resolution strategies in later stages to ensure optimal data quality. These steps will help refine our datasets, improve analytical accuracy, and enhance the reliability of findings drawn from the data. project, we have not performed any data conditioning or cleaning methods. The datasets USDOJ fentanyl precursor chemical manufacturing indictments, Fentanyl precursors, and Integrated Data and Evidence Collection (Company Reference Sheet) have been used as they were provided, without any modifications or preprocessing. This means the data may still contain inconsistencies, missing values, or duplications that could impact analysis accuracy.

As we collect evidence, we may encounter empty values in the Integrated Data and Evidence Collection dataset, particularly in company records and operational status details. The remaining datasets, USDOJ fentanyl precursor chemical manufacturing indictments and Fentanyl precursors are clean as of our knowledge, with structured chemical identifiers and legal case references appearing consistent.

We may implement validation checks, standardization techniques, and entity resolution strategies in later stages to ensure optimal data quality. These steps will help refine our datasets, improve analytical accuracy, and enhance the reliability of findings drawn from the data.

## Data Quality Assessment

The following are the data quality assessments of the datasets we have chosen

**Dataset: USDOJ fentanyl precursor chemical manufacturing indictments**

This dataset, sourced from the U.S. Department of Justice, contains records of companies and individuals indicted for manufacturing fentanyl precursors. It is employed for risk assessment and legal tracking.

* **Completeness :**
* The dataset covers key details like indictments, company names, aliases, locations, and fentanyl precursor confirmations.
* However, some fields, like "executive or employee of primary company", have a lot of "#N/A" values, which might indicate missing or unavailable data.
* Most of the important data is present, but some fields may need further verification.
* **Consistency :**
* The dataset is well-organized, but some variations in formatting exist.
* For example, company aliases are sometimes labeled as "aka" or "dba", which could be confusing.Some dates and descriptive notes are formatted differently, making it harder to ensure consistency.
* While mostly consistent, formatting needs to be standardized for clarity.
* **Uniqueness :**
* Each company is assigned a unique ISOTAC UID, which helps keep track of them accurately.
* However, some companies appear multiple times under different aliases, which might create duplicate-looking records.
* The data is unique overall, but handling aliases properly would reduce redundancy.
* **Integrity :**
* The dataset is built from official legal records, law enforcement reports, and USDOJ indictments, ensuring credibility.
* However, some company locations differ between press releases and indictment records, which might need verification.
* The dataset is reliable, but cross-checking location data and aliases would improve trustworthiness.
* **Conformity :**
* The dataset follows a structured format, but some columns use inconsistent labels (e.g., multiple ways of noting aliases).
* Some notes include extra details that might need to be structured better, rather than written in free text.
* Most of the dataset conforms to expectations, but minor formatting cleanups are needed.
* **Accuracy :**
* Since the dataset is based on government and legal sources, it is highly accurate.
* However, minor errors may exist due to inconsistent location data and potential duplicate entries under different aliases.
* Accuracy is strong, but a review of company names and locations would help eliminate small discrepancies.

**Dataset: Fentanyl Precursors**

This dataset documents various fentanyl precursor chemicals, their sources, and supply chain patterns. It helps in locating dangerous substances and organizations engaged in the manufacture of illegal drugs.

* **Completeness :**
* The dataset is almost complete with only 9 missing values in the Synonyms column and no missing values in other columns.
* Overall, the dataset is well-structured and does not have significant gaps.
* **Consistency :**
* The dataset has no empty rows or columns, which ensures structural consistency.
* However, 70 duplicate rows exist due to multiple synonyms for the same substance, which may need standardization. Mostly consistent, but duplicates should be addressed.
* **Uniqueness :**
* The CAS Registry Number serves as a unique identifier ensuring that each chemical entry is traceable.
* However, duplicate entries caused by different synonyms reduce uniqueness.
* Uniqueness is strong at the substance level but needs refinement in synonym handling.
* **Integrity :**
* The dataset maintains structural integrity, but duplicate rows across all columns (70 entries) need verification.
* The Weight column ranges from 4 to 10, suggesting controlled values, but potential validation is required.
* Integrity is generally solid but requires minor cleaning.
* **Conformity :**
* The dataset follows a structured format with properly categorized fields, ensuring conformity.
* All columns align with expected data types, and the use of CAS Registry Numbers ensures industry standard compliance.
* No major issues, but some synonym variations may require standardization.
* **Accuracy :**
* The dataset is believed to be mostly accurate, with proper chemical classification and validated sources.
* Minor verification of duplicate entries and synonym mapping may improve accuracy.

**Dataset :** **Integrated Data and Evidence Collection (Working Copy)**

This dataset is owned by TraCCC, and compiles evidence on fentanyl precursor suppliers, including indictments, sanctions, and risk factors. It is gathered through structured data integrations, open sources, and investigations.

* **Completeness :**
* Company Evidence Count and Company Evidence Weight are well-populated (263 filled rows).
* Other columns (Flagged by Matt/Andy, Sanctioned Company List, On Leads List) contain mostly zeros or a few values, making them appear complete but not necessarily useful.
* Special Alert column contains some missing (#N/A) values, requiring further review.
* Overall, while the data is present, it needs further validation to ensure meaningful completeness.
* **Consistency :**
* Due to incomplete data, it is difficult to assess consistency.
* Some columns contain only zeros, which may indicate missing data rather than true values.
* Special Alert has inconsistent values (e.g., 500, 100, 0, #N/A), which may require standardization. Overall, consistency is unclear and needs cleaning.
* **Uniqueness :**
* The Company Evidence Count and Company Evidence Weight columns show 153 and 574 unique values, respectively, which suggests a broad range of data.
* Values range from 1 to 831 (Company Evidence Count) and 0 to 68,139 (Company Evidence Weight), indicating some variation but potential outliers.
* Other columns have very few unique values (e.g., 0, 150, 10), which limits their usefulness.
* **Integrity :**
* The data appears structured but lacks clear validation rules, especially for flagged or sanctioned companies.
* Zero-filled columns may not be meaningful, suggesting that additional checks are needed to confirm accuracy.
* Integrity is questionable, requiring verification with source data.
* **Conformity:**
* The dataset follows a structured format, but the presence of #N/A values and inconsistent entries suggests the need for standardization.
* Special Alert values (500, 100, 0, #N/A) lack clear categorization, requiring normalization.
* Column definitions need to be clarified to ensure correct interpretation.
* **Accuracy :**
* Company Evidence Count and Weight seem reliable based on their unique values and range.
* Other columns have low variability (mostly zeros), which may suggest missing or incorrect data rather than accurate reporting.
* Special Alert column needs cleaning, as #N/A values indicate missing information. Overall, data accuracy needs improvement through validation and correction.

## Other Data Sources

For this project, we have exclusively used the provided datasets:

* USDOJ Fentanyl Precursor Chemical Manufacturing Indictments
* Fentanyl Precursors
* Integrated Data and Evidence Collection (Company Reference Sheet)

No additional external data sources were considered or incorporated into our analysis. These datasets were deemed sufficient for our research objectives, providing comprehensive information on fentanyl precursor chemicals, related actions, and entities involved in the supply chain.

## Storage Medium

For our project, we utilized the Microsoft ecosystem, leveraging Excel for data analysis and Teams for collaborative storage. This choice was driven by its ease of use, familiarity, and built-in features that supported efficient project management.

* **Microsoft Excel** served as our primary tool for data storage and analysis. Its robust functionalities enabled us to organize, filter, and process large datasets efficiently. Excel also provided a secure and structured repository for our datasets, reinforcing data integrity throughout the project.
* **Microsoft Teams** played a crucial role in facilitating seamless collaboration. With its centralized platform, our team could easily share files, discuss analyses, and stay updated on project developments in real time. The integration with Excel further streamlined our workflow, allowing for instant access and review of shared spreadsheets. This continuous feedback loop significantly improved our data analysis processes.

By combining these tools, we not only enhanced productivity but also ensured that our data remained secure, organized, and readily accessible, adapting to the dynamic needs of our project.

## Storage Security

Data security was a top priority throughout our project, supported by the resources available through George Mason University. We implemented stringent security measures across Microsoft Teams, and Outlook to safeguard sensitive information.

* **Microsoft Teams**: We maintained the confidentiality of our communications and shared data by enabling end-to-end encryption for chats and file sharing. Advanced permission settings ensured that only team members with specific roles could access sensitive information, further minimizing security risks.
* **Microsoft Outlook**: As our primary communication tool, Outlook’s advanced security features including malware filtering and phishing protection safeguarded email transmissions, meeting invitations, and other official correspondence. With the seamless integration of Outlook and Teams, coordinated by George Mason University’s IT services, we ensured secure and efficient project communication.

By implementing these security measures, we upheld high standards of data protection, ensuring compliance while minimizing potential risks to our project’s integrity.

## Storage Costs

Managing storage costs effectively was essential to maintaining the security infrastructure of our project. With the support of George Mason University, we were able to optimize resource allocation while keeping expenses under control.

* **Microsoft Licenses**: The university’s support significantly reduced the costs associated with Microsoft Excel, Teams, and Outlook, allowing us to benefit from their built-in security features without additional financial strain.
* **Operational Costs**: Daily expenses included maintaining security protocols, training team members on best security practices, and ensuring the secure integration of our communication tools. These costs were regularly assessed for efficiency and affordability, with George Mason University offering financial and technical support to help us achieve our security objectives.

By combining cutting-edge security solutions with strategic budget management, we ensured our project remained both financially viable and secure, aligning with our commitment to data protection and operational efficiency.

# Investigation and Exploration

## Solution Approach

This section details our investigative approach, covering our system architecture, security measures, and data flow. We emphasize data-driven evidence extraction and risk assessment for fentanyl-related crimes, leveraging machine learning and The ISO Determination Engine is designed as a structured, semi-automated system that collects, processes, and evaluates evidence to identify illicit synthetic opioid (ISO) activity. The workflow begins with evidence collection, where analysts monitor global sanctions, indictments, legal policies, and other relevant events. These events are logged as "Intake Objects," with associated attributes stored in a "Registered Attribute Log." Evidence is then extracted, anchored to specific intake objects, and compiled into an "Intake Data Log" for further analysis. From this evidence, substances and companies are processed separately. The system allows ISO Chemical Mappers to maintain and enrich substance data, including CAS numbers, synonyms, historical records, and registry sources. Substances are tagged using predefined "Substance Weighting Tags" such as CSA Schedules or indicators of illicit use. Similarly, companies are tagged with "Company Weighting Tags" based on activities like advertising ISO products or involvement in precursor chemical distribution. Both tagging systems are supported by a central Weighting Factor module, which houses tag categories, descriptions, and numerical weights used in scoring. These scores are accumulated and analyzed to assess risk, trigger alerts, and support enforcement prioritization. Evidence deconfliction, provenance tracking, and scenario-based analysis enhance the system's accuracy and transparency. The entire process is governed by flexible CRUD interfaces for tags, references, and analytics rules, allowing ISO Determiners to continuously refine the criteria and improve detection efficacy.

A diagram of a company

AI-generated content may be incorrect.

### Systems Architecture

Our system is a multi-component intelligence framework that operates through:

* **Evidence Collection:** Gathering data from legal records, law enforcement logs, and external intelligence sources.
* **Data Processing & Attribution:** Converting raw text and evidence logs into structured, actionable information.
* **Risk Assessment & Prioritization:** Leveraging weighting algorithms and machine learning models to evaluate and rank threat levels.
* **Decision Support:** Producing actionable intelligence reports to aid in decision-making.

### **Architectural Layers of Our System:**

* **Data Ingestion Layer:** Aggregates raw data from sanctions, indictments, reports, and ISO-relevant policies.
* **Preprocessing and NLP Pipeline**
  + Named Entity Recognition (NER) identifies chemical, and company mentions.
  + Dependency parsing extracts context around relationships.
* **Database Layer:** Stores structured attributes, intake objects, and evidence logs for retrieval and analysis.
* **Analysis Layer:** Models assess, score, and rank cases to determine priority levels.

**Integration of weighting factors:**

A crucial part of the architecture is the **weighted risk scoring engine**, located downstream from the evidence classification. This module uses the following **weighted formula** to compute an overall risk score:

Total risk Score=(S×w1​) +(G×w2​) +(C×w3​)

Where:

* S, G,C = Raw sub-scores derived from matched evidence.

The weights used:

* w1​=1.2 (Sanctions)
* w2​=1.1 (Government Subsidies)
* w3​=1.3 (Complicity Indicators)

These **weighting factors** reflect our prioritization of risk: we consider **complicity** the most critical signal, followed by **sanctions**, and then **government subsidies**. The architecture supports dynamic adjustment of these weights based on future expert input or evolving policy needs.

#### **UI and decision support:**

The final component of the system is a **Streamlit-based interactive dashboard**, which presents ranked company profiles alongside the evidence and score breakdown. Analysts can explore why a particular company is flagged, trace back to the evidence, and even adjust weights in real-time to see how it impacts rankings.

In conclusion, the weighting factors are not merely applied at the end of the process but are deeply embedded in the **architecture’s logic**, serving as a tunable layer that influences final decision outcomes based on parameter importance. This design ensures flexibility, interpretability, and actionability for stakeholders combating synthetic opioid trafficking.

A diagram of a company

AI-generated content may be incorrect.

### Systems Security

Due to the sensitive nature of the data involved in detecting illicit synthetic opioid activity, maintaining strong system security was a top priority throughout the development. The team incorporated multiple safeguards to protect the system’s infrastructure, data storage, and user interactions. The following measures were implemented to ensure both internal operations and external access points were secure against unauthorized access and potential data breaches:

* **Secure local environments:** All development and data processing were carried out on secure local machines to limit exposure to raw or unfiltered data.
* **Private GitHub repository:** The project's code and documentation were maintained in a private GitHub repository, accessible only to authorized team members.
* **Controlled access:** Data and models were shared only through secure collaboration tools and cloud platforms with restricted access to prevent misuse.
* **No exposure of sensitive data:** The team ensured that all datasets used in the project were free from personally identifiable information and any confidential data related to law enforcement.

### Systems Data Flows

* **Raw Data Intake** → **Intake Object Log** → **Evidence Extraction** → **Data Mapping & Weighting** → **Risk Scoring** → **Final Intelligence Output**
* Data is **transformed** into a **human & machine-readable** format to maintain provenance.

## Running the Investigation

### Preparing Data and Evidence

To support and streamline our work, we developed an Evidence Tracking Sheet to monitor and document the full lifecycle of evidence collection, from initial identification to tagging and analysis. In parallel, we established a centralized evidence repository to securely store all collected materials, such as sanction records, indictments, and other relevant artifacts. Each piece of evidence was reviewed and categorized into structured data formats, enabling easier extraction and downstream processing. These categories aligned with our system architecture, such as substance-related data, company actions, and legal source metadata.

To ensure consistency and traceability, we used this categorized data to populate a Key-Value Sheet which is our “conditioned” or “working” format. This sheet serves as a well-organized, architecture-aligned tracking interface, mapping raw evidence to specific system attributes like substance reference entries, associated synonyms, tag weights, and company identifiers. This conditioning and prepping process involved transforming unstructured information into clean, linkable formats such as resolving duplicates and aligning with internal taxonomies.

Additionally, we performed several background tasks not immediately visible: we created logic for linking intake logs to registered attributes, cleaned conflicting or overlapping data (deconfliction), and cross-referenced sources with tagging rules to support future automation. By setting up these foundational formats and workflows, we established a consistent, scalable method for turning raw evidence into actionable, tagged intelligence within our ISO determination architecture

### Algorithms & Analysis Methodology

# The process of identifying and evaluating illicit synthetic opioid (ISO) activity begins with the creation of an Object Log, which serves as the central intake repository for all incoming evidence. This log captures raw, unstructured data entries from a variety of sources, including sanctions lists, indictments, regulatory filings, and law enforcement records. Each entry in the Object Log represents an intake object containing potentially relevant information about substances, companies, or individuals. Once an intake object is registered, the system initiates attribute extraction using data-driven identification techniques to isolate key entities such as company names, chemical substances, aliases, and related metadata. These extracted attributes are then standardized and structured based on a predefined schema aligned with the Intake Data Attributes Log, ensuring the data is normalized and prepared for further analysis. After extraction, the standardized attributes are cross-referenced with existing databases and verified legal intelligence sources to accurately link companies, substances, and aliases to their respective records. Through robust entity resolution techniques, the system minimizes duplication and enhances mapping accuracy, ensuring reliable integration with regulatory frameworks, enforcement records, and industry registries. The verified records are then stored in structured formats, supporting efficient risk analysis and case prioritization. Once the entities are successfully linked, both supervised and unsupervised machine learning models are applied to assess the risk associated with each. Historical data, behavioral patterns, and inter-entity relationships are analyzed to identify suspicious activity. Algorithms such as K-means clustering help uncover hidden patterns and connections among substances, actors, and jurisdictions, enabling effective risk stratification. These models continuously improve by learning from new data and analyst feedback. Finally, the system synthesizes the extracted attributes, verified links, and model predictions to produce ranked risk assessments. These rankings are informed by weighted evidence scores, the severity and frequency of offenses, and corroborated intelligence sources. A statistical ranking framework is employed to prioritize cases, allowing law enforcement to focus resources on the most urgent and high-risk threats. This structured, data-driven output significantly enhances decision-making and supports timely, targeted enforcement interventions.

### Establishing and Maintaining Provenance

# **Each data entry is timestamped and traced** back to its source using a **provenance tracking system** that ensures data integrity. Every modification, update, or transformation is recorded in an data log, preserving the historical trail of evidence. This system enables analysts to verify the **origins, processing steps, and responsible entities** associated with each data point.

* **Transformations and updates are logged** for transparency, ensuring that every modification is tracked within an library log and object log. This logging process records all changes, including entity resolutions, and provenance updates, allowing for full traceability of data lineage.

### Accounting and Adjudicating Anomalies and Discoveries

* **Anomalies are flagged** during data intake and processing**.**
* **Investigators review flagged cases** if to determine they indicate **data errors or new criminal patterns.**

### Exploring the Solution Instrumentation (or Tooling)

We opted for unsupervised learning methods because the data we’re dealing with is unstructured and constantly evolving, making predefined labels impractical. Since fentanyl trafficking is highly dynamic, traditional supervised approaches wouldn’t be effective. Instead, we used K-means clustering and other unsupervised algorithms to uncover hidden patterns in entity relationships and case linkages. These models help our system adapt in real-time as new data comes in, enhancing anomaly detection and risk assessment. Plus, unsupervised learning allows us to scale across massive, diverse datasets without the need for extensive manual labeling.

### Instrumentation Selection

**K-means clustering was selected** over other techniques because:

* It **groups similar cases without labeled training data**, making it ideal for identifying hidden patterns in unstructured forensic datasets.
* It handles **large, unstructured datasets efficiently**, enabling rapid classification and analysis of fentanyl-related intelligence.
* It supports **dynamic data adaptation**, allowing models to update and refine risk assessments as new evidence emerges.
* Its **clustering capabilities help reveal hidden networks**, exposing potential trafficking patterns and criminal associations that may not be immediately evident.

## Running the Instrumentation

### **Data model integration**: Establishing structured connections between legal and forensic evidence by linking datasets from law enforcement, legal records, and forensic reports. This approach facilitates seamless data interoperability, ensuring that intelligence sources can be effectively cross-referenced to enhance the accuracy and reliability of case evaluations.

### Machine Learning / Artificial Intelligence (if applicable)

### Machine Learning Model Training, Evaluation, and Validation

**REPORT SECTION INSTRUCTIONS**

**For Machine Learning projects**, discuss your approach to the following with respect to the ML Model:

1. Training,
2. Evaluation, and
3. Validation of the ML Model.

**DELETE THIS TEXT BOX AFTER YOU HAVE READ AND UNDERSTOOD THE INSTRUCTIONS.**

### Model Training

### Model Evaluation

### Model Validation

### Pulling It Together

* Establishling Common Evidence Format
* Integrating the Data and Evidence
* Summarizing, or Rolling-up, data
  + Enabling Pivots and Filtering

## Testing and Validation

**REPORT SECTION INSTRUCTIONS**

For DevSecOps and MLOps projects discuss the solution approach and systems architecture.

**DELETE THIS TEXT BOX AFTER YOU HAVE READ AND UNDERSTOOD THE INSTRUCTIONS.**

### Testing

### Validation

# Outcomes and Visualizations

## Overview

The Illicit Synthetic Opioid (ISO) Determination Engine is a semi-automated intelligence system designed to support enforcement and policy efforts in combating synthetic opioid trafficking particularly fentanyl and its analogues. The platform focuses on evaluating and ranking the risk profiles of supplier companies by analyzing structured evidence of illicit activity, regulatory violations, and state complicity. Rather than solely tracking chemical substances, the system shifts focus to the behaviour of companies involved in the production of synthetic opioids. It empowers users such as DEA agents, policy analysts, and investigative researchers to make informed decisions based on data-backed risk scores and transparent evidence trails.

**Core Capabilities**

**1. End-to-End System Integration**

The ISO Engine integrates all key stages of analysis into a streamlined pipeline:

* Data Ingestion of enforcement records, sanctions, and regulatory documents.
* Secure GPT-Powered Extraction, which automates the parsing of unstructured text (e.g., press releases, case filings) into structured evidence.
* Evidence Tagging and Mapping using a centralized Evidence Tracker.
* Company Linking to associate evidence with known entities and chemical substances.
* Interactive Interface built with Streamlit for intuitive searching, comparison, and review.

This infrastructure allows near real-time evidence processing and risk analysis with minimal manual intervention.

**2. Risk Scoring and Weighting Model**

At the heart of the system is a flexible, rule-based scoring model that calculates a composite risk score for each company:

Risk Score = (Sanctions × 1.2) + (Gov. Actions × 1.1) + (Company Behavior × 1.3)

Where:

* Sanctions reflect formal legal or regulatory actions,
* Government Actions capture indirect state support or negligence,
* Company Behavior identifies indicators of trafficking, precursor handling, or regulatory failure.

Each factor is broken down into sub-categories with predefined weights (e.g., “Cooperation with traffickers” = +20), allowing for consistent and explainable scoring.

**3. Evidence-Based Justification and Traceability**

A defining feature of the ISO Engine is its commitment to transparency. Every risk score is traceable to:

* A specific piece of evidence, extracted and tagged using Secure GPT,
* Metadata such as the source type, publication date, and origin,
* A linked tag and score, maintained via a centralized Key-Value Conditioning Sheet.

This ensures scores are not black boxes, but fully explainable outcomes supported by verifiable data.

**4. Usability and Operational Flexibility**

The ISO Determination Engine is designed to support the needs of partners focused on combating illicit synthetic opioid activity through data-driven enforcement, policy, and research.

The platform can be used for:

* Support enforcement-related analysis by identifying and ranking high-risk supplier companies based on verifiable evidence and structured risk scores.
* Enable policy evaluation by allowing analysts to simulate regulatory changes through adjustable risk scoring weights.
* Facilitate research and academic inquiry by providing exportable, structured datasets that can be used for trend analysis, modelling, or integration with broader studies on transnational crime and public health.

The system is scalable and adaptable suitable for hands-on investigative work today, while also offering a foundation for future automation and expanded analytical capability

Therefore, the ISO Determination Engine delivers a powerful, user-friendly framework for **identifying high-risk actors** in the synthetic opioid ecosystem. By combining automated data extraction, structured scoring, and evidence traceability, the system provides actionable intelligence tailored to the needs of enforcement, public health, and investigative communities. This shift from monitoring individual substances to profiling high-risk companies represents a strategic evolution in how we address the opioid crisis.

## Visualizations

**REPORT SECTION INSTRUCTIONS**

For analysis projects provide the visualizations produced and their respective interpretation.

**DELETE THIS TEXT BOX AFTER YOU HAVE READ AND UNDERSTOOD THE INSTRUCTIONS.**

### Visualizations and Interpretations

# Findings

### 1. **Effective Modeling of Company-Level Risk**

Using a weighted scoring system, the team successfully modeled the risk associated with **DEA SLL chemicals** and **linked companies**. By incorporating parameters like:

* **Sanctions and indictments**
* **Government subsidies**
* **Complicity indicators**

The team demonstrated how these indicators can be **quantified and weighted** to yield a **Total Risk Score** using the formula:

Total Score=(S×w1​) +(G×w2​) +(C×w3​)

This approach enabled a **prioritized ranking** of companies in terms of their association with synthetic opioid activity.

### 2. **Entity Detection and Manual Tagging Is Effective**

Our **Named Entity Recognition (NER)** and **manual tagging interface** works. It enables users to:

* Select flagged companies or chemicals
* Label and verify associated risk indicators from source documents
* Store tagged evidence for audit trails

While manual tagging slows things down, it gives us **high precision**, which is critical for sensitive topics like DEA monitoring.

### 3. **Weight Adjustments Impact Final Rankings Significantly**

We tested different combinations of weights (w1, w2, w3) and observed that even slight changes could shift company rankings. For instance:

* Boosting the **Complicity Indicator** weight brought **less legally implicated but suspicious companies** higher in the list.
* Prioritizing **Sanctions** pushed known bad actors to the top (expected behavior).

This shows our model is flexible and we can tune it depending on DEA's priorities (e.g., legal precedent vs. suspicious behavior).

### 4. **Company vs. Chemical Focus Clarity**

**We’re prioritizing companies over chemicals**, even though chemicals feed into the risk model. As outlined in the report:

"The chemical is only a factor in calculating the score, but the entity of interest remains the company."

### 5. **UI and Backend Integration**

We’ve developed a fully functional end-to-end system that transforms raw evidence into actionable intelligence. The core workflow includes:

* Uploading documents
* Identifying entities (e.g., companies, chemicals)
* Tagging evidence based on standardized rules
* Applying a weighted risk scoring model
* Generating a sorted output of high-risk companies

While the backend handles data processing and scoring, the user interface (UI) is built to support practical, real-world interaction. The current UI allows users to:

* Perform CRUD operations on evidence and scoring parameters
* Search and retrieve company records based on risk profiles or identifiers
* Manually tag evidence associated with specific entities
* Compare companies across risk dimensions
* Adjust weighting parameters to simulate different enforcement or policy strategies

Although document processing and entity extraction are handled in the backend, the UI remains streamlined and responsive, optimized for usability by analysts and enforcement professionals.

### **Evidence-Based Justification Adds Transparency**

One of our most valuable findings is that **every risk score is backed by tagged evidence**. This makes our system not just predictive, but **explainable and defensible** something stakeholders would highly value.

We can export:

* Scores
* Raw tagged documents
* Reasoning per company

# Summary

The Illicit Synthetic Opioid (ISO) Determination Engine is a semi-automated analytical platform developed to support the detection, monitoring, and risk assessment of companies involved in the trafficking of synthetic opioids especially fentanyl and its chemical precursors. The system was created in collaboration with partners such as TRACCC at George Mason University and MITRE, who identified the critical need for scalable tools to track and prioritize high-risk entities in global supply chains.

At its core, the system ingests publicly available regulatory and enforcement data, including DOJ press releases, OFAC sanctions lists, DEA Special Surveillance List (SSL) entries, and trade reports. Using a combination of Secure GPT-powered extraction and rule-based processing, it identifies references to relevant companies, chemicals, and countries. The extracted evidence is then tagged manually or semi-automatically to highlight indicators of illicit activity, such as involvement in trafficking networks, regulatory evasion, or state-linked protection.

A custom-built weighted risk scoring model calculates a Total Risk Score for each company based on three core factors:

* Sanctions and Indictments (S)
* Government Subsidies and Complicity (G)
* Behavioral Indicators (C)

The scoring formula is configurable, allowing users to adjust the weights and explore how different priorities affect the ranking of companies. These scores are presented through a user-friendly Streamlit dashboard, which supports CRUD operations, search functionality, company comparisons, and weight adjustments.

One of the system’s most valuable features is its commitment to transparency and explainability. Each risk score is fully traceable to specific pieces of tagged evidence, complete with source metadata such as publication date, issuing agency, and document type. This makes the tool not only operationally useful but also defensible for legal, policy, and audit purposes.

While the system is not fully automated, its semi-automated architecture strikes a practical balance between machine-driven efficiency and human judgment particularly when interpreting complex legal language or context-specific risk factors. This approach reduces false positives while maintaining analytical rigor.

The ISO Determination Engine provides a strong proof-of-concept for how data science and open-source intelligence can enhance the visibility of illicit supply chains. It lays a solid foundation for future enhancements, such as real-time monitoring, automated tagging pipelines, and integration with larger threat intelligence systems.

# Future Work

While the ISO Determination Engine establishes a solid foundation for semi-automated risk assessment in synthetic opioid trafficking, several technical, operational, and strategic enhancements are necessary to maximize its real-world impact. Below are the key areas for future development:

**1. Automated Evidence Collection**

* Implement full automation of data collection from structured and unstructured government sources.
* Develop real-time scraping and ingestion pipelines using secured scrapping tools.
* Expand coverage to include news articles, court filings, academic reports, and trade documents using advanced NLP methods.

**2. Risk Model Expansion**

* Apply data-driven weight optimization techniques (e.g., regression, AHP).
* Incorporate confidence intervals or uncertainty scoring to better reflect evidence strength and data quality.

**3. Multilingual & Cross-Border Entity Resolution**

* Build advanced entity resolution models to link aliases and name variations across jurisdictions.
* Extend Named Entity Recognition (NER) to process documents in key non-English languages (e.g., Chinese, Spanish, Russian).
* Develop tools for global supply chain tracking of precursor chemicals.

**4. Operational Readiness & User Roles**

* Implement secure, role-based access control for multi-user environments (analysts, reviewers, admins).
* Enable audit trails for all system actions to support traceability and accountability.
* Support collaborative workflows for tagging, verification, and peer review.

**5. User Interface Enhancements**

* Develop dynamic dashboards for temporal risk tracking and trend analysis.
* Enable real-time weight simulations and exportable stakeholder reports for decision-making support.

**6. Security, Ethics, and Legal Compliance**

* Ensure compliance with privacy laws (e.g., GDPR, CCPA) in handling sensitive data.
* Apply robust security protocols to protect user data, evidence records, and backend systems.
* Implement fairness checks and bias audits to uphold ethical standards and prevent unjust risk labeling.

**7. Long-Term Vision: Full Intelligence Engine**

* Introduce real-time alert systems for high-risk activity detection.
* Enable investigator-specific dashboards and custom risk profiles.
* Explore blockchain integration for tamper-proof tracking of precursor chemical movements.
* Partner with border agencies and customs officials to feed insights into enforcement workflows.

The proposed roadmap reflects both the challenges and opportunities in scaling the ISO Determination Engine into a robust intelligence platform. Future work should prioritize automation, machine learning integration, multilingual support, and real-world operational readiness to ensure the system continues to deliver high-impact, actionable intelligence to enforcement, policy, and research communities.

**This page intentionally left blank**

Appendix

Appendix A: Domain Background

I. Introduction

Fentanyl, a synthetic opioid, has become a major public health crisis due to its extreme potency, low production cost, and widespread illegal distribution. Originally developed as a pain reliever for cancer patients, fentanyl is now a leading cause of overdose-related deaths worldwide. Unlike naturally derived opioids like morphine and heroin, fentanyl is entirely synthetic and 50 to 100 times stronger than morphine [8].

The crisis is fueled by illicitly manufactured fentanyl, which is frequently mixed with heroin, cocaine, or counterfeit prescription pills. These mixtures make it difficult for users to determine the strength of the drugs they consume, significantly increasing the risk of fatal overdoses [9]. The primary supply chain for illicit fentanyl begins with precursor chemicals from China, which are then processed in Mexico before being smuggled into the United States through well-established drug trafficking networks [10]. Despite significant efforts by U.S. law enforcement agencies, including Customs and Border Protection (CBP) and the Drug Enforcement Administration (DEA), drug traffickers continue to adapt, taking advantage of regulatory loopholes and online marketplaces to expand their operations [8].

**II. The Fentanyl Crisis and Its Origins**

**A. Medical and Illicit Use of Fentanyl**

Fentanyl was first synthesized in 1960 by Dr. Paul Janssen as a powerful painkiller for patients undergoing surgery, cancer treatments, and palliative care. While fentanyl remains valuable in medical settings, its potency also makes it highly addictive and dangerous when misused.

Illicitly manufactured fentanyl is produced without the strict safety and quality controls applied to pharmaceutical fentanyl, leading to extreme variations in potency. Users may unknowingly consume doses that are far stronger than their tolerance, leading to a sharp rise in overdose deaths. According to the Centers for Disease Control and Prevention (CDC), synthetic opioids like fentanyl are involved in more than 70% of opioid-related overdose deaths in the United States [9].

**B. The Role of China and Mexico in the Supply Chain**

China has historically been the largest supplier of fentanyl precursor chemicals. Despite efforts by Chinese authorities to regulate these substances, many unregulated chemical analogs are still produced and exported under different names [10]. These chemicals are then trafficked to Mexico, where drug cartels manufacture fentanyl before smuggling it into the United States.

Mexican drug cartels, such as the Sinaloa Cartel and Jalisco New Generation Cartel (CJNG), play a central role in fentanyl production and distribution. These organizations use sophisticated smuggling operations to evade law enforcement and supply the U.S. drug market. Once inside the United States, fentanyl is often mixed with other illicit drugs, further increasing the risk of accidental overdoses.

The following table summarizes the roles of each region in the fentanyl supply chain:

| **Region** | **Role in Supply Chain** |
| --- | --- |
| **China** | Main producer of precursor chemicals; some chemicals are legally exported but diverted to illicit use. |
| **Mexico** | Manufacturing hub for synthetic opioids like fentanyl using Chinese precursors. |
| **U.S.** | Primary consumer market: enforcement efforts focus on intercepting drug shipments and dismantling trafficking networks. |

**C. Distribution and Trafficking Methods**

To evade law enforcement, traffickers use various smuggling methods, including:

1. Dark Web and Cryptocurrencies – Online marketplaces allow anonymous drug sales, making it harder to trace transactions.
2. International Postal Services – Small packages containing fentanyl or precursor chemicals are shipped through standard mail services to avoid detection.
3. Border Smuggling through Mexico – Fentanyl is transported in hidden compartments inside vehicles or mixed with legitimate cargo shipments.
4. Counterfeit Prescription Pills – Traffickers press fentanyl into fake pills resembling prescription opioids such as Oxycodone or Xanax, misleading users [8].

**III. Public Health Impact of the Fentanyl Crisis**

**A. Overdose Trends and Mortality Rates**

The fentanyl crisis has led to a significant increase in overdose deaths in the United States. The CDC reports that over 70,000 overdose deaths in 2022 were linked to synthetic opioids, primarily fentanyl [9].

Fentanyl’s extreme potency means that even 2 milligrams—equivalent to a few grains of salt—can be a fatal dose for many individuals. Overdose symptoms include:

* Severe respiratory depression (slowed or stopped breathing)
* Loss of consciousness
* Confusion and drowsiness
* Blue-tinted skin (cyanosis) due to lack of oxygen

To counteract fentanyl overdoses, Naloxone (Narcan) is widely used as an emergency treatment. However, because fentanyl is much stronger than heroin, multiple doses of naloxone may be needed to reverse an overdose.

**IV. Economic and Social Impact**

The fentanyl crisis extends beyond public health it has severe economic and social consequences:

* Strain on Healthcare Systems - Hospitals and emergency responders struggle to handle the increasing number of overdose cases.
* Law Enforcement Costs - Drug enforcement requires massive resources, increasing the burden on law enforcement agencies.
* Workforce and Productivity Losses - Many fentanyl overdose victims are between ages 18 - 45, leading to a decline in economic productivity.
* Community Devastation - Families and local communities suffer from addiction-related crime, unemployment, and social instability [10].

**V. Law Enforcement and Regulatory Responses**

**A. U.S. Customs and Border Protection (CBP)**

* CBP has strengthened border security using advanced drug detection technologies such as X-ray scanners, K-9 units, and chemical analysis tools.
* In 2023, CBP seized over 27,000 pounds of fentanyl, preventing millions of lethal doses from reaching U.S. communities [8].

**B. Drug Enforcement Administration (DEA)**

* The DEA’s "One Pill Can Kill" initiative raises awareness about counterfeit pills containing fentanyl.
* DEA task forces work with Mexican law enforcement to shut down illicit fentanyl production labs.
* The DEA Special Surveillance List (DEA SLL) helps track high-risk precursor chemicals.

**C. Government Accountability Office (GAO) Recommendations**

* A GAO report emphasized the need for stronger interagency cooperation and enhanced tracking of synthetic opioids across postal and commercial shipping networks [10].
* The GAO has recommended expanding forensic analysis capabilities to improve tracking of fentanyl production methods.

**D. Legislative and Policy Measures**

* The FEND Off Fentanyl Act (S. 1271) proposes stricter penalties for fentanyl trafficking and aims to block financial transactions linked to drug cartels.
* The U.S. government continues to engage in diplomatic efforts with China and Mexico to reduce the flow of precursor chemicals [11].

**VI. Work done by Previous Semester Teams**

Before we, the students of Spring 2025, started working on this project, students from two previous semesters Spring 2024 and Fall 2024 had already contributed to it. The following explanation provides a brief introduction to their work.

In the Spring of 2024, the research teams deepened their investigation into government involvement and corporate deception within the fentanyl trade. Their analysis identified 22 companies engaged in government-supported production, with 14 key connections emerging from TraCCC data and two major corporations linked to both PRC Known Associations and active fentanyl manufacturers. By mapping over 140 corporate aliases, the teams exposed deliberate efforts to obscure identities and evade detection. Further network analysis revealed more than 76 identity linkages, shedding light on the complex network of fentanyl producers. The project’s impact extended beyond individual company investigations, as heatmaps and visualizations highlighted major fentanyl hubs in cities like Shanghai, Wuhan, and Jiangsu. Through structured datasets and scalable data models, the research team equipped law enforcement and policymakers with essential tools to monitor, track, and dismantle illicit fentanyl supply chains, contributing a significant step forward in combating the global opioid crisis.

Expanding on this work Fall 2024, multiple teams collaborated to address the rising issue of illicit fentanyl trade using advanced data analytics and investigative techniques. Their research involved extracting key evidence from images and unstructured text, uncovering 131 organizations, 34 suspicious keywords, and 31 foreign trade company links. By utilizing Optical Character Recognition (OCR) and Natural Language Processing (NLP), they automated the identification of fentanyl-related entities, ensuring scalability for future investigations. Additionally, network analysis revealed shared phone numbers among various companies, exposing coordinated operations within the fentanylsupply chain. Through the standardization and integration of company and contact data, the teams enhanced traceability and developed an interactive, structured database for deeper analysis. Their work resulted in actionable intelligence aimed at aiding law enforcement in dismantling illegal networks and strengthening interdiction efforts.

**VII. Methodologies for Applying ISO Determiners in Illicit Drug Trade Analysis**

To effectively utilize ISO determiners in analyzing illicit drug trade, the following methodologies are employed:

* Data Scraping & Cross-referencing: Extracting data from regulatory databases such as DEA Special Surveillance Lists and TraCCC research reports [12].
* Heuristic Modeling: Developing algorithms that assess substances based on structural similarities to known illicit drugs and their precursors.
* Supply Chain Analysis: Mapping the flow of precursor chemicals from manufacturers to distributors to identify suspicious activity.
* Surveillance Integration: Using intelligence reports, forensic data, and law enforcement case studies to validate ISO determiner classifications.

**VIII. Challenges in Implementing ISO Determiners**

Despite their potential, several challenges exist in applying ISO determiners effectively:

* Evolving Chemical Structures: Illicit drug manufacturers continuously modify fentanyl analogs to evade regulations, making detection and classification difficult [14].
* Limited International Cooperation: Regulatory misalignment between China, Mexico, and the U.S. hinders enforcement efforts. Need for better data-sharing agreements between agencies.
* Dark Web and Cryptocurrency Transactions: Many fentanyl transactions occur via dark web marketplaces, using cryptocurrency payments, making it harder for authorities to track illicit financial flows.

**IX. Conclusion and Future Research**

The fentanyl crisis is a rapidly evolving challenge, requiring continuous efforts from law enforcement, policymakers, and researchers. Future research priorities include:

1. Using AI-driven models to classify new fentanyl analogs and precursors.
2. Enhancing cross-border cooperation between China, Mexico, and the U.S. to regulate precursor chemicals.
3. Improving financial tracking of illicit drug proceeds through cryptocurrency and online transactions.

By implementing ISO determiners, machine learning, and improved regulatory measures, law enforcement agencies can more effectively track, intercept, and prevent fentanyl trafficking, ultimately saving lives.

**Appendix B: Glossary**

|  |  |
| --- | --- |
| **Term** | **Definition** |
| **AI (Artificial Intelligence)** | Systems that mimic human intelligence; used here for classifying fentanyl analogs and identifying risk patterns. |
| **Anomaly Detection** | Machine learning technique to identify unusual patterns or outliers, e.g., suspicious entities in datasets. |
| **CBP (Customs and Border Protection)** | U.S. agency responsible for securing borders and preventing illicit trafficking, including fentanyl. |
| **CDC (Centers for Disease Control and Prevention)** | U.S. health agency that tracks opioid-related mortality and overdose trends. |
| **Complicity Indicators** | Behavioral or contextual clues indicating company involvement in illicit activity (e.g., trafficking cooperation). |
| **Cryptocurrency** | Digital currency used for anonymous online transactions, often linked to dark web fentanyl sales. |
| **DEA (Drug Enforcement Administration)** | U.S. federal agency combating drug trafficking and controlling substances like fentanyl. |
| **DEA SLL (Special Surveillance List)** | DEA list of precursor chemicals frequently used in the illicit production of controlled substances. |
| **Dark Web** | A hidden part of the internet used for untraceable, illegal transactions including fentanyl sales. |
| **Entity Resolution** | Linking different representations of the same company or entity across datasets (e.g., aliases). |
| **FEND Off Fentanyl Act** | U.S. legislation aimed at increasing penalties for fentanyl trafficking and disrupting cartel finances. |
| **Fentanyl** | A synthetic opioid, 50–100 times stronger than morphine, responsible for a large number of overdose deaths. |
| **GNN (Graph Neural Network)** | A machine learning model that uncovers complex relationships between entities (e.g., companies, chemicals). |
| **Heuristic Modeling** | Rule-based method to evaluate substances and entities based on observed behaviors or features. |
| **ISO Determiners** | A system developed to classify and assess risk of entities involved in synthetic opioid trafficking. |
| **ISOs (Illicit Synthetic Opioids)** | Synthetic opioids like fentanyl, illegally produced and distributed. |
| **NER (Named Entity Recognition)** | NLP technique used to extract names of people, organizations, chemicals from text. |
| **NLP (Natural Language Processing)** | AI field used to analyze and extract insights from human language in documents. |
| **Naloxone (Narcan)** | Emergency medication used to reverse opioid overdoses, including fentanyl. |
| **OFAC (Office of Foreign Assets Control)** | U.S. agency managing sanctions; a key source for identifying high-risk entities. |
| **OCR (Optical Character Recognition)** | Converts images or scanned documents into machine-readable text. |
| **One Pill Can Kill** | DEA initiative to raise awareness of counterfeit pills containing fentanyl. |
| **Precursor Chemicals** | Raw materials used to manufacture synthetic opioids like fentanyl. |
| **Random Forest / XGBoost** | Machine learning algorithms used for classification and prediction in risk modeling. |
| **Risk Scoring** | Assigning numerical risk levels to entities based on defined indicators. |
| **Sanctions** | Government-imposed penalties on individuals or companies involved in illicit activities. |
| **Secure GPT** | GPT-powered tool used to securely extract and process evidence from legal/regulatory text. |
| **Sinaloa / CJNG Cartels** | Major Mexican drug trafficking organizations central to fentanyl production and smuggling. |
| **Structured vs. Unstructured Data** | Organized (tables, databases) vs. raw, unformatted data (text, documents, images). |
| **Synthetic Opioids** | Man-made opioids designed to mimic effects of natural opioids; often stronger and riskier. |
| **TraCCC (Terrorism, Transnational Crime and Corruption Center)** | A research institute at George Mason University contributing to data and analysis on synthetic opioid trafficking. |

Appendix C: GitHub Repository

Overview

**REPORT SECTION INSTRUCTIONS**

Provide a GitHub Link and the README.MD content. Do not just provide a link to the GitHub repository but provide a narrative paragraph which introduces the project. This section should mirror the look and feel of a well-documented professional GitHub site.

**DELETE THIS TEXT BOX AFTER YOU HAVE READ AND UNDERSTOOD THE INSTRUCTIONS.**

GitHub Repository Link

GitHub Repository Contents

Appendix D: Risks

Sprint 1 Risks

During this sprint1, the team undertook a structured risk assessment at the outset, refining it continuously based on technical progress and user feedback. Risks were identified through collaborative brainstorming involving data scientists, domain experts, and system engineers. This cross-functional approach proved essential in surfacing risks from technical, analytical, and operational angles. Mitigation strategies were drawn from industry best practices and prior experience with similar modeling initiatives.

The team’s early identification of critical risks such as data reliability and algorithmic accuracy—ensured that safeguards like data validation and fairness audits were embedded into development workflows from the beginning. Additionally, continuous testing of the weighted model allowed the team to detect and resolve issues in edge cases before deployment. However, not everything went perfectly. The initial structure of the weighting model lacked defined caps and minimums, leading to inconsistent outputs during early testing. With a more thoroughly defined framework up front, some rework could have been avoided. Moreover, user-facing concerns such as model interpretability and transparency weren’t fully addressed until later in the sprint, when they emerged more clearly during stakeholder feedback sessions.

On the whole, the team accurately identified most major technical and data-related risks, but underestimated some operational challenges particularly the reliance on timely updates from external data sources, which led to unforeseen delays. One unanticipated issue was the latency involved in cross-checking external chemical databases, which impacted data ingestion speed. Another was the inconsistent chemical nomenclature across data sources, which introduced complexity in integration not initially accounted for. Despite these surprises, the risk assessment process laid a solid foundation for managing complexity, and future teams are encouraged to apply a similarly proactive and multidisciplinary approach.

A table with text on it

AI-generated content may be incorrect.

Table 2: Sprint 1 Risks

Sprint 2 Risks

During the sprint2, the team carefully identified and documented potential risks associated with managing and processing external data sources. This process began with collaborative planning sessions involving data engineers, project managers, and security specialists to capture both technical and operational risks. These sessions were instrumental in surfacing concerns early and defining mitigation strategies rooted in practical experience and existing workflows.

One of the major strengths of this sprint was the team’s awareness of data inconsistency issues. Recognizing that integrating data from various sources could introduce format and structural discrepancies, the team wisely implemented standardization and cleaning procedures upfront, which significantly improved downstream reliability. The proactive approach to **version control** was also a success. By deploying a lightweight version control system and clearly documenting changes, confusion was minimized despite multiple contributors updating files simultaneously.

The team also correctly anticipated risks around missing or incomplete data, putting checks in place to validate and supplement gaps using secondary sources. Security and storage were prioritized, with strong access controls and cloud-based encryption mechanisms adopted early on, minimizing exposure to potential breaches.

However, there were areas where the team could have improved. The manual nature of early data validation created significant delays before automation scripts were developed. Had automation been prioritized sooner, the team could have saved valuable time and resources. Additionally, while most risks were well-identified, there was some underestimation of the effort needed for maintaining automation scripts and scheduling periodic updates, particularly as new data formats emerged.

An unanticipated risk emerged when different teams used inconsistent naming conventions for evidence types, which wasn’t caught until later in the sprint. This caused minor disruptions in the integration pipeline and required rework to harmonize terminology across modules.

In hindsight, this sprint validated the importance of proactive planning, especially around data handling and security. Future project teams are encouraged to prioritize automation earlier in the sprint, maintain consistent naming conventions across teams, and invest time in creating robust documentation and validation layers from day one. These actions can dramatically reduce rework and improve project velocity and data integrity across similar efforts.

A screenshot of a computer

AI-generated content may be incorrect.

Table 3: Sprint 2 Risks

Sprint 3 Risks

A table with text on it

AI-generated content may be incorrect.

Table 4: Sprint 3 Risks

In this sprint, the team identified several critical risks tied to the structure, security, and handling of data. The risk assessment was conducted during the planning phase through working sessions with system architects, backend developers, and analysts, ensuring a comprehensive understanding of potential pitfalls. The risks were not only identified early but also continually re-evaluated throughout development as new technical requirements surfaced.

A standout success in this sprint was the early recognition that the existing data structure was too complex for scalability. The team addressed this by beginning work on a more modular and adaptive structure that could evolve with the project. This forward-thinking decision helped reduce long-term rework and enabled smoother integration with other components.

Another key area the team got right was version control. While frequent file updates initially created confusion, implementing a proper version control system mid-sprint significantly improved traceability and collaboration. Similarly, security concerns were proactively managed by leveraging cloud-based storage solutions with access controls preventing any security breaches or loss of data.

Despite these successes, there were areas for improvement. The manual extraction and validation of evidence took longer than expected early in the sprint. Automation was discussed but not prioritized soon enough. Once implemented, automation scripts drastically improved efficiency—highlighting the importance of addressing manual workflows at the start of the sprint. Additionally, while risks were accurately forecasted overall, the complexity of adapting a new data structure in a live development environment was slightly underestimated, causing short delays in implementation and testing.

No major unanticipated risks emerged, but some integration friction was encountered as the new structure had to be reconciled with legacy components still in use. This emphasized the need for better communication between architecture and development sub-teams regarding system-wide changes.

Sprint 4 Risks

**REPORT SECTION INSTRUCTIONS**

Include the risk table associated with the Sprint. Below the risk table provide a narrative description of how the risks and mitigation plans were identified, what the team got correct, what the team could have done differently, how accurate was the team in identifying the risks, did the team encounter any unanticipated risks, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.

**DELETE THIS TEXT BOX AFTER YOU HAVE READ AND UNDERSTOOD THE INSTRUCTIONS.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Risk | Description | Probability | Impact | Mitigation |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Table 5: Sprint 4 Risks

Massa eget egestas purus viverra accumsan in nisl nisi. Amet consectetur adipiscing elit pellentesque habitant morbi. Amet consectetur adipiscing elit duis tristique sollicitudin nibh sit amet. A erat nam at lectus. Pulvinar elementum integer enim neque volutpat ac tincidunt vitae semper. Id neque aliquam vestibulum morbi. Nisl pretium fusce id velit ut tortor pretium viverra suspendisse. Mauris cursus mattis molestie a iaculis at erat pellentesque adipiscing. Id semper risus in hendrerit gravida rutrum. Sit amet mattis vulputate enim nulla aliquet porttitor lacus. A arcu cursus vitae congue mauris rhoncus aenean. Est ultricies integer quis auctor elit sed vulputate. Lectus arcu bibendum at varius vel pharetra vel. Sed elementum tempus egestas sed sed. Lacus sed turpis tincidunt id.

Sprint 5 Risks

**REPORT SECTION INSTRUCTIONS**

Include the risk table associated with the Sprint. Below the risk table provide a narrative description of how the risks and mitigation plans were identified, what the team got correct, what the team could have done differently, how accurate was the team in identifying the risks, did the team encounter any unanticipated risks, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.

**DELETE THIS TEXT BOX AFTER YOU HAVE READ AND UNDERSTOOD THE INSTRUCTIONS.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Risk | Description | Probability | Impact | Mitigation |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Table 6: Sprint 5 Risks

Sollicitudin tempor id eu nisl nunc mi ipsum faucibus vitae. Leo in vitae turpis massa sed elementum tempus egestas sed. Semper eget duis at tellus. Tempus iaculis urna id volutpat lacus laoreet non curabitur gravida. Enim nunc faucibus a pellentesque. Odio euismod lacinia at quis risus sed vulputate odio. Tempus iaculis urna id volutpat lacus laoreet non curabitur gravida. Odio tempor orci dapibus ultrices. Gravida cum sociis natoque penatibus et magnis dis. Ultricies tristique nulla aliquet enim tortor at auctor urna. Non consectetur a erat nam at lectus urna.

Appendix E: Agile Development

Scrum Framework Team Approach

**REPORT SECTION INSTRUCTIONS**

Provide a narrative of the team efforts in adapting a scrum framework for a data analytics engineering project. Describe how easy/difficult was it to adapt to the Scrum framework. Did the team conduct a daily standup? If not, how often did the team conduct a standup. Describe how easy/difficult it was to use the YouTrack tool to manage the project. Don’t be limited to just these questions. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.

**DELETE THIS TEXT BOX AFTER YOU HAVE READ AND UNDERSTOOD THE INSTRUCTIONS.**

A diagram of a timeline

AI-generated content may be incorrect.

Figure 3: Sprint project dates.

Ac orci phasellus egestas tellus. Faucibus interdum posuere lorem ipsum dolor. Dolor sed viverra ipsum nunc aliquet. Vitae et leo duis ut diam quam nulla porttitor. Lectus sit amet est placerat in egestas erat imperdiet. Blandit volutpat maecenas volutpat blandit aliquam etiam erat velit scelerisque. Nunc sed velit dignissim sodales ut eu sem integer. Massa sapien faucibus et molestie ac feugiat sed. Donec pretium vulputate sapien nec. Risus ultricies tristique nulla aliquet enim tortor at auctor urna. Tincidunt tortor aliquam nulla facilisi cras fermentum odio. Tortor vitae purus faucibus ornare. Tincidunt ornare massa eget egestas purus viverra accumsan in. Sed adipiscing diam donec adipiscing tristique. At lectus urna duis convallis convallis tellus id interdum. Nisl purus in mollis nunc sed id semper. Arcu cursus euismod quis viverra nibh cras pulvinar. Neque convallis a cras semper auctor. Semper feugiat nibh sed pulvinar proin gravida hendrerit lectus. Quam viverra orci sagittis eu volutpat odio.

Sprint 1 Lessons Learned

**REPORT SECTION INSTRUCTIONS**

Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.

**DELETE THIS TEXT BOX AFTER YOU HAVE READ AND UNDERSTOOD THE INSTRUCTIONS.**

Quis lectus nulla at volutpat diam ut venenatis tellus. Diam sollicitudin tempor id eu nisl nunc mi ipsum. Nisi vitae suscipit tellus mauris a diam maecenas. Habitant morbi tristique senectus et netus et. Cras adipiscing enim eu turpis egestas. Nisl nunc mi ipsum faucibus vitae aliquet. Eu facilisis sed odio morbi quis commodo odio aenean. Est ante in nibh mauris cursus mattis molestie a. Ipsum suspendisse ultrices gravida dictum fusce ut. Adipiscing elit duis tristique sollicitudin nibh sit amet commodo. Laoreet sit amet cursus sit amet dictum.

Sprint 2 Lessons Learned

**REPORT SECTION INSTRUCTIONS**

Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.

**DELETE THIS TEXT BOX AFTER YOU HAVE READ AND UNDERSTOOD THE INSTRUCTIONS.**

Volutpat blandit aliquam etiam erat velit scelerisque in dictum. Nam at lectus urna duis. Tristique senectus et netus et malesuada fames ac turpis. Nulla facilisi morbi tempus iaculis urna id volutpat lacus laoreet. A arcu cursus vitae congue mauris. Semper auctor neque vitae tempus quam pellentesque nec. Ut lectus arcu bibendum at varius vel pharetra. Risus at ultrices mi tempus imperdiet. Pellentesque habitant morbi tristique senectus et netus. Dictum varius duis at consectetur lorem donec massa sapien faucibus. Phasellus vestibulum lorem sed risus ultricies. Nunc aliquet bibendum enim facilisis gravida neque. Velit euismod in pellentesque massa. Enim facilisis gravida neque convallis a. Diam sollicitudin tempor id eu nisl nunc mi. Et sollicitudin ac orci phasellus egestas tellus rutrum tellus pellentesque. Integer quis auctor elit sed vulputate mi sit amet mauris.

Sprint 3 Lessons Learned

**INSTRUCTIONS**

Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.

**DELETE THIS TEXT BOX AFTER YOU HAVE READ AND UNDERSTOOD THE INSTRUCTIONS.**

Convallis posuere morbi leo urna. Tellus rutrum tellus pellentesque eu tincidunt tortor aliquam nulla facilisi. Sit amet nisl suscipit adipiscing bibendum est ultricies. Sed adipiscing diam donec adipiscing tristique risus nec feugiat in. Nullam non nisi est sit amet facilisis. Viverra orci sagittis eu volutpat odio facilisis. Justo laoreet sit amet cursus. Tortor consequat id porta nibh venenatis cras. In dictum non consectetur a erat nam at lectus urna. Leo duis ut diam quam nulla porttitor massa id.

Sprint 4 Lessons Learned

**REPORT SECTION INSTRUCTIONS**

Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.

**DELETE THIS TEXT BOX AFTER YOU HAVE READ AND UNDERSTOOD THE INSTRUCTIONS.**

Pellentesque elit eget gravida cum sociis natoque penatibus. Vivamus arcu felis bibendum ut tristique et egestas. Mauris commodo quis imperdiet massa tincidunt nunc. Euismod nisi porta lorem mollis aliquam ut porttitor leo a. Id cursus metus aliquam eleifend mi in nulla posuere. Sit amet risus nullam eget. Commodo viverra maecenas accumsan lacus vel facilisis. Dui ut ornare lectus sit amet. Scelerisque varius morbi enim nunc faucibus a pellentesque. Mi proin sed libero enim sed faucibus. Quam adipiscing vitae proin sagittis nisl rhoncus mattis.

Sprint 5 Lessons Learned

**REPORT SECTION INSTRUCTIONS**

Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.

**DELETE THIS TEXT BOX AFTER YOU HAVE READ AND UNDERSTOOD THE INSTRUCTIONS.**

Fermentum iaculis eu non diam phasellus vestibulum lorem sed. Donec et odio pellentesque diam volutpat commodo sed egestas egestas. Feugiat vivamus at augue eget. Tortor posuere ac ut consequat semper viverra. Vitae tortor condimentum lacinia quis vel. Tortor condimentum lacinia quis vel eros. In pellentesque massa placerat duis. Quam viverra orci sagittis eu volutpat odio facilisis mauris. Et netus et malesuada fames ac. Mauris ultrices eros in cursus turpis massa tincidunt dui ut. Neque egestas congue quisque egestas diam in arcu cursus. Sed faucibus turpis in eu mi bibendum. Sit amet venenatis urna cursus eget nunc scelerisque viverra mauris. Mi sit amet mauris commodo quis. Egestas sed tempus urna et pharetra. Ac auctor augue mauris augue. Blandit volutpat maecenas volutpat blandit aliquam etiam erat. Dignissim sodales ut eu sem integer vitae justo. Risus pretium quam vulputate dignissim suspendisse.

References

[1] pvilardo3164 and [Author removed at the request of original publisher], “7.1 drug use in history,” Social Problems Continuity and Change, <https://pressbooks.howardcc.edu/soci102/chapter/7-1-drug-use-in-history/> (accessed Feb. 2, 2025).

[2] “Facts about fentanyl,” DEA, <https://www.dea.gov/resources/facts-about-fentanyl> (accessed Feb. 2, 2025).

[3] F. Dews and V. Felbab-Brown, “The fentanyl pipeline and China’s role in the US opioid crisis,” *Brookings*, Oct. 2024. <https://www.brookings.edu/articles/the-fentanyl-pipeline-and-chinas-role-in-the-us-opioid-crisis> (accessed Feb. 03, 2025).

[4] We bought what’s needed to make millions of fentanyl pills–for $3,600, <https://www.reuters.com/investigates/special-report/drugs-fentanyl-supplychain/> (accessed Feb. 3, 2025).

 [5] Zinnia Health, “How much does fentanyl cost? (The street Prices) | Zinnia Health,” *Zinnia Health*, Sep. 13, 2023. <https://zinniahealth.com/substance-use/fentanyl/costs>

[6] USA Facts, “Are fentanyl overdose deaths rising in the US?,” *USA Facts*, Sep. 27, 2023. [Online]. Available: <https://usafacts.org/articles/are-fentanyl-overdose-deaths-rising-in-the-us/>

[7] “The fentanyl crisis requires ongoing and strategic federal action,” *ASTHO*. <https://www.astho.org/communications/blog/the-fentanyl-crisis-requires-ongoing-and-strategic-federal-action/#:~:text=These%20strategies%20include%20increasing%20access,to%20treatment%20and%20recovery%20services>.

[8] “Frontline Against Fentanyl | U.S. Customs and Border Protection,” *www.cbp.gov*, 2024. <https://www.cbp.gov/border-security/frontline-against-fentanyl>

[9] “Illicitly manufactured fentanyl–involved overdose deaths with detected xylazine - United States, January 2019–June 2022,” Centers for Disease Control and Prevention, <https://www.cdc.gov/mmwr/volumes/72/wr/mm7226a4.htm> (accessed Feb. 6, 2025).

[10] U. S. G. A. Office, “Illicit opioids: While greater attention given to combating synthetic opioids, agencies need to better assess their efforts,” Illicit Opioids: While Greater Attention Given to Combating Synthetic Opioids, Agencies Need to Better Assess their Efforts | U.S. GAO, <https://www.gao.gov/products/gao-18-205> (accessed Feb. 6, 2025).

[11] M. P. Prekupec, P. A. Mansky, and M. H. Baumann, “Misuse of novel Synthetic opioids: A deadly new trend,” Journal of addiction medicine, <https://pmc.ncbi.nlm.nih.gov/articles/PMC5537029/> (accessed Feb. 6, 2025).

[12] “United Nations Synthetic Drugs,” United Nations : Synthetic Drugs, <https://syntheticdrugs.unodc.org/> (accessed Feb. 6, 2025).

[13] “The Federal Register,” Federal Register :: Request Access, <https://www.federalregister.gov/documents/2023/10/24/2023-23478/special-surveillance-list-of-chemicals-products-materials-and-equipment-used-in-the-manufacture-of> (accessed Feb. 6, 2025).

[14] Defense, security, and justice | rand, <https://www.rand.org/randeurope/focus-areas/defence-security.html> (accessed Feb. 7, 2025).

**This page intentionally left blank**