

# Identifying Fugitives Through Data Fusion and PAI/CAI

Machine learning streamlines the analysis of publicly and commercially available intelligence (PAI/CAI)—speeding up data processing by 65x—giving law enforcement novel capabilities for locating and investigating wanted fugitives.



*Striveworks developed a method for using open-source intelligence (OSINT) data to find potential fugitives and provided multisource analytics that fused tabular data with text from social media to locate known, wanted individuals.*

## Manual Analysis Struggles to Handle an Abundance of Semi-Structured Data

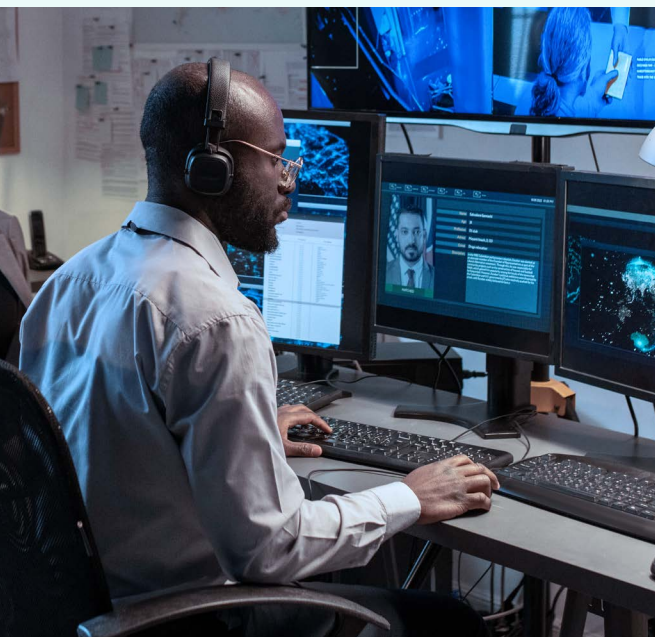
Law enforcement organizations have a standing mission to apprehend fugitives from justice—a major challenge that involves locating and identifying individuals hiding in plain sight among millions of upstanding citizens. In these efforts, agencies use publicly and commercially available data on people and businesses: public corporate records, social media profiles, collateral telemetry data, travel records, and other sources. These records hold invaluable information for directing police investigations on the ground. But **the enormous volume and disaggregation of the datasets** make it difficult and time-consuming for analysts to manually connect the dots and find relevant, actionable information on fugitives' whereabouts.

## Data Fusion Uncovers Corroborating Evidence to Identify Potential Fugitives

AI offered a way to streamline and scale the analysis of multisource data. Striveworks and two partner companies were brought in to apply machine learning to process vast amounts of openly available data in search of evidence to enrich fugitive profiles, ultimately narrowing in on the right individuals for further investigation.

Using broad demographic and professional profiles supplied by the client, Striveworks developed a complete analytic platform to meet these objectives. The tool was designed to surface three tiers of candidates for more focused investigation:

- 1 Companies potentially employing expatriates
- 2 Office locations of those companies
- 3 Potential fugitives



Because the available data included a mix of free text, geospatial data, profile images, semi-structured data, and other data sources, Striveworks originated a first-of-its-kind AI-powered data fusion workflow to analyze it. The workflow used AI models created by Striveworks to process scads of collateral telemetry data, as well as natural language processing models to perform named entity recognition and text classification on semi-structured business records and social media profiles. Striveworks then developed a **novel fusion algorithm to merge insights from these models and rank the probability** of specific corporations, office locations, and employees as worthy of further investigation.

## Deduplication Strengthens Predictions for Investigation Targets

Multisource data often includes a large number of duplicate records—a huge obstruction to machine learning and other automated analyses, which require clean and precise data. Many organizations have similar names, which may or may not refer to the same enterprise. (For example, Marco's Pizza in New York and Marco's Pizza in Atlanta are two unrelated businesses. Yet, Pizza Hut in New York and Pizza Hut in Atlanta belong to the same corporate entity.)

Striveworks cleaned the available data on corporate entities and individuals by developing an automated deduplication process to combine the repeat entries, including aliases, “doing business as” names, and other relevant associations. This deduplication delivered accurate, easy-to-understand details that the client could use to select and prioritize candidate investigations.

## Multisource Analytics: A Never-Before-Seen Tool for Apprehending Fugitives

Striveworks **created effective AI to harness large, disaggregated sets of PAI/CAI data** to automatically identify and locate fugitives. This multisource analytic workflow provided the customer with vital information about individuals, corporate affiliations, financial relationships, and potential employers, empowering them to direct their limited resources to high-probability targets.

## New Runtime Processes Data 65x Faster

The volume of data—over 12 million text records, plus similar amounts of tabular data—presented a major challenge. Early projections indicated that processing the text data would take 2.4 years—if the system didn't fail in the attempt.



Striveworks developed a new runtime that optimized this data processing, condensing the time needed to just 13.5 days—approximately **65 times faster** than the original runtime.

### WHAT IS DATA FUSION?

Data fusion combines information from multiple, varied sources to produce more detailed and accurate insights. This process can be difficult since each data source has its own unique characteristics (such as format, structure, and quality), which can cause conflicts and inconsistencies. Successfully implementing data fusion requires careful configuration, adept data management, principled mathematical modeling, and considerable computational resources.

This custom law enforcement application fused several data types to generate its predictions:

- Geospatial collateral telemetry data
- Free text from social media profiles
- Free text from company profiles and financial records
- Semi-structured data on corporate relationships, leadership, financials, and facilities

**STRIVEWORKS.com**  
sales@striveworks.com

