BABELSTREET

Uncover the Hidden Weaknesses Putting Your Borders & Country at Risk

Honest Guide to AI for Border Security

Introduction

Imagine this: you and your roll-aboard bag arrive at Toronto Pearson International Airport, rushing for a flight to Sydney. New imaging systems automatically X-ray your car and your luggage. Before entering the terminal, you stand at a biometrics station. It scans your retina. With your identity confirmed and your possessions examined, you simply walk onto the plane. No security lines. No passport checks. No need to take off your shoes.

Or suppose you're an importer/exporter whose trucks regularly travel from the United States to Mexico and back. You enter a government program that labels your company a trusted international trader. Your cargo may be automatically scanned at the border, but your drivers roll right through.

You're envisioning a "frictionless border" — one that enables smooth passage for the sea of people and cargo traversing international boundaries each day for completely legitimate purposes. Easy crossings are vital to national economies that increasingly depend on tourism and international trade. The simpler you make legitimate passage, the stronger your country's financial health. Hence the dream of frictionless borders.

Future technology may one day automate passage for legitimate travel and cargo. But today's AI-powered name matching and entity resolution technologies can streamline current customs and border processes while laying the groundwork for future innovation. They accomplish this by fostering interagency data sharing and helping border organizations to better screen people and companies before they reach national boundaries. In doing so, they also aid in spotting the people and businesses who cross borders to cause harm.

"Ninety-nine percent of all traffic may be completely legitimate, but the risks associated with that small element of malefactors is huge," says Declan Trezise, Vice President of Global Solutions Engineering for Babel Street. These are the risks posed by terrorists, human traffickers, drug runners, arms dealers, and purveyors of counterfeit goods — all of whom seek international passage alongside business travelers and families going on vacation.

New border solutions ease entry processes, improve security

Today, customs and border security agencies seek to ease passage and improve security by pre-screening travelers and devising integrated border management (IBM) strategies.

"Pushing the border out" is a term used for implementing technology that can empower customs and border officials to assess risk before people or goods arrive at a nation's land crossings, airports, or maritime borders. Obtaining this information early helps officials make entry decisions in advance, improving national security while easing entry as appropriate.

Michael Ivahnenko, CEO of Advoco LLC and former senior advisor to the United States Customs and Border Protection agency, points to land crossing as an example. "Systems that enable you to identify and verify what's in a container or who's in a car well before it arrives at your border are key facilitators of this concept," he says.

Integrated border management is another emerging approach for facilitating legitimate passage while improving national security.

IBM is an initiative that enables significant collaboration among customs and border security agencies by integrating border-related information, systems, processes, and people.

This collaboration can take place among different agencies working at the national, regional, and local levels of government.

The United States was among the first countries to implement a more integrated approach to border management. "Immigration" and "customs" operated as separate agencies before the terrorist attacks of September 11, 2001. The Federal government merged these agencies into United States Immigration and Customs Enforcement after publication of the *9/11 Commission Report*. This report detailed how lack of interagency collaboration and data sharing made it easier for terrorists to carry out their attacks. Since then, IBM has begun to take hold: different countries around the world now strive to coalesce border operations. (See sidebar, *Who's in charge?*)
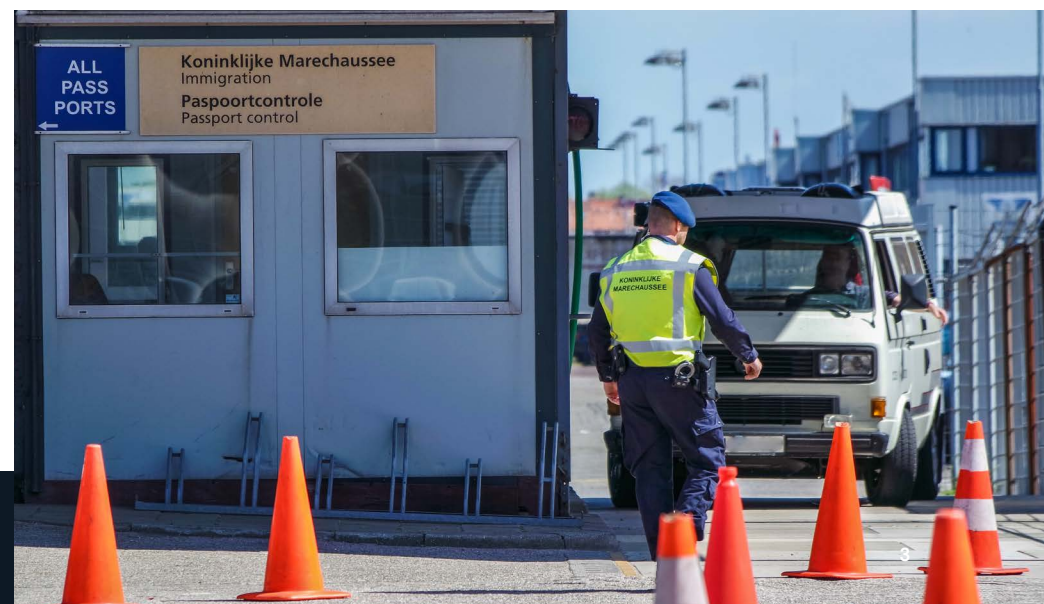
Increasingly, different forms of IBM are also taking shape not just among one nation's agencies, but among neighboring countries, and across different regions. Europol, the European Union Agency for Law Enforcement Cooperation, is a prime example. Europol serves as a central hub for sharing intelligence and combatting terrorism and organized crime among the EU's 27 member nations. Tony Smith, CBE, managing director of Fortinus Global and former Director General of the United Kingdom Border Force, believes that as the world shrinks, this type of collaboration grows more important.

"My own view is that international collaboration these days is critical to effective border management," Smith says. "The better you can work with your friends and neighbors on joint risk assessments, the better the chance of securing your border."

Who's in charge?

Despite integration efforts after the 9/11 terrorist attacks, the United States still divides the responsibility for border security and customs enforcement among a plethora of agencies. These include:

Federal agencies: U.S. Customs and Border Protection secures and facilitates operations at 328 ports of entry across the county, processing nearly 300 million visitors each year.[1] The Department of State issues visas, including nearly 7 million non-immigrant visas annually.[2] They are joined in their efforts by the Department of Defense; the Department of Justice/Federal Bureau of Investigation; and by Department of Homeland Security agencies including Immigration and Customs Enforcement, the United States Border Patrol, the Federal Emergency Management Agency, and the Transportation Security Administration.

State agencies: State agencies involved with border security and customs enforcement include state police departments, the National Guard, and state departments of public policy.

Local and regional agencies: These include county sheriffs' offices, municipal police departments, and regional port and airport authorities.

The United States is not alone in its disjointed approach to customs and border security. According to Smith, in the United Kingdom, at least 27 different agencies have some role to play in border management. These include security, policy, health, immigration, and customs agencies.

New border operations require new technologies

Consular officials, border agents, immigration officials, intelligence analysts, and others can use AI-powered name matching and publicly available information/commercially available information (PAI/CAI) platforms to more quickly and effectively assess the risks travelers and goods present before they reach the border. These officials can then scan and compile information from disparate databases to provide a single view of a person or corporation to better understand the risk they present. These capabilities help customs and border agencies "push out the border" and, through data sharing, take the first steps toward integrated border management.

Unfortunately, these efforts are too often hampered by legacy systems that inadequately match names; that glean only limited insight from PAI/CAI; and that cannot effectively communicate with each other.

Let's take a closer look at each of these challenges.

Inadequate name matching

Too many customs and border security agents rely on outdated search platforms to match names in structured text, such as when comparing names of incoming travelers against watchlists.

The name matching capabilities of these search engines lie somewhere between binary match/no match determinations and fuzzy matching — a computing approach that improves upon binary processes by considering degrees of truth. Returning only exact or near-exact matches, search platforms are fuzzy enough for general searches, but not expansive or fast enough for optimized name matching. Many accommodate only a limited number of languages, making it difficult to match for translated names, transliterated names, and names rendered in non-Latin scripts. (See sidebar, *The name-match challenge, and why AI matters*.)
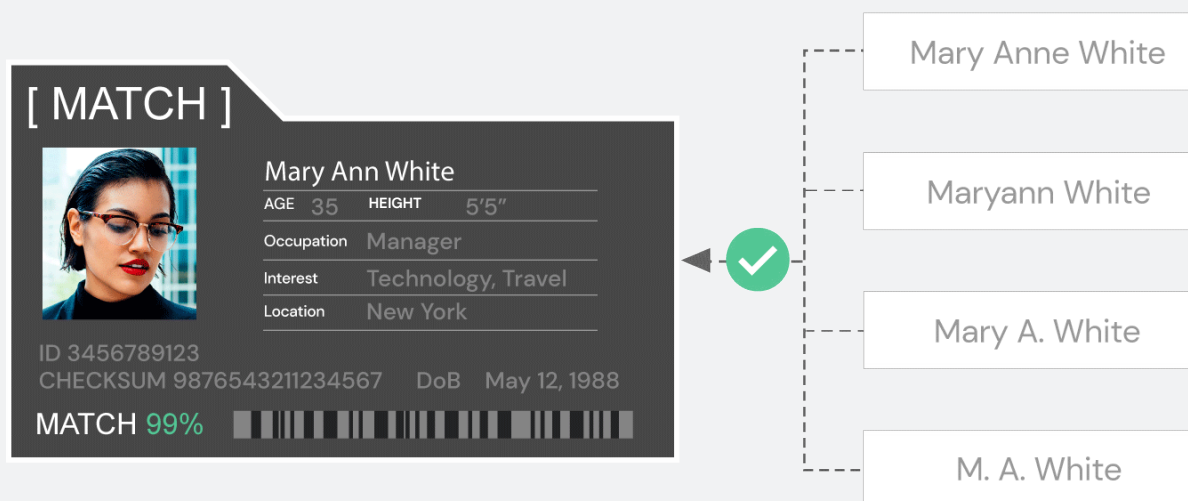
Match/no match processes also fail to spot aliases, nicknames, misspellings, honorifics, or out-of-order names. Due to these inadequacies, customs and border officials miss too many matches, allowing entry of criminals and contraband. Conversely, by returning too many false positives, these systems lead to unnecessary security alerts and inhibit the movement of legitimate travelers and goods.
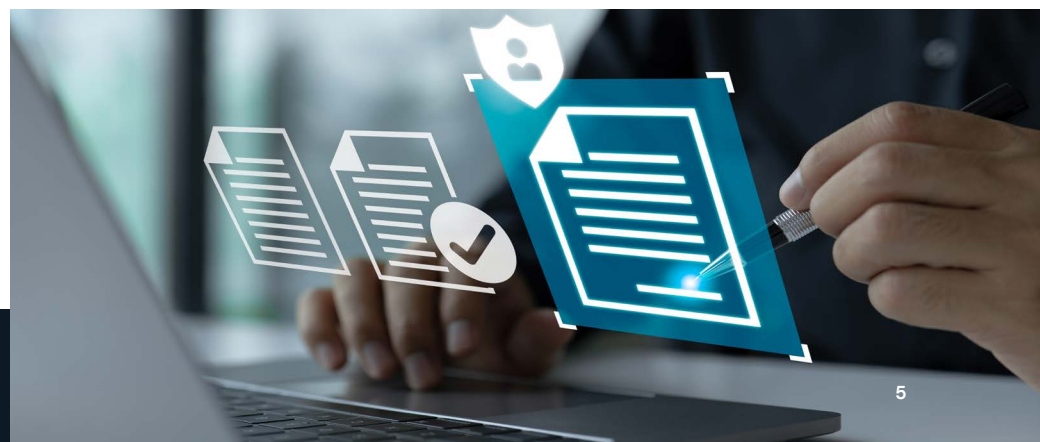
The name-match challenge, and why AI matters

Imagine you're screening for a known criminal named Mary Ann White. Full-text searches will find exact matches of her full name, spelled correctly, in English. They may also find slight variations of the spelling of her name, but not in any context-aware sense. Instead, full-text searches base matches on something called "edit distance" — how many letters differ between each version of the name. AI-powered fuzzy name matching technologies, meanwhile, can return variations of the name: Mary Anne White, Maryann White, Mary A. White, M.A. White, or White, Mary Ann.

These technologies also find misspellings such as Mary Ayn White, and variations of the first name "Mary Ann" that can be considered nicknames. (Mimi or Mae, as examples.) Cross-lingual name-matching can even match transliterated versions of names among languages using non-Latin scripts. If there exists in your database or list an Arabic, Chinese, Japanese, Chinese or Korean-transliterated version of "Mary Ann White," this version will also be matched.

[MATCH]

Mary Ann White

AGE 35	HEIGHT 5'5"
Occupation	Manager
Interest	Technology, Travel
Location	New York

ID 3456789123
CHECKSUM 9876543211234567   DoB   May 12, 1988

MATCH 99%

Mary Anne White

Maryann White

Mary A. White

M. A. White

AI-powered fuzzy name matching can help customs and border security agencies overcome the challenges of matching names in structured text. Automated NLP algorithms use a variety of criteria to quickly, accurately, and intelligently match and disambiguate names of people, organizations, and locations across a broad array of languages, scripts, and databases.

Limited insight from PAI

PAI is any data that is freely accessible by the public. This data includes social media posts, news stories and videos, information appearing on web sites, and more. In a world where people spend 27 percent of their time online[3], creating more than 2.5 quintillion bytes of data daily[4], there is an incalculable amount of PAI available for search.

Searching, monitoring, and analyzing PAI in real time can help customs and border security officials better identify potential threats. (Insight gleaned from PAI is called "open-source intelligence," or "OSINT.") These officials can use PAI systems to detect and track illegal cross-border activity; monitor the movements of individuals and groups of interest; and aid in real-time threat intelligence and response planning.

PAI capabilities help border officials better manage both potential and present threats. Consider a State Department official pre-screening travelers for United States visas. That official can use PAI platforms to examine social media posts and other content to determine whether the applicant is in any way associated with a criminal appearing on a watch list. Benefits are even clearer to those charged with preventing present danger. If a PAI system detects someone tweeting, "Just saw a woman abandon a bag at Liverpool John Lennon Airport, Gate 8," it can trigger an alert to airport authorities.

While some customs and border security organizations currently use PAI systems, those systems are often suboptimal. They suffer from:

- Poor data quality: The PAI searched can be incomplete or outdated.
- Inaccurate data: Improperly curated PAI can contain errors, biases, and inconsistencies that negatively affect the validity of insights derived from it.
- Insufficient data sources: Too many PAI systems collect data from a limited number of sources, leading to limited insight.
- An inability to accommodate data volume: PAI data is generated in huge volumes — too much for some PAI systems to appropriately search, process, and analyze.

Outdated legacy systems

Efforts to pre-screen travelers and integrate border management are further hampered by legacy systems' inability to communicate with each other. One outdated system simply cannot work well with another to find and present the data needed to pre-screen travelers and businesses, then communicate findings among agencies. Replacing these systems, or retooling them to communicate better, can be prohibitively expensive.

What to look for in a combined name matching/ PAI solution

To ease border processes and better secure their countries, customs and borders security agencies need a best-in-class combination of name matching and PAI solutions. But there are a number of name matching and PAI solutions on the market. What should you look for?

Your combined name matching and PAI solution must be capable of finding, analyzing, and coalescing data wherever it lives. To accomplish this in a cost-effective manner, you should consider an API-based solution, one that works on top of legacy systems to facilitate sharing from one application or data silo to another — avoiding the need to replace or re-tool older systems.

Your name matching and PAI systems themselves need to be interoperable. You cannot obtain the benefits of a combined system if your name matching and PAI platforms don't work well together, and with your existing technologies. For the easiest deployment possible, containerized delivery should be available. Deployment options should be up to you: cloud-based or on-site, depending on your organization's needs.

For PAI filtering, look for an automated solution that can access all layers of the internet, including the deep and dark web. Choose a PAI platform that includes a large and diverse library of enriched data, originating from a broad array of web sites; commercially available sources; and real-world interactions generated on chats, social media posts, and online comments. For the most up-to-date insight, it should provide persistent search capabilities. (See sidebar, *What is persistent search?*) In order for users in a variety of job roles to obtain immediate and actionable insight from your PAI platform, it should present findings via an easy-to-use, single pane of glass interface.

PAI platforms can only search effectively if they're looking for the right people. That's why combining your PAI solution with an AI-powered name matching system is so important. This system should be speedier and more accurate than existing systems. It should give you clear match scores, and the ability to adjust scoring parameters to meet your needs.

In addition, your combined PAI/name matching solution should offer translation and entity resolution capabilities.

Translation capabilities

All the OSINT in the world does border and customs officials no good if it's presented in a language they can't understand. This is obviously true for news articles, social media posts, and other

pieces of PAI. It is also true for names. It's a safe bet that very few American customs and border officials would recognize the name "Владимир Путин" as "Vladimir Putin."

The best PAI and name matching solutions automatically translate content from an array of different languages, helping customs and border security officials to monitor names and online content from across the globe.

Entity resolution

Entity resolution is the process of examining names appearing in unstructured text, then matching those names to entities appearing in a public knowledge base or the knowledge bases maintained by your organization. This capability helps you distinguish among multiple entities with the same or similar names.

Why is this capability important? There are 8 billion people on the planet. A lot of us are going to have the same name. Without entity resolution capabilities, State Department officials issuing B-1 business visas will struggle to distinguish between Wei Zhāng, the founder of a tech startup visiting Los Angeles to meet with potential investors, and Wei Zhāng, the owner of a factory that illicitly manufactures fentanyl to cut heroin.

Entity resolution capabilities automatically append identifying data to each name to distinguish between one Wei Zhāng and another. This data includes biographical information such as age, gender, street addresses, email addresses, and telephone numbers. It can also include information on the traveler's family members, employment, and education.

Similar capabilities distinguish between corporate identities. Many companies have similar names. For example, a simple web search of the corporate name "John's Burgers" may reveal a plethora of companies operating with this name in the United States and Europe. Your PAI/name matching solution should be able to link the right data to the right "John's Burger's."

In addition, the corporate world commonly uses initialisms and nicknames "PennyLuck Pharmaceuticals" may often be referred to as "PennyLuck Drugs." You may regularly dine at your favorite chain restaurant, "Bobby D's," or pick up your groceries at "VFG." The right name matching solutions can link these initialisms to the company's official names, "Robert D'Amico Foods, Inc." and "Very Fine Groceries, LLC."

What is persistent search?

Persistent search is a technology that keeps a search operation open whether someone is using it or not, recording updates and changes to the search term. Imagine a search for John Marshall of Sydney, Australia. A search may initially return information on his date of birth, his street address, his profession, the name of his spouse and children, and other data. If persistent search later uncovers an arrest for drug dealing, this information will be automatically appended to Marshall's record.

Conclusion

Too many existing name matching and PAI systems inadequately address pressing needs for advanced screening and name matching. They are slow, incomplete, and inaccurate. API-based, AI-powered, interoperable PAI and name matching solutions can help customs and border security officials ease passage for legitimate travelers and cargo, while better protecting their nations and laying the groundwork for future innovation. Use advanced name-matching and PAI systems to:

- Streamline border processing
- Reduce instances of false positive matches by up to 90% while concurrently missing fewer matches
- Screen for threats with greater context and understanding
- More quickly respond to and mitigate threats
- Improve analysis and collaboration capabilities

Babel Street's AI-powered software empowers customs and border security officials with the name matching and PAI-filtering capabilities needed to uncover critical insights across the massive digital landscape. Our software is used for more than half a billion watchlist checks each day. With more than three billion data sources in more than 200 languages, we offer the world's most comprehensive PAI library — with insight presented in the user's language of choice. That's why 84 percent of United States national security agencies along with similar agencies worldwide have partnered with us.

Start improving customs and border protection today. Learn more at www.babelstreet.com.

Endnotes

1. U.S. Customs and Border Protection, "Traveler and Conveyance Statistics," accessed May 2023, https://www.cbp.gov/newsroom/stats/travel

2. U.S. Department of State, "Report of the Visa Office 2022," accessed May 2023, https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics/annual-reports/report-of-the-visa-office-2022.html

3. Salim, S, Digital Information World, "More than six hours of our day is spent online – Digital 2019 Reports," accessed January 2023, https://www.digitalinformationworld.com/2019/02/internet-users-spend-more-than-a-quarter-of-their-lives-online.html

4. SG Analytics, "2.5 Quintillion Bytes of Data Generated Everyday — Top Data Science Trends 2020," accessed May 2023, https://us.sganalytics.com/blog/2-5-quintillion-bytes-of-data-generated-everyday-top-data-science-trends-2020/

Babel Street is the trusted technology partner for the world's most advanced identity intelligence and risk operations. The Babel Street Insights platform delivers advanced AI and data analytics solutions to close the Risk-Confidence Gap.

Babel Street provides unmatched, analysis-ready data regardless of language, proactive risk identification, 360-degree insights, high-speed automation, and seamless integration into existing systems. We empower government and commercial organizations to transform high-stakes identity and risk operations into a strategic advantage.

Learn more at babelstreet.com

BABELSTREET