The Tradeoffs That NPOs Must Make to Utilize Data to Defend Citizens

By Brandon Miner

Introduction

Imagine driving on I-5 a couple miles from Oregon. The road is calm until you see flashing red and blue lights in your mirror. You pull over, expecting a routine interaction. Instead, minutes stretch into half an hour as officers question you and ask to search through your car which you agree to, hoping that it will make this interaction end. With every passing moment of the stop, the fear and uncertainty grow stronger. You begin wondering why you were stopped, whether you are safe, and if your identity played a role in the encounter. For many Asian Americans traveling through Siskiyou County, California, this fear became a repeated reality.

In 2022, the American Civil Liberties Union (ACLU) led a civil rights class action lawsuit against Siskiyou County for the discriminatory actions against Asian Americans. The ACLU found that Asian Americans were 17 times more likely to be stopped and 25 times more likely to be searched by Siskiyou County deputies than groups. In 2021 alone, one conducted 71% of his traffic stops on Asian Americans. These numbers point to more than isolated incidents; they reveal a systemic pattern of racial profiling hidden within thousands of traffic stop records and body camera videos. In California alone, there were 5.1 million vehicle and pedestrian stops in 2024, generating nearly a century's worth of raw, continuous video footage. How can we be sure that there aren't other similar incidents hidden in the sheer volume of footage? For civil rights organizations like the ACLU, manually reviewing Siskiyou County footage alone proved to be a

daunting enough task to utilize data science to get the statistics mentioned prior using RIPA reports.

Artificial intelligence offers a powerful solution by making it possible to process and analyze massive amounts of footage without the need to train an in-house model, and increase the number of signals that can be analyzed. However, given that bodycam footage often has civilians giving their name, driver's license, insurance information, and more there is personally identifiable information (PII) that civil rights groups want to keep private. This introduces a profound paradox: how do we build an automated system capable of auditing law enforcement data without systematically compromising the privacy and civil liberties of the very people it is meant to protect?

The Resource Gap

The greatest hurdle facing civil rights nonprofits is the overwhelming imbalance of resources. Law enforcement agencies don't concern themselves with processing this data themselves. They hold it all in storage only for when required to give it out legally. This means that the computation load falls on the civil rights organizations that rely heavily on grants and donations. This means that the hardware available is limited to a few laptops rather than a proper server necessary to do the data processing required.

This hardware deficit becomes especially critical when dealing with modern policing data. Reviewing spreadsheets or written reports is manageable, but analyzing hundreds of hours of body camera footage and audio requires enormous computational power. But as mentioned before, meaningful oversight becomes nearly impossible without a form of automation. As a result, nonprofits are pushed toward artificial intelligence and machine learning systems capable of processing data at a scale humans cannot even with this imbalance because it is the best they can do.

However, this necessity creates a brutal infrastructure tradeoff. Do organizations invest scarce resources into building secure, localized servers capable of protecting sensitive data? Or do they rely on third-party cloud APIs that offer no hardware requirements at the potential cost of privacy, security, and control over deeply personal information?

Local Open-Weights vs. Proprietary APIs

To process this sensitive data, engineers must choose where the "brain" of the operation will live. The options are either locally — in the office, or remotely — On the distributor's servers. This forces a choice between two distinct architectural paths, each with its own fatal flaw:

The Local Route (Open-Weights)

Downloading an open-weight model is like buying the physical game disc and the massive console required to play it. You are literally downloading gigabytes of "knowledge" (parameters/weights) onto your own hard drive. Because it lives entirely in your house, your privacy is perfectly secure—but you have to foot the bill for the expensive hardware (GPUs) to actually run it.

- **The Pros:** The raw, uncensored PII of vulnerable citizens never leaves the organization's control, ensuring the highest level of privacy, security, and legal compliance. Sensitive body camera footage can be processed without exposing citizens' information to outside companies or cloud providers.
- **The Cons:** The capital expenditure is staggering. Most nonprofits cannot afford even one GPU, let alone a storage system or server infrastructure required to run these models efficiently at the scale demanded. Because organizations are often limited to smaller, less capable models that can fit within their hardware constraints, the system's reasoning power, speed, and accuracy become fundamentally weaker. On top of that, there is significant operational overhead involved in maintaining a local server, managing cybersecurity risks, handling software updates, and ensuring long-term system reliability.

The API Route (Proprietary Models)

Using an API is like streaming a 4K video game on your phone via the cloud. Your phone doesn't need to be powerful because a massive supercomputer in a data center is doing all the heavy lifting and just sending you the video feed. It's cheap and accessible, but you have to constantly send your data over the internet.

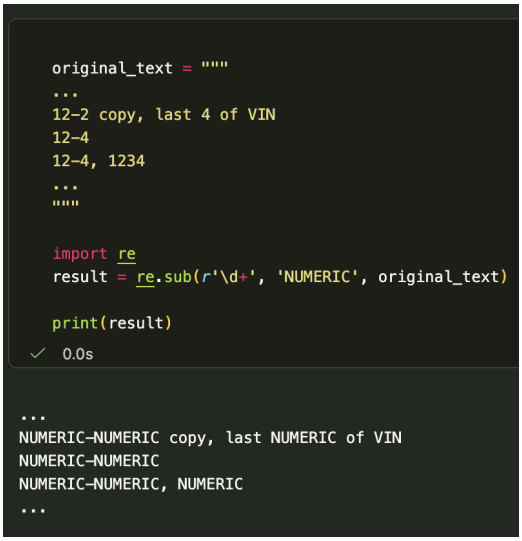
- **The Pros:**The hardware barrier is effectively gone. A single consumer-grade laptop is often enough to access the latest AI models. These models provide significantly stronger language understanding, pattern recognition, and analytical capabilities than most locally hosted alternatives, allowing nonprofits to process massive amounts of data far more efficiently and affordably.
- **The Cons:** Using proprietary APIs means transmitting sensitive information to third-party corporate servers. Because this data may contain uncensored civilian PII, including faces, names, addresses, license plates, etc., it cannot ethically or legally be sent in raw form. Before any data reaches a proprietary model, it must first go through extensive anonymization and redaction processes designed to protect citizen privacy.

This creates an engineering compromise: choosing the API route means building an anonymization pipeline powerful enough to safeguard identities while still preserving the contextual details necessary to detect discrimination, misconduct, or abuse. Yet as more information is stripped away to protect privacy, the system may also lose the very context needed to defend the citizens it was built to protect.

Anonymization Cascade

The first line of defense in protecting civil liberties is often a local anonymization model. Because Named Entity Recognition (NER) models are relatively lightweight, a civil rights organization can run them locally to scan and redact text without relying on external servers. However, these are statistical models that operate on probability rather than absolute certainty. This reality forces engineers into a dangerous balancing act between two failure modes. The most immediately damaging is the false negative: the model misses a sensitive detail, leaking raw civilian PII to an external, proprietary API.

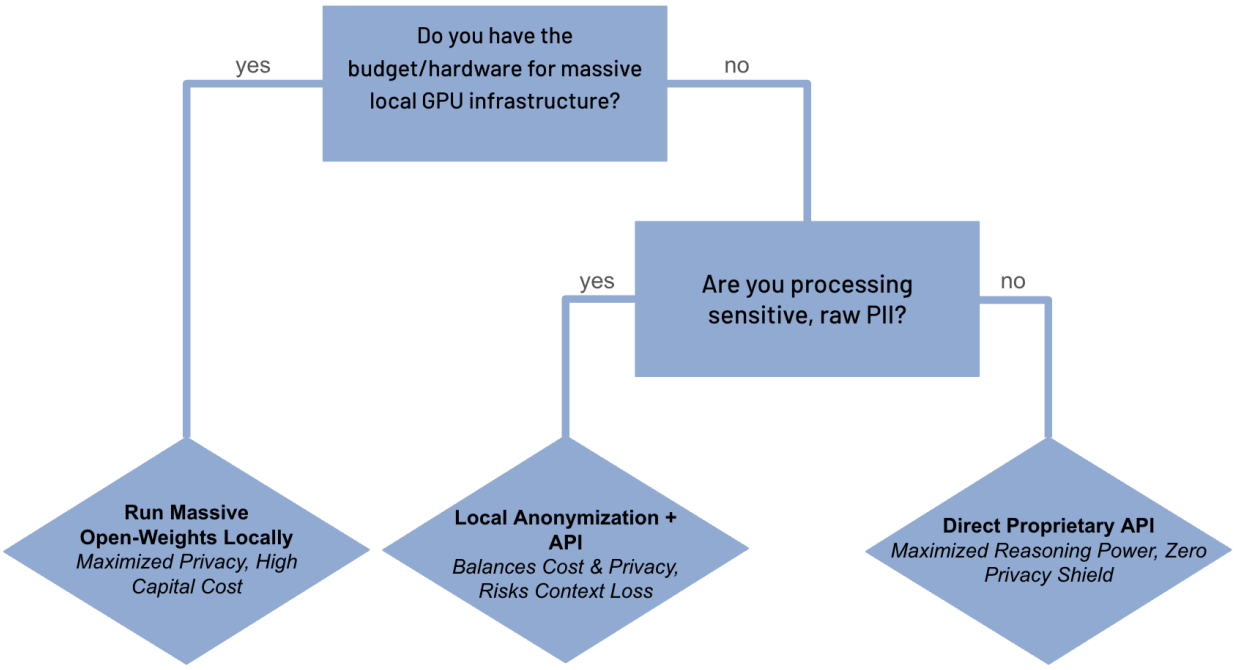
To aggressively prevent this, engineers deploy regular expressions (regex) as a blunt-force safety net. While a regex script can efficiently replace “123-456-7890” with “[PHONE_NUMBER],” it often acts as a “chainsaw” rather than a “scalpel”. When a broad regex command replaces every remaining digit with “[ALPHANUMERIC],” it introduces the second failure mode: the false positive. It violently strips away the context of the interaction. For example, take a look at the image below. The original text loses a lot of context due to the aggressive regex operation that makes sure no UID can possibly leak through. By the time the data is deemed safe enough to send to an API, the transcript looks like a heavily redacted court document, drastically reducing the artificial intelligence's ability to accurately detect a civil rights violation



The Ultimate Threshold

Suppose a nonprofit just doesn’t have the means to use a model locally, and set up an anonymization pipeline on a laptop that works perfectly. The citizen (and police officer) is fully protected. Yet in doing so, the system begins to strip away the very context needed to understand what actually happened. If the AI is effectively blindfolded, it may no longer be capable of accurately detecting the violations it was designed to surface. This is worsened by the fact that real bodycam footage is often literally noisy due to the sounds of traffic.

This leads to a deeply uncomfortable question: is a heavily censored, probabilistically limited AI audit better than manual review? For civil rights organizations confronting hundreds of hours of unreviewed footage, the answer is often a cautious yes. Even imperfect automated triage can surface patterns, prioritize cases, and reveal signals that would otherwise remain buried. In practice, it becomes the only viable pipeline that balances scale, cost, and a minimal baseline of privacy protection, even if it is fundamentally incomplete. AI-use at the very least can be paired with limited-manpower by directing what videos to watch.



A Path Forward

Navigating this resource and privacy gap is daunting, but civil rights organizations can take practical steps to utilize AI to its fullest without compromising their ethical obligations:

1. Implement Human-in-the-Loop (HITL) Auditing: No automated pipeline is flawless. nonprofits should mandate that human reviewers manually spot-check a random 5% sample of the AI's redacted transcripts. If it looks good, the pipeline proceeds; if not, this serves as an immediate feedback loop to catch bugs and tune the anonymization models before context is permanently lost.
2. Commit to Algorithmic Transparency: When NPOs publish findings based on automated analysis, they must explicitly disclose the anonymization tools, API pipelines, and prompt structures they used, acknowledging the inherent limitations and potential false positives of their system.

Takeaway & Conclusion

Civil rights organizations face a cruel reality: they urgently need data systems to help vulnerable communities, yet they lack the funding and infrastructure to build them. With dedicated data scientists and large-scale computing systems out of reach, these organizations are left dependent on imperfect tools to process overwhelming amounts of evidence.

This case study highlights a hard truth in engineering for the public good: perfect privacy and perfect utility cannot coexist. Any system that meaningfully analyzes sensitive data must make tradeoffs between accuracy, scale, cost, and privacy protection. The responsibility of data systems engineers, then, is not to eliminate these tensions but to make them explicit. By clearly defining tradeoffs, understanding the limitations of each pipeline, and designing systems with care and transparency, we can make the most responsible choices available even when none of them are ideal.

References

- https://www.aclunorcal.org/press-releases/page/
- https://oag.ca.gov/news/press-releases/california-racial-and-identity-profiling-advisory-board-releases-report-2024
- https://krcrtv.com/news/local/siskiyou-county-settles-aclu-lawsuit-over-alleged-racial-bias-targeting-asian-americans