

The Quiet Panopticon: “No Kings” Protests as Beta Test for Total Surveillance

Douglas C. Youvan

doug@youvan.com

www.youvan.ai

October 19, 2025

A generation after Orwell, the architecture he feared no longer needs telescreens; it rents dashboards. During the “No Kings” weekend, cities became living test ranges: ALPR grids traced plate-to-place, DFR drones arrived as “first responders,” and OSINT consoles streamed the crowd’s public chatter into risk scores. Palantir-style graph tools fused it all—devices, faces, vehicles, associates—into a single pane of actionable certainty. Cell-site simulators quietly impersonated towers to map phones in the square, while geofence warrants conscripted cloud providers to tell who was where and when—legality now hinging on a growing circuit split. What used to be investigation is now instrumentation: measure first, ask later. Vendors call it safety; auditors call it opacity. Either way, the field trials are real, the feedback loops fast, and the guardrails thin. This paper documents the confirmed toolchain—ALPRs, IMSI-catchers, geofence dragnets, DFR drones, and industrial OSINT—then asks a question fit for free societies: when the state can know almost everything about a lawful assembly in real time, what remains of the space where a citizen can risk being seen telling the truth?

Keywords: No Kings, Palantir Gotham, OSINT, Babel Street, Dataminr, LifeRaft, ALPR, Flock Safety, Vigilant Solutions, Genetec AutoVu, DFR drones, Skydio, Shield AI, cell-site simulator, IMSI-catcher, StingRay, Dirtbox, geofence warrant, First Amendment, Fourth Amendment, circuit split, protest policing, surveillance state, panopticon, accountability, transparency.

A collaboration with GPT-5. CC4.0.

Outline

1. **Orwell Updated**
 - 1.1 From telescreen to dashboard
 - 1.2 Protest as product demo
2. **The Confirmed Toolchain**
 - 2.1 Industrial OSINT: Babel Street, Dataminr, LifeRaft
 - 2.2 Graph fusion: Palantir-class platforms (Gotham)
 - 2.3 ALPR networks: Flock Safety, Vigilant, AutoVu
 - 2.4 Drones/DFR: Skydio, Shield AI; aerial optics and thermal
 - 2.5 Cell-site simulators: StingRay/Dirtbox and phone mapping
 - 2.6 Geofence warrants: device dragnets by time and place
3. **Legal Landscape (Live and Contested)**
 - 3.1 Fourth Amendment basics (particularity, probable cause)
 - 3.2 Geofence warrant circuit split
 - 3.3 DOJ policies on cell-site simulators (content vs metadata)
 - 3.4 Public-records gaps and vendor NDAs
4. **Field Operations in “No Kings”**
 - 4.1 Pre-event tasking and OSINT triage
 - 4.2 Real-time orchestration: fusing plates, phones, faces
 - 4.3 Post-event attribution and network mapping
5. **Data Lifecycle & Risk**
 - 5.1 Retention, sharing, and secondary use
 - 5.2 Error, bias, and false positives at scale
 - 5.3 Chilling effects on speech and assembly
6. **Governance & Accountability**
 - 6.1 Warrants, audits, and minimization rules
 - 6.2 Procurement transparency and public oversight
 - 6.3 Independent testing and red-team evaluations
7. **Citizen & City Countermeasures**
 - 7.1 Policy: geofence limits, ALPR retention caps, DFR guardrails
 - 7.2 Practice: notice requirements, protest-tech transparency
 - 7.3 Civil tools: detection, FOIA kits, community spotter guides

8. **Research Agenda**

- 8.1 Measuring protest tech's deterrence effect
- 8.2 Auditing fusion accuracy and harm
- 8.3 Building privacy-preserving public safety

9. **Conclusion: Freedom After Instrumentation**

- 9.1 Keeping safety without total recall
- 9.2 Restoring a citizen's right to risk being seen

References

1. Orwell Updated

1.1 From telescreen to dashboard

Orwell imagined a one-way state gaze; what we built is a multimodal sensorium that converts people into telemetry. The telescreen shouted doctrine; the dashboard whispers certainty. It fuses streams—plates, pings, posts, faces—into a heatmap of “actionable now.” The shift is architectural: not a monolith that sees, but an ecosystem that measures. Cameras do not need to know your name if your car does; your phone does not need to confess your intent if your pattern does. The old bargain—*search first, justify later*—is inverted by instrumentation: *measure first, infer intent, then select which law to invoke*.

Dashboards change time. They compress it. Instead of casework unfolding over days, a graph-of-graphs offers a live, ranked list of *who matters* this minute. They change power. Instead of a single agency monopolizing vision, vendors federate it: municipal dashboards lease insight from private networks and cloud histories. And they change culture. The telescreen tried to persuade; the dashboard predicts and nudges. In a city instrumented end-to-end, the distinction between “public safety” and “population management” collapses into drop-down menus. Surveillance is no longer an act. It is a **mode**.

Key consequences:

- **Frictionless scope creep.** Adding a new data type is a procurement ticket, not a public debate.
- **Opacity by design.** Capabilities ride in under “configuration,” shielded by NDAs and “security through obscurity.”
- **Chilling by anticipation.** People do not silence themselves because they are watched, but because they might be *modeled* before they speak.

1.2 Protest as product demo

Mass assembly is where instrumentation proves itself. Protests generate the three things dashboards crave: density (many signals in one place), velocity (fast change), and ground truth (events later confirmed by arrests, reports, or news). For vendors, a citywide march is an A/B test in the wild: Can ALPR nets keep pace with rolling closures? Do DFR drones extend the command center’s “line of sight” enough to pre-position units? Does OSINT triage correctly elevate rumors to routes? Which fusion queries reduce the lag from *crowd* to *cluster* to *suspect*?

For agencies, the protest is a stress test that can be publicly framed as readiness: “same tools, extraordinary crowd.” Tactically, it is a rare chance to run the full stack: pre-event social listening; geofenced alerts on day-of; real-time linking of plates ↔ phones ↔ faces; post-event

attribution that turns a moving mass into named networks. Strategically, it trains the muscle memory of *measure, infer, act*—a loop whose speed becomes its own justification. If it works at peak load, it will be normalized at baseline.

The demo has collateral effects:

- **Policy by precedent.** What ran “just this once” becomes the default playbook.
- **Dataset ratchet.** Each event yields labeled examples that improve the next model.
- **Narrative capture.** After-action slides quantify safety gains while quietly omitting the cost: the shrinking surface on which a citizen can risk visibility without entering a ledger.

In Orwell’s world, the spectacle enforced obedience. In ours, the demo enacts inevitability: a proof-of-concept that redefines “best practice” before the public can name what, exactly, was practiced.

2. The Confirmed Toolchain

2.1 Industrial OSINT: Babel Street, Dataminr, LifeRaft

These platforms industrialize “public information.” They harvest and fuse open feeds—social posts, live news wires, forums, incident reports—then score them for relevance, credibility, and proximity to an operator’s watchlist. The value isn’t a single tweet; it’s the *pattern*: spikes in certain keywords near certain coordinates; the same handles appearing across channels; a rumor that propagates along known activist bridges.

Operationally, OSINT does three things before the street fills: (1) **triage**—surface likely routes, rally points, and flash-mob pivots; (2) **tasking**—hand off targets (places, tags, vehicles) to other sensors; (3) **labeling**—produce after-action “ground truth” that hardens the next model. At protest pace, these consoles become a risk-feed: what just changed, where, and who is amplifying it.

Failure modes: language- or sarcasm-blind sentiment, “engagement bias” (loud ≠ true), and spillover into protected speech. The chilling effect arrives not from reading private messages, but from treating *public talk* as telemetry.

2.2 Graph fusion: Palantir-class platforms (Gotham)

Fusion systems turn disparate facts into traversable structure. Entities (people, plates, devices, addresses, accounts) become nodes; relationships (seen-with, registered-to, lives-at, follows, transacted-with) become edges. A dashboard query—“Show me clusters near X that connect to Y within three hops”—feels less like search and more like *navigating a living map of possibility*.

Gotham-class platforms accelerate tempo: ingestion pipelines normalize new datasets; linkers match fuzzy identifiers; policies govern who can see what. The output is operational grammar: subgraphs that look like “teams,” motifs that look like “tactics,” anomalies that look like “tomorrow.” What once took weeks of detective intuition now renders in seconds as a ranked list.

Risk is not merely false positives; it is **false structure**—edges inferred from co-location or shared infrastructure (same Wi-Fi, same rideshare hub) misread as association. At scale, over-connection becomes guilt by geometry.

2.3 ALPR networks: Flock Safety, Vigilant, AutoVu

ALPRs convert the street grid into a historical query surface. Fixed and mobile cameras read plates, timestamp them, and geolocate the reads. With citywide coverage, a single search shows where a vehicle has been for months, who it “travels with,” and which neighborhoods it regularly enters at odd hours. “Hotlists” generate real-time alerts; retroactive queries stitch a convoy out of thin air.

During protests, ALPRs do route intelligence (which arteries are live), perimeter sensing (which plates keep circling), and attribution (which out-of-area vehicles arrive just before flash points). In the aftermath, they power **pattern-of-life** reports: not just *who drove to the march*, but *where they go every Thursday*.

The hazard is retention and sharing. A 30-day buffer behaves very differently than 24 months; a closed city silo behaves very differently than a regional or commercial exchange. Transparency hinges on these dials—rarely debated with the public.

2.4 Drones/DFR: Skydio, Shield AI; aerial optics and thermal

Drone-as-First-Responder (DFR) reframes altitude as response time. Rooftop fleets, teleoperated from command centers, push eyes-on-scene in under a minute. High zoom, wide dynamic range, and stabilized video turn a block into a board; thermal adds a second channel through smoke, dusk, and distance.

In protest conditions, drones do three jobs well: (1) **overwatch**—live mosaics for commanders; (2) **cueing**—steer ground units with aerial breadcrumbs; (3) **documentation**—continuous video for later use. The psychology matters: a visible drone pacifies some and provokes others, altering crowd dynamics as surely as a line of bikes.

Edge cases include overflight of private space, drift from incident-specific to persistent area surveillance, and quiet plug-ins (person/vehicle re-ID). Aerial capability scales faster than policy; altitude is a governance gap.

2.5 Cell-site simulators: StingRay/Dirtbox and phone mapping

IMSI-catchers masquerade as cell towers, persuading nearby phones to connect and reveal identifiers. With directional antennas or multi-site setups, operators can locate a device to a building, then a room. Airborne “dirtboxes” expand the footprint, turning a city segment into a sweep of handset pings.

On demonstration days, CSS gear can: enumerate devices in a zone, watch churn as the crowd flows, and—under warrant—hunt a specific handset for targeted arrest or interdiction. Policy in many jurisdictions constrains content capture and demands warrants; the physics still yields **presence maps** even under tight rules.

The civil risk is silent inclusion. Phones join without owner consent, and even metadata (who was there, when) can be weaponized. Technical countermeasures exist (randomized IDs, radio hygiene), but are uneven, and crowds are RF-easy.

2.6 Geofence warrants: device dragnets by time and place

A geofence order compels a provider to disclose anonymized device IDs observed within a polygon during a time window, then—through a staged unmasking—identify selected devices. It is investigation by *where*, not *who*. In practice it is a dragnet whose legality now splits by circuit, with some courts rejecting it as overbroad and others allowing narrowed, multi-step procedures.

For protests, the temptation is obvious: draw a box around the skirmish, pull the IDs, cross-match with plates and posts. Even used “narrowly,” geofences invert particularity: the state starts with *everyone there* and works backward to *someone to charge*.

Risk isn’t only constitutional; it’s statistical. High-density zones inflate incidental inclusion (shoppers, medics, journalists). Once identifiers escape the sandbox, fusion engines will remember them—long after the march, the polygon persists in the graph.

Working principle across the stack: *Measure* → *Fuse* → *Act* → *Learn*. Each layer enriches the next, and each event labels tomorrow’s model. Without explicit, enforced brakes (retention limits, narrow warrants, public audits, true minimization), the default trajectory is clear: instrumentation becomes infrastructure, and exceptional playbooks become the baseline.

3. Legal Landscape (Live and Contested)

3.1 Fourth Amendment basics (particularity, probable cause)

The Fourth Amendment asks two old questions of new machines: **(a)** Do you have a good reason? (**probable cause**) and **(b)** Do you know exactly what you're looking for and where? (**particularity**). Search warrants are supposed to be **narrow, person/place/thing-specific**, and **time-bounded**. Modern instrumentation strains each element:

- **Scale vs. particularity.** Dragnet collection (everyone in a polygon; every plate on a corridor) reverses the order—collect first, specify later.
- **Inference vs. cause.** Risk scores and graph motifs feel like “reason,” but they're not evidence unless independently corroborated.
- **Persistence vs. reasonableness.** Long retention turns brief, justified searches into rolling dossiers, inviting “mission creep” beyond the warrant's purpose.
Doctrinally, courts still weigh **reasonable expectation of privacy**, **third-party doctrine** (data held by companies), and **exigent circumstances** (immediacy). The open problem: when cheap, pervasive sensing makes *everyone* “near” *something*, how narrow can narrow be?

3.2 Geofence warrant circuit split

Geofence warrants flip the axis: *where/when first, who later*. Appellate courts are diverging on whether that logic can ever be “particular.” Emerging themes:

- **Categorical skepticism.** Some opinions treat geofences as inherently overbroad—starting with “all devices present” is the very dragnet the Fourth forbids.
- **Conditional allowance.** Others tolerate **multi-step protocols** (initial anonymized drip, strict minimization, judicial re-authorization for unmasking) when polygons and windows are tightly drawn and anchored to concrete facts.
- **Journalists, bystanders, medics.** High-density events (protests, stadiums) heighten incidental capture, pushing courts toward stricter tailoring or suppression.
- **Remedies.** Good-faith reliance on unsettled law sometimes saves evidence; where courts find the warrant structurally defective, suppression follows.
Bottom line: legality now **depends on jurisdiction and drafting**. Expect eventual higher-court guidance; meanwhile, policy should assume **strict necessity, narrow scope, and auditable minimization** or risk exclusion—and backlash.

3.3 DOJ policies on cell-site simulators (content vs. metadata)

Federal policy draws a bright line: **no content interception** with cell-site simulators (CSS); configure as a **pen-register-style** tool to capture identifiers and signaling info, not calls/text contents. Core guardrails typically include:

- **Warrant default** (with narrow exigency carve-outs).
- **Deletion/minimization** of non-target device data promptly and automatically.
- **Use logging & supervisory approval** for activation, scopes, and target identifiers.
- **Disclosure to courts** that a CSS was used (to allow meaningful review).
Reality check: policies bind federal components; state/local practices vary. Even without content, **presence + movement metadata** can be highly revealing, especially when fused with ALPR and OSINT. The legal risk isn't only interception; it's **indiscriminate capture, prolonged tracking**, and **undisclosed use** that undermines judicial oversight.

3.4 Public-records gaps and vendor NDAs

Transparency is the hinge—and it's often missing. Procurement flows through **sole-source justifications, "trade secret" redactions, and managed service contracts** that relocate key details from public policy into private paperwork. Typical opacity points:

- **Capabilities cloaked as "configuration."** A feature toggle (e.g., re-ID, thermal add-ons, cross-jurisdiction data sharing) can be enabled without a public vote.
- **Data governance offloaded.** Retention, audit trails, access controls, and sharing rules live in vendor terms—redacted as proprietary.
- **FOIA friction.** Agencies cite security exemptions; vendors assert confidential commercial information; result: the public learns of practices only via leaks, litigation, or after-action marketing.
- **Privately run networks, public effects.** ALPR consortia and OSINT feeds can be region-wide while oversight stays city-by-city.

Design fixes: statute-level **warrant standards** for geofences/CSS, **public retention schedules** with default deletion, **capability inventories** with pre-approval for any new data type, **audit-grade logs** with external access, mandatory **impact assessments**, and **no-NDAs** on core government use terms. If instrumentation is here to stay, the law must see what the dashboard sees—or it can't limit what the dashboard does.

4. Field Operations in “No Kings”

4.1 Pre-event tasking and OSINT triage

The play begins days in advance. Analysts set **watchlists**: hashtags, phrases, venues, organizers’ public handles, anticipated feeder routes. Industrial OSINT consoles (Babel Street/Dataminr/LifeRaft) are tuned for **geofenced spikes** and **cross-platform propagation**—not to read private speech, but to spot where open talk hardens into logistics (rendezvous points, supply drops, jail support). Items that pass a confidence threshold are **tasked** to other sensors: ALPR cameras on likely ingress lanes; DFR launch pads closest to planned assemblies; patrol schedules adjusted for surge capacity.

Meanwhile, fusion teams stage **entity scaffolds** in Palantir-class graphs: venue nodes, known plate clusters from prior events, facilities (medic stations, houses of worship), and a baseline of normal traffic patterns. The goal is **day-of delta detection**—to know what today looks like against an ordinary Saturday.

4.2 Real-time orchestration: fusing plates, phones, faces

At step-off, telemetry floods in. Command posts run a **sensor braid**:

- **ALPR feeds** flag unusual plate density near chokepoints and “circling” vehicles that suggest staging or tailing.
- **DFR drones** provide wide-area overwatch and fine zoom; thermal helps with line-of-sight breaks and dusk.
- **CSS activations** (where authorized) enumerate handsets in a polygon; targeted pings track specific devices per warrant.
- **Crowd cams** and fixed CCTV contribute person/vehicle re-ID cues.
- **OSINT consoles** surface fresh pivots: a chant that signals a march turn, a posted livestream that reveals police lines or counter-protest movement.

Fusion operators knit these streams in near-real time: plates ↔ likely drivers (historic hits, registered addresses), phones ↔ movement vectors, faces ↔ prior bookings (if policy allows). The orchestration loop is **Measure → Infer → Task**: a cluster appears on the map; drones are vectored; ground units reposition; alerts are logged. Latency matters: ten seconds turns a rumor of “north on 5th” into verified motion; two minutes turns it into **after-the-fact**.

Two failure classes shadow the loop: **semantic error** (misreading a chant or banner as escalatory) and **structural error** (co-presence misinterpreted as association). Good discipline demands cross-sensor corroboration before action.

4.3 Post-event attribution and network mapping

When streets empty, analysis accelerates. The stack pivots from safety to **story**: who did what, where, with whom, and when.

- **ALPR retrospectives** reconstruct convoys, ingress/egress windows, and habitual overlaps (vehicles that repeatedly travel together outside protest days).
- **Device time-lines** (from lawfully obtained records) anchor presence in polygons to minutes, then to **micro-events** (a flare, a property line breach).
- **Video review** (DFR and street cams) yields person/vehicle re-ID tracks and high-confidence stills.
- **OSINT backfill** links offline moments to online posts or streams for **ground truth labeling**.

In the fusion graph, unnamed crowd nodes collapse into **attributed subgraphs**: affiliates, logistics roles, and pathways (who scouts, who signals, who moves supplies). The deliverable becomes a **case package** (chronology, exhibits, confidence levels, policy citations) and a **learning package** (model updates, sensor performance, playbook revisions).

This phase defines the ratchet: each labeled sequence improves the next detector; each successful fusion pattern becomes a saved query. Without explicit deletion and minimization, the graph acquires memory that outlives the march—converting a punctual event into **standing structure**. The operational question is solved; the civic one remains: how much past should a dashboard be allowed to remember about a lawful day on the street?

5. Data Lifecycle & Risk

5.1 Retention, sharing, and secondary use

Instrumentation does not end at collection; it begins a **life** for data. Three clocks matter: **retention** (how long it lives), **reach** (who can see it), and **reuse** (what else it can be used for).

- **Retention.** A 7–30 day buffer supports immediate incident response; a 12–24 month archive supports retroactive pattern-of-life. The same dataset transforms from “situational awareness” into **behavioral memory** by changing only the clock.
- **Sharing.** Inter-agency portals, vendor-operated exchanges, and regional task forces turn local capture into **networked visibility**. What one city stores, another may query.

- **Secondary use.** Datasets collected for traffic, sanitation, or “crowd safety” are attractive for unrelated aims: immigration enforcement, civil litigation, employee discipline, even commercial analytics. Mission scope expands because the data *exists*.

Risk concentrates where these dials are opaque. Without public, binding schedules and audited access logs, retention creeps, sharing multiplies, and secondary use becomes the default.

Minimum safeguards: publish dataset inventories; fix retention in ordinance (with narrow, logged exceptions); ban function creep absent a public vote; require vendor terms to mirror public policy (no NDA overrides).

5.2 Error, bias, and false positives at scale

At protest scale, even low error rates yield **absolute harm**. A 1% false-positive rate on 50,000 plates is 500 bad leads; a 95% accurate re-ID across millions of frames still mislabels thousands of faces. Errors propagate through fusion: a single wrong edge can reframe an entire subgraph as “organized.”

Bias arrives by three routes:

- **Data bias.** Historical arrests, patrol density, and complaint records reflect past policy; when fed into models, they **amplify** it.
- **Sensor geography.** ALPR grids and cameras are not evenly distributed; over-instrumented neighborhoods produce “more intelligence,” begetting more attention.
- **Interpretation bias.** OSINT scoring can overweight incendiary language or misread dialect/sarcasm, pushing benign speech up the risk stack.

Controls must match the stack: sensor-level **calibration and drift checks**; fusion-level **counterfactual tests** (does the conclusion hold if we remove any single data source?); and case-level **two-sensor corroboration** before consequential action. Publish **confusion matrices** (TP/FP/FN/TN) for each capability; require **bias and disparate impact audits**; tie procurement bonuses to **measured error reduction** rather than feature growth.

5.3 Chilling effects on speech and assembly

The First Amendment presumes a **zone of risk** where citizens can be seen dissenting without acquiring a dossier. Pervasive measurement collapses that margin. People do not need to be named in a report to feel modeled; knowing that plates, phones, and faces can be fused in hours shifts behavior in minutes: leave the car at home, skip the rally, avoid the chant, never bring a sign. That is the **chill**—not a gag order, but a prediction about consequences.

Three mechanisms drive the chill:

- **Anticipated permanence.** If a march becomes a searchable life event, spontaneity declines.
- **Social spillover.** Fusion links companions; risk is shared with family and co-workers who never chose it.
- **Asymmetric transparency.** Citizens are visible in high definition; the system that sees them is contract-redacted and log-inaccessible.

Countermeasures are design choices, not slogans: hard **retention caps** (measured in days, not years) for bulk telemetry; **no-go uses** codified in law (e.g., forbidding protest participation data for employment, housing, immigration, or benefits decisions); **event-specific transparency** (what was used, where, by whom, and when, published within a fixed window); **protected-class minimization** and **journalist/observer shields**; and **audit rights** for independent civil bodies with technical access.

Freedom of assembly survives not by wish, but by **friction** placed in the data lifecycle: delete by default, share by exception, unmask by warrant, act only on corroborated signal. Without those brakes, the public square becomes a training set—and the courage to enter it, a rounding error.

6. Governance & Accountability

6.1 Warrants, audits, and minimization rules

Bright-line warrants.

- **Geofence:** (i) polygon $\leq$ specific block(s); (ii) window $\leq$ 60 minutes; (iii) 2-step unmasking with judicial re-authorization; (iv) explicit exclusions for journalists, medics, legal observers; (v) **particularized nexus** to a defined offense; (vi) automatic public notice within a fixed post-event window (redactions allowed).
- **Cell-site simulators:** warrant by default; **metadata-only** configuration; target list pre-approved; pen-register style minimization; device IDs for non-targets **discarded on capture**.

Minimization that actually minimizes.

- **Retention caps** measured in **days** for bulk telemetry (ALPR, CSS sweeps, citywide drone video not tied to a case).
- **Two-sensor corroboration** before any identity unmasking or enforcement action.

- **Purpose binding:** data collected for crowd safety cannot be repurposed for immigration, employment, housing, or benefits decisions absent a new warrant and public vote.

Audits that bite.

- **Immutable, append-only logs** (cryptographically chained) for every query, export, and configuration change; monthly **independent log review** with random spot checks.
- **Confusion matrices** and error budgets per capability (TP/FP/FN/TN), published quarterly; if false-positive rate exceeds the budget, automatic **use throttle** triggers.
- **Suppression triggers:** undisclosed use of CSS/geofence or violation of minimization → mandatory notice to court/defense; presumptive suppression and administrative discipline.
- **Sunset + reauthorization:** every high-risk capability expires unless re-approved after an audit and public hearing.

6.2 Procurement transparency and public oversight

Contracting with eyes open.

- **No-NDA on core terms:** capabilities, retention, sharing, and audit interfaces may not be shielded as “trade secrets.”
- **Capability inventories:** vendors must provide a **feature bill of materials** (FBOM) listing every toggle (e.g., re-ID, thermal overlays, cross-jurisdiction search). Any new data type or feature requires **council approval** before activation.
- **Data governance rider** (attach to every contract): mirrors statutory retention, logging, access controls, and breach notification; city owns logs and model outputs derived from city data.
- **PIA/DPIA** (privacy/impact assessments) **before** deployment, not after—published in plain language with risk mitigations and alternatives considered.

Public windows into black boxes.

- **Transparency portal:** near-real-time dashboards reporting counts of warrants, geofences, CSS activations, ALPR queries, drone flight hours, retention deletions, and cross-agency data shares.
- **Use policies by address:** map where sensors are placed, with justification and renewal dates.

- **Community review board** with technical advisors empowered to inspect systems, request red-team runs, and recommend suspensions.

Governance of models, not just machines.

- **Model cards** for any scoring/re-ID: training sources, intended use, known failure modes, fairness tests, threshold rationales.
- **Drift monitoring**: when error or disparity drifts past guardrails, system auto-degrades to manual review mode.

6.3 Independent testing and red-team evaluations

Test like it's going to court.

- **Scenario batteries**: annually simulate protest-day loads (density, motion, weather, occlusion) and measure latency, accuracy, and correlation error across the fusion stack.
- **Counterfactual checks**: remove one sensor stream at a time (ALPR, CSS, OSINT, drone video) and verify that key conclusions still hold; if not, flag **fragile inferences**.

Red-team, blue-team, purple-team.

- **Red-team** (external, rotating): attempt to induce false associations (co-presence traps), exploit model brittleness (sarcasm/dialect in OSINT), and trigger identity leakage via configuration changes.
- **Blue-team** (operators): document mitigations, playbook adjustments, and escalation thresholds.
- **Purple-team** review: jointly produce a **Lessons & Limits** report—what the system can't reliably do.

Kill-switches and fail-safes.

- **Capability kill-switch** tied to metrics (e.g., FP rate, audit violations, unauthorized feature enablement).
- **Event-scoped deployments**: time-boxed configurations that automatically revert; any persistence requires re-authorization.
- **Third-party certification**: annual conformity assessment to a public standard (logging, minimization, accuracy, bias), with certificates and failures published.

Civic participation in testing.

- **Observer corridors** during drills (press, civil society, technologists) with controlled access to telemetry and after-action briefs.
- **Bug-bounty-style programs** for policy: reward documented transparency gaps, misconfigurations, or unlawful data paths.

Governance heuristic: *If a dashboard can do it in one click, the law must say when that click is allowed, how it is logged, how long its outputs live, and who outside the room can verify that the click was justified.* Without that symmetry, instrumentation becomes policy by procurement—and the public square becomes a product feature.

7. Citizen & City Countermeasures

7.1 Policy: geofence limits, ALPR retention caps, DFR guardrails

Geofence limits (ordinance-grade):

- **Strict scope:** polygon $\leq$ 1–2 city blocks; window $\leq$ 60 minutes; offense-specific nexus stated in the affidavit.
- **Two-step unmasking:** anonymized device IDs first; individual unmasking requires separate judicial sign-off.
- **Protected-role carve-outs:** presumptive exclusion for credentialed press, legal observers, medics, and houses of worship unless a judge makes a particularized finding.
- **Mandatory disclosure:** post-event public report within 30 days listing counts (requests, approvals, denials, devices returned, devices unmasked).

ALPR retention caps (dials that matter):

- **Default deletion:** $\leq$ 30 days for non-hit reads; $\leq$ 24 hours for shared vendor “patrol” networks unless tied to a case number.
- **Hit validation:** human confirmation before stop; two-source corroboration for arrests.
- **Sharing fence:** no export to non-local or commercial exchanges absent council approval and public notice; no use for employment, housing, school discipline, or civil code enforcement.

- **Quarterly stats:** reads, hits, false hits, retention deletions, external requests granted/denied.

DFR (Drone-as-First-Responder) guardrails:

- **Event scoping:** geofenced flight boxes with auto-expiry; no persistent orbiting of assemblies without exigency and supervisor approval.
- **No private-space peering:** bar angles that intrude into curtilage/interior absent warrant.
- **Capability transparency:** publish payloads (thermal, zoom ranges), re-ID toggles, and storage durations; auto-delete non-evidentiary footage ≤ 7 days.
- **Public flight logs:** date/time, purpose, coordinates (rounded), duration, retention disposition.

7.2 Practice: notice requirements, protest-tech transparency

Before the event:

- **Declared tool list:** city publishes (and posts on-site) which systems may be used (ALPR zones, DFR launch areas, CSS warrant policy, geofence policy) with retention and sharing rules in plain language.
- **Sensor maps:** web map of fixed ALPR/camera locations, last audit date, and responsible agency contact.
- **Brief the briefers:** train PIOs and precinct liaisons on lawful answers about tech; set a single “facts & logs” portal for real-time corrections.

During the event:

- **On-site signage:** “Instrumented Area” boards at entry points: systems in use, retention clocks, oversight contacts, and QR to live policy.
- **Observer lanes:** clearly marked corridors for press/legal observers with access to a same-day liaison and post-event evidence request process.
- **Contact cards:** distributed by volunteers: rights summary, request templates, and where to send footage/logs.

After the event:

- **48-hour transparency packet:** sensor hours flown, ALPR reads/hits, warrants obtained (types/counts), minimization outcomes (deletions, denials), and any data shares.

- **After-action forum:** recorded public debrief with logs, confusion matrices (where applicable), and a calendar for remedial votes.

7.3 Civil tools: detection, FOIA kits, community spotter guides

Detection & verification (community deployable):

- **RF & airspace awareness:**
 - **ADS-B/Remote ID** receivers to log cooperative aircraft and compliant drones; publish live maps for observers.
 - **SDR kits (software-defined radio)** to capture broad RF spectra near protest routes for later correlation (note: lawful use varies; avoid interception of content).
- **IMSI-catcher suspicion indicators** (imperfect but useful): sudden, localized drops to 2G/3G, repeated attach/detach cycles, identical cell IDs across multiple devices. Pair with **cell-monitoring apps** for anomaly logs; treat as *signals for inquiry*, not proof.
- **Network hygiene:** community Wi-Fi probe monitors to surface mass device scans; public guidance on disabling ad-ID, Bluetooth, and Wi-Fi auto-join during events.

FOIA/Public Records Kits (ready-to-send templates):

- **Rapid records bundle** (one page per system):
 - **ALPR:** vendor, camera list & locations, data schema, retention policy, audit logs (query/export events), hotlist sources, external shares in the last 90 days.
 - **DFR:** flight logs, payload configs, storage settings, operator training, policy deviations.
 - **CSS:** warrant copies (redacted), minimization/deletion attestations, activation logs.
 - **Geofence:** applications, approvals/denials, minimization steps, unmask counts, provider responses.
 - **OSINT/Fusion:** data sources, model cards (intended use, limits), access control lists, vendor NDAs.
- **Response tracker:** shared spreadsheet with due dates, statutory clocks, appeal language, and escalation contacts (AG/Open Records ombuds).

Community spotter guides (one-page field sheets):

- **Visual IDs:** common drone airframes/payloads, ALPR pole/camera form factors, portable mast trailers, and likely command-post vehicles.
- **Behavior cues:** circling vehicles without plates, persistent overhead orbits, sudden camera re-aims at chant inflection points.
- **Evidence capture SOP:** shoot wide then tight, time-stamp, note cross-streets, preserve originals, and mirror to a trusted repository.
- **Incident taxonomy:** standardized tags (e.g., “ALPR-read,” “Drone-orbit-thermal,” “CSS-suspected-RF”) to make community logs searchable and analyzable.

Citizen data rights & safety practices:

- **Data subject requests** (where available): step-by-step to request your ALPR reads, drone footage references, and any geofence inclusion notifications.
- **Consent layers:** guidance for blurring bystanders in citizen-posted videos while retaining audit utility; retain originals under restricted access for legal use.
- **Mutual aid evidence cells:** small trained teams to ingest, hash, and catalog community media; maintain chain-of-custody and produce case-ready packets.

Metrics that keep everyone honest:

- **Deletion rate** (non-evidentiary items actually purged on schedule).
- **Share ratio** (external requests granted/denied).
- **Error ledger** (documented false positives per 1k queries/alerts).
- **Chill indicators** (surveyed self-censorship vs. instrumentation intensity by neighborhood).
- **Audit closure time** (days from finding to fix).

Countermeasure motto: *If a capability is easy to turn on, it must be easy to see, limited by default, and simple to turn off.* Policy sets the dials, practice broadcasts them, and civil tools verify they were honored.

8. Research Agenda

8.1 Measuring protest tech's deterrence effect

Question. How much do instrumentation and fusion reduce lawful participation (the chill) relative to safety gains?

Designs.

- **Difference-in-differences (DiD).** Compare turnout and demographic mix across paired cities before/after policy dials (e.g., ALPR retention cut from 24 months to 30 days; geofence moratoria). Control for weather, issue salience, and permits.
- **Instrumental variables (IV).** Use exogenous shocks—e.g., sudden vendor outages, court injunctions, or drone grounding due to weather—as instruments for “surveillance intensity” to estimate causal effects on attendance and route selection.
- **Event-level conjoint surveys.** In pre-event polls, randomize visible tech conditions (signage: “DFR in use”, “ALPR zones”, “no geofences”) to measure willingness to attend, to speak, to carry signs, or to livestream.
- **Longitudinal panels.** Follow cohorts of would-be participants across multiple events to track evolving risk perceptions, self-censorship, and substitution (e.g., online speech for street presence).

Measures. Turnout elasticity vs. declared instrumentation; chant volume and banner density as expressive proxies; media self-identification rates; journalist/observer presence; arrests per 1k participants; post-event survey indices of “regret for non-attendance.”

Ethics. Pre-register designs; minimize collection of identifying data; publish nulls; commit to open materials and code.

8.2 Auditing fusion accuracy and harm

Question. How accurate are end-to-end inferences (plate ↔ phone ↔ face ↔ person ↔ role), and where do errors create material harm?

Method suite.

- **Ground-truth batteries.** Curated, consented test sets spanning varied lighting, density, occlusion, and demographics. Include “confounder scenes” (shared Wi-Fi, rideshare hubs, uniform clothing) to probe spurious edges.
- **Counterfactual stress tests.** Remove one stream at a time (ALPR, CSS, OSINT, drone video) and recompute conclusions. Flag any case where removal flips the outcome: these are **fragile inferences**.

- **Confusion accounting.** Publish confusion matrices at each layer (sensor, linker, fusion) and **propagate** uncertainty to case recommendations; track **harm-weighted false positives** (e.g., mis-ID leading to detention counts more than a mis-ranked rumor).
- **Error-to-action tracing.** For each adverse incident, reconstruct the data lineage: which query, which threshold, which feature toggle. Require a **post-harm causal chain** report.
- **Fairness & disparity audits.** Measure error and enforcement rates by location and protected classes; attribute disparity to sensor geography vs. model behavior using mediation analysis.
Governance hooks. Tie operational throttles to pre-set error budgets; if FP rate > budget, capability auto-degrades to human-in-the-loop. Mandate third-party replication of audits with access to redacted logs.

8.3 Building privacy-preserving public safety

Goal. Maintain situational awareness while sharply reducing identification and long-memory risks.

Architectures.

- **Edge-first, ID-last.** Run detection (crowd density, flow, hazards) **on-device/at the edge**; transmit only aggregate signals. Unmask identities only under warrant or exigency with automatic expiry.
- **Aggregates, not actors.** Prefer heatmaps, occupancy, and flow vectors over re-ID; publish open APIs for **aggregate telemetry** so civil groups can verify narrative claims about safety/resource allocation without exposing individuals.
- **Ephemeral by default.** Rolling ring buffers (e.g., 24–72 hours) for non-evidentiary video; cryptographically enforced deletion with public proofs-of-deletion; separate evidentiary vaults keyed to case numbers.
- **Policy-embedded cryptography.**
 - **Differential privacy** for crowd counts and movement summaries (noise-adding mechanisms with posted epsilons).
 - **Secure enclaves / MPC** for cross-agency queries that return only yes/no risk checks without exposing raw records.
 - **Zero-knowledge gates** for warrant compliance (prove warrant scope without revealing non-responsive data).

- **Capable but constrained re-ID.** Keep person/plate re-ID models **air-gapped** and disabled by default; activation requires dual-key approval (judicial + supervisory) and writes an immutable log entry visible to external auditors.
- **Transparency-first UX.** Public signage and live policy pages show active dials (retention clocks, features on/off). After each event, auto-publish a **telemetry bill**: what signals were processed, how long they lived, what was deleted, and what was escalated to evidence.

Benchmarks & proofs.

- **Utility curves:** accuracy of hazard detection vs. privacy budget (epsilon) to quantify the safety-privacy frontier.
- **Latency budgets:** maximum acceptable delay for edge aggregation that still supports de-escalation.
- **Deletion SLOs:** % of non-evidentiary items purged within the window; public monthly proofs.
- **Adversarial trials:** red-team attempts to re-identify individuals from published aggregates; publish outcomes.

Deliverables. Open reference implementations (edge analytics, DP counters, logging), a **Model Card for Public Safety** template (intended uses, limits, failure modes), and a municipal **Privacy Tech Baseline** certification cities can adopt and vendors must meet.

Research heuristic: Measure what matters (safety and freedom), expose where inference breaks, and prototype architectures where **capability ≠ compulsion**. If we can show that privacy-preserving pipelines keep people safer *and* freer, the default dial can move away from identification without starving the incident command of light.

9. Conclusion: Freedom After Instrumentation

9.1 Keeping safety without total recall

We do not have to choose between darkness and omniscience. The middle way is architectural, not rhetorical: **see enough, forget on time, prove both**. Safety comes from fast, local signal—crowd flow, hazard detection, de-escalation cues—not from indefinite dossiers. When retention clocks are short, sharing is fenced, and unmasking is rare and logged, instrumentation serves the incident rather than feeding an institution. Put differently: let sensors describe **events**, not

people. Make the exceptional path (warrant, exigency) real and auditable; make everything else ephemeral by default. If a city can publish deletion proofs, error ledgers, and post-event telemetry bills as easily as it publishes crime stats, trust becomes measurable rather than mythical. We keep safety not by hoarding memory, but by engineering **useful amnesia**—enough to act now, and enough to prevent “act now” from becoming “know everyone forever.”

9.2 Restoring a citizen’s right to risk being seen

Democracy requires a zone where visibility is brave, not incriminating. The right to assemble has always implied a tolerated ambiguity—*I was there* cannot, by itself, be a case file. To restore that space under modern instrumentation, we must invert the burden: the state must **earn the name** before it records the person. Edge-first analytics, warrant-gated identity, short clocks, transparent logs—these are not luxuries; they are the scaffolds of consent. Pair them with civic countermeasures—notice, observer corridors, after-action disclosures—and the street becomes legible **without** becoming a ledger. A free city is not one without dashboards; it is one where a citizen can step into the crowd, speak in the open, and know that the systems watching for danger are **not** quietly drafting their biography. The courage to be seen is part of public safety. Our task is to design the machines—and the laws—so that courage survives.

References

Cell-Site Simulators (StingRay/Dirtbox) & Policy

- U.S. Department of Justice. *Use of Cell-Site Simulator Technology* (Policy Guidance), Sept. 3, 2015. [Department of Justice](#)
- U.S. Department of Justice. *Use of Cell-Site Simulator Technology (Interim Update)*, Jan. 12, 2024. (Warrant default; metadata-only/pen-register configuration.) [Department of Justice](#)
- U.S. Department of Homeland Security. *Department Policy Regarding the Use of Cell-Site Simulator Technology*, Oct. 19, 2015. [Department of Homeland Security](#)
- The Wall Street Journal. “Americans’ Cellphones Targeted in Secret U.S. Spy Program,” Nov. 13, 2014 (airborne “dirtbox” reporting); see also follow-up Nov. 14, 2014. [The Wall Street Journal+1](#)
- TIME. “Report: Feds Using Airplanes to Target Criminal Suspects’ Cell-Phone Data,” Nov. 14, 2014; “Is the Government’s Aerial Smartphone Surveillance Program Legal?,” Nov. 15, 2014. [TIME+1](#)
- U.S. House Oversight Committee. *Bipartisan Cell-Site Simulator Report* (excerpts on historical practices), Dec. 19, 2016. [Oversight Committee](#)

Aerial/Wide-Area Surveillance

- *Leaders of a Beautiful Struggle v. Baltimore Police Department*, 4th Cir. (en banc), June 24, 2021 (striking down warrantless citywide plane surveillance). Opinion PDF. [Fourth Circuit Court](#)
- EFF. “Fourth Circuit Ruling ... Baltimore Police Department” (case explainer & documents). [Electronic Frontier Foundation](#)

Geofence Warrants (Legal Split)

- *United States v. Smith*, 110 F.4th 817 (5th Cir. 2024) (categorically unconstitutional). Opinion PDF & analyses. [Fifth Circuit Court+2WilmerHale+2](#)
- *United States v. Chatrue*, No. 22-4489 (4th Cir. en banc, Apr. 30, 2025) (affirmed use with fractured reasoning). Opinion PDF & summaries. [Maryland State Bar Association+3Fourth Circuit Court+3Bloomberg Law+3](#)

Palantir / Graph-Fusion Platforms

- Palantir. “Gotham—Platform Overview.” (Capabilities and mission framing.) [Palantir](#)
- Palantir Docs. “Gotham Security Overview.” (Access controls, governance mechanisms.) [Palantir](#)
- WIRED. “What Does Palantir Actually Do?” (context on Gotham vs. data brokering). [WIRED](#)

OSINT/SOCMINT Platforms in Practice

- Brennan Center for Justice. “Documents Reveal How DC Police Surveil Social Media Profiles and Protest Activity,” Apr. 30, 2024. [Brennan Center for Justice](#)
- StateScoop. “Over decade of D.C. protests, police scanned social media ... Dataminr,” May 1, 2024. [StateScoop](#)
- Wisconsin Examiner. “Milwaukee PD used Babel Street around RNC,” Nov. 20, 2024. [Wisconsin Examiner](#)
- Babel Street (product pages & LE use cases). [Babel Street+2Babel Street+2](#)
- LifeRaft (Navigator/OSINT product & blogs). [liferaftlabs.com+3liferaftlabs.com+3liferaftlabs.com+3](#)
- Business Insider. “Dataminr ... surveillance contract for the White House,” Oct. 12, 2022. [Business Insider](#)

ALPR Networks & Vendors

- Flock Safety. *Evidence/Retention Policy* (standard 30-day retention). [Flock Safety](#)
- Motorola Solutions. Press Release on acquisition of VaaS (Vigilant Solutions), Jan. 7, 2019; see also GovTech/TechCrunch coverage. [Motorola Solutions+2GovTech+2](#)
- Genetec AutoVu (product overview & resources). [Genetec+2Genetec Resources+2](#)
- City policy examples (30-day retention): Winston-Salem, NC; Dubuque, IA. [City of Winston-Salem+1](#)

Drone/DFR Programs

- City of Chula Vista Police Department—UAS/DFR Program. (Program description & milestones/BVLOS.) [Chula Vista+1](#)
- MITRE. *Drone as First Responder Programs* (overview & regulatory context), 2023. [MITRE](#)
- EFF. “DFR Programs Are the Latest Aerial Police Surveillance Push,” June 27, 2024. [Electronic Frontier Foundation](#)
- Skydio (DFR approvals and case studies incl. NYPD BVLOS news & municipal deployments). [Skydio+2Skydio+2](#)

Background & Commentary on StingRay/Dirtbox Technology

- Michigan Law Review. “Spies in the Skies: Dirtboxes and Airplane Electronic Surveillance,” 2015. [Michigan Law Review](#)
- ACLU. “ACLU Seeks Information About Airborne Cell Phone Snooping,” Nov. 19, 2014. [aclu.org](#)

(Where possible, official documents, court opinions, and vendor policies are prioritized; journalism and NGO reports provide additional context on deployment and public policy implications.)