PITTSBURGH AI POLICY HACKATHON

POLICY BRIEF / ALGORITHMIC SURVEILLANCE ACCOUNTABILITY ACT

The CoT Mirror.

Capturing human reasoning in AI surveillance — to eliminate racial profiling and sharpen objective threat detection.

RIZALDY UTOMO · BRACKET 2 FINALIST

bracket 2 · civil liberties

posner hall a35 · apr 19

07:00 · final pitch

SCENE · JFK INTERNATIONAL · TERMINAL 4

“

You step off a flight. Before you reach the gate agent, a camera has already captured your face, matched it against a government database, and cleared or flagged you.

no officer

no warrant

no suspicion required

CBP FACIAL COMPARISON · DEPLOYED TODAY

The Fourth Amendment was not written for this scale.

238

AIRPORTS

running facial comparison

39

SEAPORTS

plus land border expansion

75 yrs

RETENTION

non-citizen photos kept

Old border rules assumed friction — officers, time, limited scope. AI removes all three.

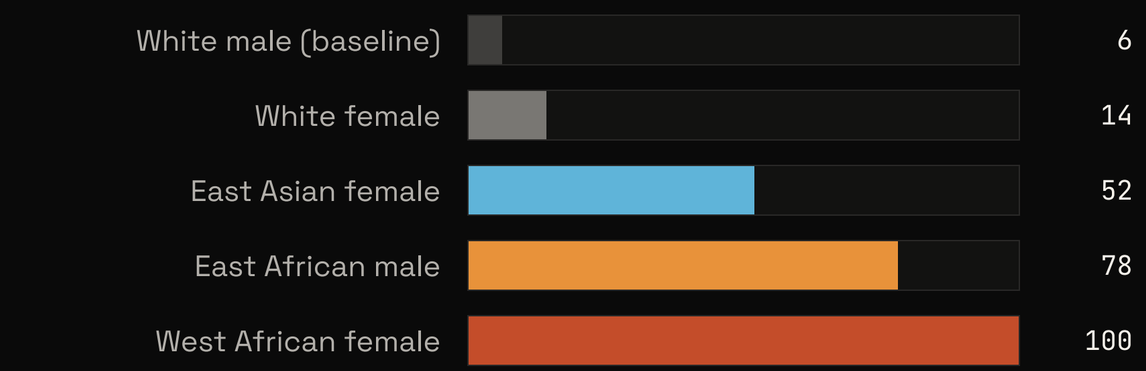
NIST FRVT · 189 ALGORITHMS · 18M IMAGES

10–100×

False-positive rates vary by this factor across demographic groups.

Highest error rates: West African, East African, East Asian faces.

RELATIVE FALSE-POSITIVE RATE



Directional illustration based on NIST FRVT Part 3 (NISTIR 8280).

Every known wrongful arrest victim is Black. Six arrests. All of them.

CASE · DETROIT · 2020

Robert Williams

Arrested on his front lawn in front of his two daughters after a facial-recognition match.

CASE · DETROIT · 2023

Porcha Woodruff

Eight months pregnant. Held 11 hours. Matched to an eight-year-old mugshot.

DETECTIVE, TO WILLIAMS: ***“The computer says it’s you.”***

THE PROBLEM IS NOT JUST A BIASED ALGORITHM.

It is an unaccountable human who clicks “approve” without leaving a trace of reasoning.

WHY “HUMAN IN THE LOOP” IS CEREMONIAL

We can audit the machine. We cannot audit the person who pulls the trigger.

SKITKA, MOSIER & BURDICK · 1999

~100%

error rate when officers follow wrong AI recommendations — even against their own judgment.

GREEN · 2022 · 41 GOVT ALGORITHMS

“Human oversight policies provide a false sense of security.”

The click is logged. The reasoning is not.

CHAIN-OF-THOUGHT · MACHINE VS HUMAN

Only one side of the decision leaves a reasoning trail.

INSPECTABLE

The Machine

- Confidence score
- Candidate ranking
- Model version
- Image-quality assessment
- Watch-list source

VS

OPAQUE

The Reviewer

- ~~What they weighed~~
- ~~Contradicting evidence~~
- ~~Why they approved~~
- ~~Override rationale~~
- ~~— nothing —~~

ALGORITHMIC SURVEILLANCE ACCOUNTABILITY ACT

Three pillars.

01 define what's allowed

02 who watches the watchers

03 the CoT Mirror

DRAWING THE LINE CONGRESS HAS NEVER DRAWN

Four tiers. Defined predicate, judicial approval, measurable floor.

TIER 0	Prohibited	Live mass biometric ID in public spaces. No exceptions.
TIER 1	Warrant Required	Targeted ID of a named person. Federal court warrant, probable cause.
TIER 2	Restricted Border Use	1:1 only. False-positive rate <=0.1% per group. Error gap <=1.5x across groups.
TIER 3	National Security	Court-approved. 30-day cap. No protest monitoring. No ICE use.

← PROHIBITED

SCOPE OF PERMITTED USE

NARROWLY EXEMPTED →

BIDIRECTIONAL ACCOUNTABILITY LOGGING

If the model’s reasoning is inspectable, the reviewer’s should be too.

A structured log entry — confidence score, override rationale, visual differences, contrary evidence — turns the “approve” click into an auditable decision.

CBP REVIEWER DECISION PANEL

Both sides. Same audit trail.

jfk t4 | 2026-04-08 14:32 utc | reviewer · redacted

SYSTEM OUTPUTS — MACHINE COT

CONFIDENCE	72.3%
CANDIDATE RANK	3 of 8
MODEL	TVS-FR v3.2.1
SOURCE LIST	government watchlist / visa overstay
IMAGE QUALITY	0.67 · low light
THRESHOLD ALERT	human review required

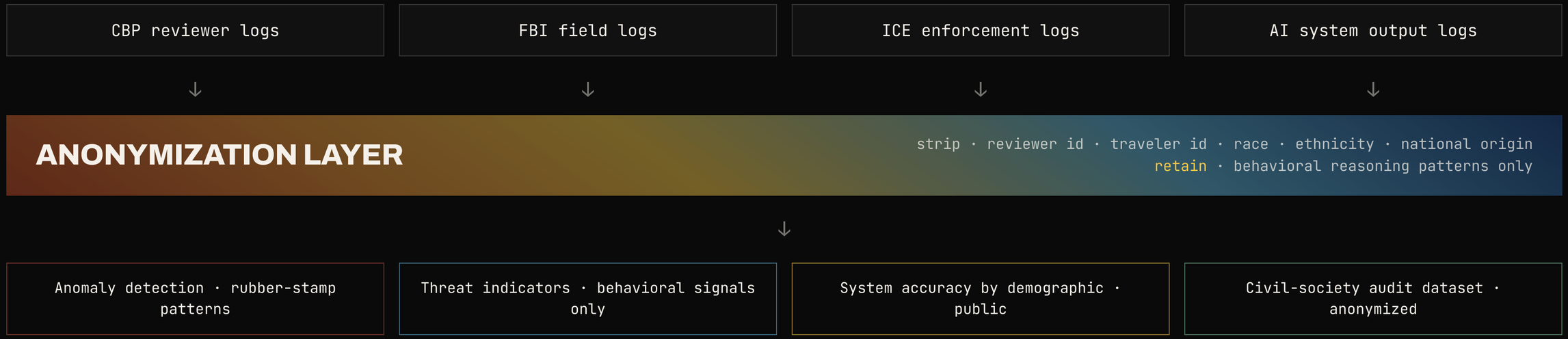
REVIEWER JUSTIFICATION — HUMAN COT · NEW

DECISION	Reject match
VISUAL DIFFERENCES	scar absent; weight mismatch
CONTRARY EVIDENCE	low-light image; third-ranked
RATIONALE	features do not correspond
NEXT ACTION	traveler cleared
SIGNED	cred-check ✓ · tamper-proof ID: 0x7a…e4c

AUDIT · ACCESS watchdog · GAO · IG only · public: quarterly · anonymized · retention: 7 years · cannot be used to penalize officers

FIGURE 03 · BEHAVIORAL INTENT, NOT DEMOGRAPHIC PROXY

The same logs that catch wrongful arrests also catch real threats.



Security and rights come from the same architecture — replace demographic proxies with documented intent.

EVERY ACCOUNTABILITY REFORM HAS REQUIRED A PAPER TRAIL

The surveillance officer is the last high-stakes federal decision-maker without one.

Pilot	FLIGHT DATA RECORDER	FAA requires pilots to log 88 flight parameters.
Police	BODY-WORN CAMERA	Mandatory capture · defined retention · tamper-proof audit.
Judge	WRITTEN OPINION	Reasoning becomes the record appellate courts review.

Surveillance officer COT MIRROR → Structured reasoning fields · compartmented access · 7-year retention.

THE ASK

AI already explains itself. We owe the same standard to the humans who act on it.

A tiered boundary. A forced sunset. The CoT Mirror.
The tools Congress needs to draw the line before the next wrongful arrest — not after.