

Product Requirements Document

HumanityAI'd Patrol Sight — v1.1.7

July 2026

CONFIDENTIAL — HUMANITYAI'D

1. Product overview

Patrol Sight gives a security officer on foot the recognition capability that until now only a fixed camera and a control room had. The officer wears lightweight AR glasses; the glasses stream what the officer is looking at to an **on-premises GPU server**, which detects and identifies faces against the organisation's own watchlist and returns a colour-coded label into the officer's lens — **in about 68 milliseconds**, before the person has walked past.

Everything — detection, recognition, the watchlist, the audit trail and the analytics — runs **inside the customer's network**. No face image, faceprint or alert leaves the building. In a region where biometric data is classed as sensitive personal data and in-country processing is the legal default, that is not a feature; it is the entry requirement.

2. Problem statement

Today's pain	Consequence
Recognition lives on fixed cameras; the officer on the ground has none	The control room identifies someone the officer already walked past
Officers carry phones or tablets to check a face	Hands occupied, eyes down, and the check happens after the encounter
Watchlist checks depend on an officer remembering a briefing photo	Recognition accuracy varies by shift, fatigue and crowd density
Cloud face-recognition services cannot be used for regulated biometric data	Procurement stalls, or the capability is simply never deployed
Where wearable recognition exists at all, it is a closed foreign platform	Sovereignty, supply-chain and support exposure

3. Target users & personas

- **The Patrol Officer** — needs an identification they can trust arriving in their line of sight, hands free, with a clear indication of *category* (wanted, banned, VIP, staff, missing person) rather than a raw name, and no risk of being distracted by a flickering overlay.
- **The Control-Room Operator** — needs live visibility of every deployed device, the alert stream, tracking history, and the ability to push watchlist changes to the whole fleet immediately.

- **The Watchlist Owner / Investigator** — needs to enrol and categorise people, in bulk, with provenance and an audit trail of who searched for whom.
- **The Security Director** — buys on operational outcome, officer safety, defensibility, and the ability to answer a regulator's questions.
- **The IT / Information-Security Authority** — approves on data residency, network posture, credential management, auditability and supply-chain provenance.

4. Core user journeys

1. **Patrol.** Officer collects glasses at shift start; they pair to the site's dedicated 5 GHz network and begin streaming. As people pass, faces are detected, identified and labelled in the lens with a category colour. Nothing requires a button press.
2. **Alert and verify.** A high-confidence match on a watchlisted category surfaces to the officer and simultaneously publishes to the control room with the device, timestamp and confidence. **The officer verifies the person before acting — the system informs a human decision, it never authorises one.**
3. **Watchlist lifecycle.** The watchlist owner enrolls a person (single or bulk), assigns a category and colour, and the index rebuilds and hot-reloads into the live engine without restarting anything. New entries are active for every deployed device within seconds.
4. **Review.** Control room reviews tracking history per person and per device, alert outcomes, and the audit log of searches — the record that makes the deployment defensible.

5. Functional requirements

5.1 Officer experience (glasses)

- F-1 Hands-free continuous face detection and identification from the live camera stream; **target < 150 ms p95 from face-in-view to label-in-lens** (measured: ~68 ms typical, 49-103 ms range)
- F-2 In-lens label showing name and **category colour** via a compact encoded format; category-only display mode where names must not be shown to the wearer
- F-3 Stable overlays — temporal smoothing, stale-label eviction and duplicate-track suppression so labels do not flicker or double
- F-4 Recognition envelope stated and enforced: reliable to ~5 m, maximum 6-7 m in good light, long-range mode for 8-10 m at reduced frame rate; faces below the minimum pixel size are **rejected, not guessed**
- F-5 Robust to the real world: off-angle faces to 90° yaw, and a link that survives brief wireless interruption without ending the session
- F-6 One-command device provisioning; per-device session identity

5.2 Recognition engine

- F-7 SCRFD face detection + ArcFace 512-dimension embeddings executed on TensorRT FP16
- F-8 GPU FAISS IVF-PQ search over the enrolled population, with adaptive search-breadth tuning and a documented similarity calibration
- F-9 Two-layer tracking (spatial + identity cache) so continuous streams do not re-search every frame
- F-10 Configurable recognition and detection thresholds per deployment, recorded per site
- F-11 Live index rebuild and hot reload on enrolment — no restart, no retraining
- F-12 Face-quality assessment to suppress unusable frames
- F-13 Backpressure by design: newest-frame-wins ingest and a frame-rate-capped worker so the GPU never accumulates a queue

5.3 Watchlist, command & control

- F-14 Enrolment of people with categories, labels and colours; bulk enrolment at **40-60 images/second**
- F-15 Control console: live device monitor (per-device frame rate, latency, face counts, annotated preview), alert stream, tracking history, per-device and per-person views
- F-16 Alert publication on the internal message bus for integration with the customer's existing command-and-control or video-management system
- F-17 Role-based operator accounts, activity logs, and an audit trail of searches and matches
- F-18 Data lifecycle controls: retention configuration, purge, and export for regulatory review

5.4 Platform & deployment

- F-19 Fully on-premises single-GPU-host deployment as one orchestrated stack; no cloud dependency in the recognition path
- F-20 One-command deploy, health, update and teardown tooling; automatic recovery on reboot including GPU-driver race handling
- F-21 All customer data on external volumes that survive teardown and upgrade
- F-22 Authenticated, encrypted internal service communication (per-service bus authentication; AES-256-GCM payloads between device hub and console)
- F-23 Privacy mode: transmit the 512-number faceprint instead of the image where imagery must not traverse the network

6. Non-functional requirements

Requirement	Target	Status
Glasses-to-label latency	< 150 ms p95	Measured ~68 ms typical, 49-103 ms; WebRTC path p50 26-32 ms, p95 63-85 ms
Effective recognition rate per stream	≥ 12 fps	Measured 14-20 fps
Watchlist capacity	100k on the reference host; 13M documented	93,728 persons in production; 13M → 80-130 ms measured
Index memory efficiency	Compressed	962 MB at 13M vectors (27× vs flat)
Off-angle accuracy	Usable to 90° yaw	R50 model: ~60% fewer false negatives at extreme pose than lightweight alternative
Enrolment throughput	≥ 30 images/s	Measured 40-60 images/s
Concurrent devices per GPU	To be measured per GPU class	Not yet published — measured during pilot (declared gap)
Availability	99.5% during operating hours	Auto-restart, boot recovery, health probes
Data residency	No egress of biometric data	On-premises by construction
Auditability	Every search and match attributable	Audit middleware + operator audit view

7. Proven reference deployment (July 2026)

Patrol Sight runs today on HumanityAI'd's own infrastructure as a live 34-service stack on a single **RTX 4060 Ti (8 GB)**, sharing that GPU with other production workloads:

- **93,728 enrolled persons** / 95,667 vectors in one GPU FAISS shard
- **~68 ms** typical glasses-to-label, stage-by-stage measured; **0% packet loss**, ~4 ms network round trip on 5 GHz WiFi 6
- **14-20 fps** effective recognition per stream; FAISS search P95 **8.5 ms** over 1,931 real searches
- RayNeo X3 Pro glasses streaming hardware-encoded H.264 at 2560×1440 / 20 Mbps
- Recognition threshold 0.50, detection 0.65, tuned on this deployment after measuring false-detection rates

8. Out of scope (v1)

- **Autonomous action on a match** — deliberately excluded. Patrol Sight informs an officer; it does not open a gate, dispatch a unit or record an enforcement decision on its own.
- Third-party watchlist connectors (national or interoperable databases) — roadmap v1.5; scoped and priced per engagement until then.

- Body-worn evidential video recording — Patrol Sight is a recognition system, not a body-camera; it integrates alongside one.
- Behaviour, weapon or anomaly detection — roadmap v2.
- Emotion or demographic inference — deliberately excluded on ethical grounds.
- Covert operation — the product assumes visible, disclosed deployment consistent with the customer's legal authority.

9. Success metrics

- Officer-verified true-positive matches per deployment week, with false-positive rate tracked and reviewed
- $\geq 95\%$ of watchlist checks resolved in the lens rather than by radio or phone
- Watchlist owner can enrol and activate a person, unaided, in under 2 minutes
- Zero biometric-data egress events (verified by network audit)
- Officer-reported confidence and comfort ratings $\geq 4/5$ over a full shift

10. Dependencies & risks

Risk	Mitigation
Model licensing — the detection/recognition weights derive from a research-licensed lineage	Resolve before any paid deployment: commercial licence or replacement weights. Tracked as the top-priority item in the Technical Due-Diligence Report
Wireless quality determines latency more than the AI does	Dedicated 5 GHz SSID, power-save disabled, site survey in the pilot; 20 Mbps per device budgeted
Plaintext glasses transport	Isolated glasses VLAN with no uplink as a contractual control until encrypted transport ships (Phase 2 hardening)
Single GPU per site; concurrent-device ceiling unpublished	Measure during pilot and commit to the measured figure; multi-GPU path pre-built
Regulatory and reputational scrutiny of facial recognition	Customer-owned watchlists, no scraped data, human verification required before action, full audit trail, support for a Data Protection Impact Assessment
Glasses hardware supply and generational change	Vendor SDK isolated; the app has already been run against more than one AR glasses platform
Accuracy variance in real conditions (lighting, pose, crowd)	Published recognition envelope, per-site threshold calibration, face-quality gating, and pilot-phase accuracy measurement in the customer's own environment