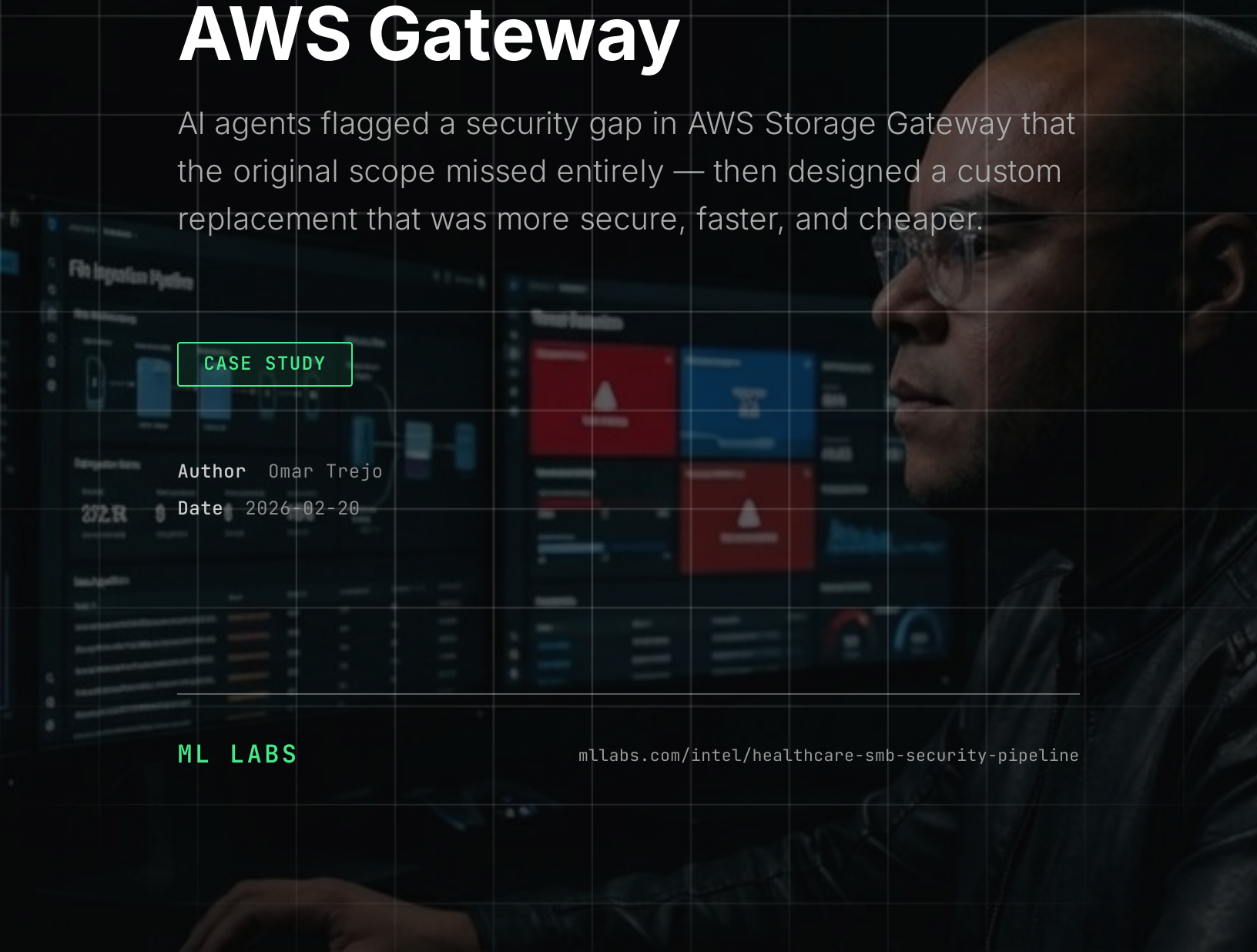
ML LABS INTELLIGENCE

Securing Healthcare File Ingestion Beyond AWS Gateway

AI agents flagged a security gap in AWS Storage Gateway that the original scope missed entirely — then designed a custom replacement that was more secure, faster, and cheaper.



CASE STUDY

Author Omar Trejo

Date 2026-02-20

A healthcare AI platform needed to ingest ECG recordings from dozens of clinic sites into a cloud processing system. The original scope called for AWS Storage Gateway — the standard managed service for bridging SMB file shares to S3 storage. It was the obvious choice until AI agents analyzing the architecture flagged a fundamental security gap that had not been part of the original requirements review.

This discovery expanded the engagement scope significantly. What started as a straightforward infrastructure task became a custom security engineering project — one that ultimately delivered a system that was more secure, 20% faster, and 73% cheaper than the AWS service it replaced. The value surfaced beyond the original scope is the reason this case study exists.

The Security Gap

The core issue is structural, not configurational. AWS Storage Gateway acts as a proxy between the SMB protocol and S3. When a clinic connects to the gateway, the gateway authenticates the SMB session and then issues S3 API calls on the clinic's behalf. But every S3 call originates from the gateway instance itself.

This creates cascading isolation failures:

- **Shared authentication surface.** All organizations connect to the same SMB gateway using guest authentication. Any organization that discovers another's share name can mount it — there is no mechanism to restrict share access by source identity at the protocol level.
- **IP-based access controls are blind.** S3 condition keys like `aws:SourceIp` and `aws:SourceVpc` resolve to the gateway's IP, not the clinic's. Bucket policies that attempt to restrict access by clinic IP address have no effect.
- **No per-organization isolation at the storage layer.** IAM roles attached to the gateway apply to all shares. Per-share IAM role mapping does not exist. Every

share routes through the same credentials to the same S3 backend.

For a platform handling protected health information across multiple healthcare organizations, this is not a theoretical concern — it is a data isolation violation waiting to happen. A single misconfigured share name exposes one organization's clinical data to another.

Alternatives Evaluated and Rejected

Before building a custom solution, ML LABS systematically evaluated every AWS-native approach to closing the isolation gap. AI agents were used to accelerate this evaluation — compressing what would have been weeks of security architecture research across AWS documentation, IAM policy semantics, and SMB protocol specifications into hours of structured analysis.

- **Per-share IAM roles.** Storage Gateway does not support per-share IAM role mapping. All shares use the gateway's single IAM role.
- **S3 Access Points.** Access Points provide per-application entry points to S3 buckets, but the source IP is still the gateway's. Access Point policies cannot distinguish between clinics.
- **Active Directory integration.** AD would provide per-user authentication on the SMB side, but adds $110-350/month in AWS Directory Service costs, requires clinic-side domain joining or credential distribution, and introduces significant operational complexity for an ingestion path that needs to be zero-touch for clinics.
- **Separate Storage Gateway per organization.** Each gateway instance costs approximately $550/month. For a platform serving dozens of organizations, this scales to tens of thousands in monthly infrastructure cost — for a file sync function.

- **AWS Transfer Family (SFTP).** Transfer Family supports per-user IAM roles and would solve the isolation problem, but requires every clinic to change from SMB to SFTP. Clinic IT teams will not change network protocols for a vendor integration.

Every managed AWS service either failed to provide per-organization isolation or required clinics to change their network configuration. The constraint was clear: the solution must speak SMB, must isolate organizations at the protocol level, and must not require clinics to change anything about their existing network setup.

Custom SMB Server Architecture

ML LABS built MI-SMB — a self-managed Samba server on EC2 that provides what AWS Storage Gateway cannot: per-organization access control enforced at the protocol level, before authentication even begins.

Protocol-Level Isolation

Each organization gets a dedicated Samba share configured with IP-based access control using Samba's `hosts allow` and `hosts deny` directives. When a connection arrives, Samba checks the source IP against the share's allowed list. If the IP doesn't match, the connection is refused before any credential exchange occurs.

This is a fundamentally different security model than gateway-proxied access:

- **Connection refused before authentication.** An unauthorized IP cannot even begin an SMB handshake with a share it doesn't belong to.
- **Per-organization credentials.** Each share has its own Samba user and password, scoped to that organization's storage only.
- **Per-organization S3 destination.** Each share syncs to an organization-specific S3 bucket, enforcing storage isolation at the infrastructure layer.

Real-Time File Sync

Storage Gateway uses a batched approach to syncing files from SMB shares to S3. Files land on the gateway's local cache and are flushed to S3 on a schedule or when cache pressure forces a write. This introduces latency between when a clinic writes a file and when it's available for AI processing.

MI-SMB uses `inotify` — the Linux kernel's file system event notification mechanism — to detect new files the instant they're written. An event-driven sync process uploads files to the organization's S3 bucket within 1-2 seconds of write completion. Clinical recordings are available for AI processing almost immediately after the clinician saves them.

Multi-Region Deployment

The platform operates in both US and UK regions to satisfy GDPR data residency requirements. MI-SMB is deployed in both regions with identical configuration management, ensuring that a UK clinic's data never transits through US infrastructure.

Monitoring and Audit

Every connection attempt, file sync event, and access denial is logged to CloudWatch. Automated alarms fire on:

- Failed connection attempts from unauthorized IPs
- Sync failures or delays exceeding thresholds
- Disk utilization approaching capacity on the SMB server
- Share configuration changes outside the deployment pipeline

The audit trail satisfies healthcare data protection requirements for access logging and monitoring, providing a complete record of which organization accessed which share, when, and from what IP address.

Performance and Cost

MI-SMB delivers measurably better performance at lower cost than the managed service it replaced:

- **20% faster data movement.** Real-time `inotify`-based sync delivers files to S3 within 1-2 seconds, compared to Storage Gateway's batched approach that introduces variable-length delays.
- **~$400/month savings per instance.** An EC2 instance running Samba costs approximately $150/month versus ~$550/month for a Storage Gateway instance — with better isolation guarantees.
- **Zero cross-organization data leakage by design.** Protocol-level IP restriction means unauthorized access is impossible, not merely improbable.

The cost differential compounds with scale. A platform serving 20 organizations needs one MI-SMB instance (with failover), not 20 separate Storage Gateways.

When Managed Services Create Security Gaps

Managed services are the right default for most infrastructure decisions. But they encode assumptions about access patterns and trust boundaries that may not match your system's actual security requirements. AWS Storage Gateway assumes that all SMB clients connecting through a single gateway can share a trust boundary. For a multi-tenant healthcare platform, that assumption creates a security gap that no IAM policy can close.

The key insight is that AI agents helped identify this gap by systematically evaluating every combination of AWS access control mechanisms against the actual isolation requirements — a task that would have required deep expertise across IAM policy semantics, SMB protocol behavior, and Storage Gateway internals. The structured evaluation compressed weeks of architecture research into hours and produced high confidence that no AWS-native solution existed before committing to a custom build.

First Steps

If your platform ingests data from multiple organizations through shared infrastructure, audit whether your isolation guarantees hold at every layer — not just at the application level, but at the protocol and network levels where managed services make assumptions about trust boundaries.

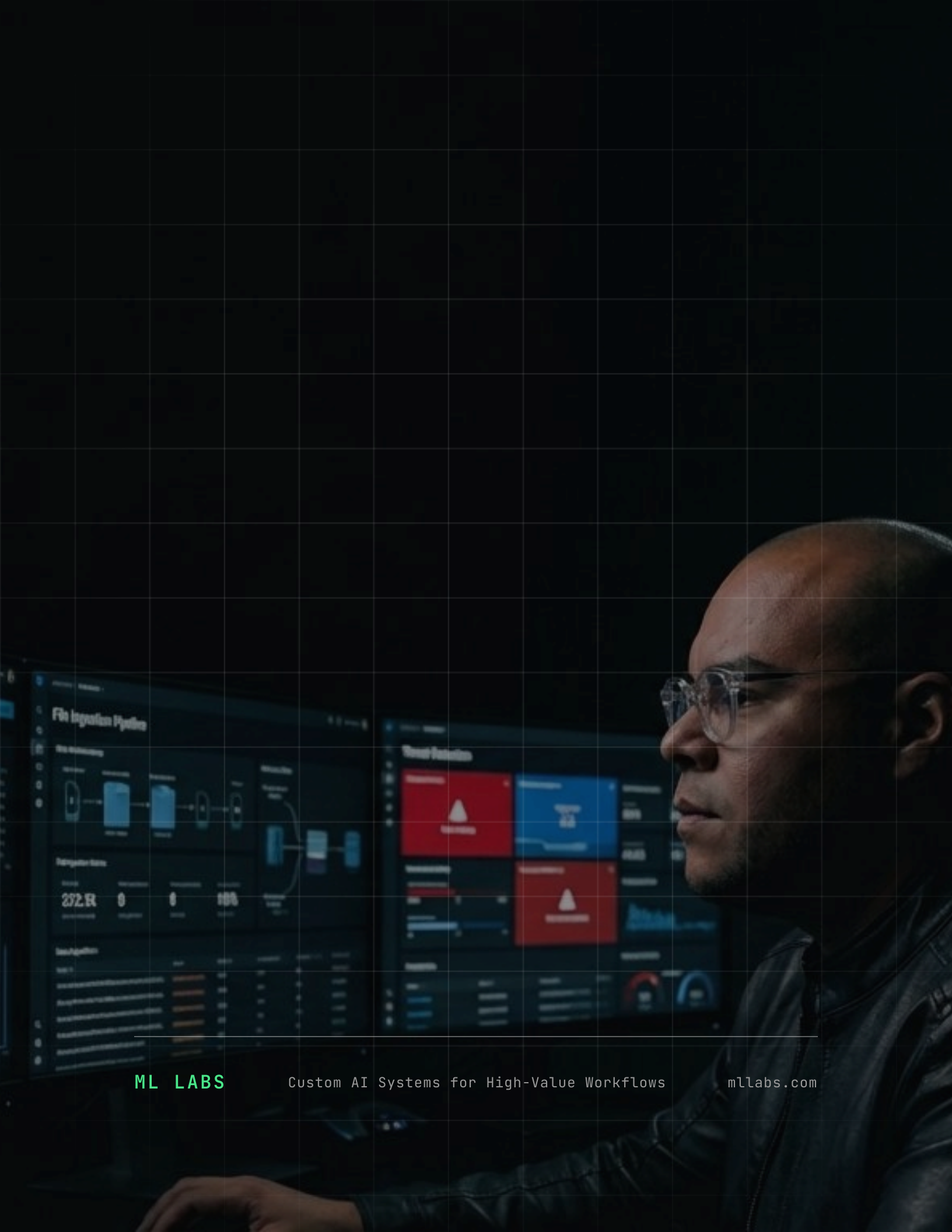
Start with the access pattern: trace a connection from the client through every proxy, gateway, and service boundary to the storage layer, and verify that the client's identity is preserved at each hop. If any layer collapses multiple clients into a single identity, that's where your isolation breaks. If your organization needs a security architecture review of existing ingestion infrastructure, a **Technical Assessment** maps the gaps. If the architecture is already scoped and needs to reach production, **AI Workflow Integration** is the direct build path.

References

1. AWS. **What Is AWS Storage Gateway?**. *AWS Documentation*, 2025.
2. AWS. **IAM Condition Context Keys**. *AWS Documentation*, 2025.
3. Samba Team. **smb.conf — The Configuration File for the Samba Suite**. *Samba*

Documentation, 2025.

4. HHS. **Healthcare Security Rule — Technical Safeguards**. *U.S. Department of Health and Human Services*, 2024.
5. NIST. **Healthcare Security Rule Guidance**. *National Institute of Standards and Technology*, 2024.
6. AWS. **Security Best Practices for Amazon S3**. *AWS Documentation*, 2025.
7. Linux man-pages. **inotify — Monitoring File System Events**. *Linux Programmer's Manual*, 2024.



ML LABS

Custom AI Systems for High-Value Workflows

mllabs.com