DocFlow — AI Governance Assessment

NIST AI RMF Alignment Report

System: Intelligent Document Processing Pipeline (DocFlow) **Version**: 1.0
Assessed: March 2026 **Assessor**: Carlandra Williams, Sr. FinOps Practitioner

---

Executive Summary

DocFlow is a serverless AWS pipeline that automates the extraction of structured data from business invoices and receipts. This assessment maps DocFlow's current governance controls against the NIST AI Risk Management Framework (AI RMF 1.0) across four functions: Govern, Map, Measure, and Manage.

Overall Assessment: Partially Aligned

DocFlow demonstrates strong operational governance practices — cost controls, resource tagging, audit trails, and documented limitations. It has material gaps in areas specific to AI governance: bias evaluation, data provenance, explainability documentation, and formal risk categorization.

NIST Function	Status	Notes
Govern	🟡 Partial	Accountability and tagging in place; no formal AI governance policy
Map	🟡 Partial	Use case and limitations documented; no formal risk categorization
Measure	🟡 Partial	Performance metrics tracked; no bias or fairness evaluation
Manage	🟢 Mostly Aligned	Controls, guardrails, and remediation paths in place

AIRC Core Subcategory Mapping

This table maps each identified gap to its corresponding NIST AI RMF 1.0 Core subcategory. Subcategory IDs reference Tables 1–4 of the AI RMF Core (AIRC).

Gap ID	Gap Description	AIRC Subcategory	Subcategory Description
GOV-01	No formal AI governance policy	Govern 1.1	Legal and regulatory requirements involving AI are understood, managed, and documented.
GOV-01	No formal AI governance policy	Govern 1.4	The risk management process and its outcomes are established through transparent policies, procedures, and other controls.
GOV-02	No designated AI risk owner	Govern 2.1	Roles and responsibilities and lines of communication related to mapping, measuring, and managing AI risks are documented and clear.
GOV-02	No designated AI risk owner	Govern 2.3	Executive leadership takes responsibility for decisions about risks associated with AI system development and deployment.
GOV-03	No user communication about AI involvement	Govern 4.2	Organizational teams document the risks and potential impacts of the AI technology they deploy and communicate about the impacts more broadly.
GOV-04	No incident response plan for AI failures	Govern 4.3	Organizational practices are in place to enable AI testing, identification of incidents, and information sharing.
GOV-05	No third-party audit	Govern 6.1	Policies and procedures are in place that address AI risks associated with third-party entities.
MAP-01	No formal risk categorization	Map 1.5	Organizational risk tolerances are determined and documented.
MAP-01	No formal risk categorization	Map 2.1	The specific tasks and methods used to implement the tasks that the AI system will support are defined.
MAP-02	No training data provenance	Map 2.3	Scientific integrity and TEVV considerations are identified and documented, including those related to data collection and selection.

Gap ID	Gap Description	AIRC Subcategory	Subcategory Description
MAP-02	No training data provenance	Map 4.1	Approaches for mapping AI technology and legal risks of its components — including use of third-party data or software — are in place and documented.
MAP-03	No stakeholder impact assessment	Map 5.1	Likelihood and magnitude of each identified impact based on expected use are identified and documented.
MAP-04	No prohibited use documentation	Map 1.1	Intended purposes, potentially beneficial uses, and prospective settings in which the AI system will be deployed are understood and documented.
MAP-04	No prohibited use documentation	Map 3.3	Targeted application scope is specified and documented based on the system's capability and established context.
MEA-01	No bias evaluation	Measure 2.11	Fairness and bias — as identified in the map function — are evaluated and results are documented.
MEA-02	No confidence calibration analysis	Measure 2.3	AI system performance or assurance criteria are measured qualitatively or quantitatively and demonstrated for conditions similar to deployment setting(s).
MEA-02	No confidence calibration analysis	Measure 2.5	The AI system to be deployed is demonstrated to be valid and reliable. Limitations of generalizability are documented.
MEA-03a	No drift monitoring	Measure 3.1	Approaches and documentation are in place to regularly identify and track existing, unanticipated, and emergent AI risks.
MEA-03b	No drift monitoring	Measure 2.4	The functionality and behavior of the AI system and its components are monitored when in production.

Gap ID	Gap Description	AIRC Subcategory	Subcategory Description
MEA-04	Sentiment analysis not validated	Measure 2.3	AI system performance criteria are measured and demonstrated for conditions similar to deployment setting(s).
MEA-05	No adversarial testing	Measure 2.6	The AI system is evaluated regularly for safety risks. The system can fail safely, particularly if made to operate beyond its knowledge limits.
MEA-06	No explainability documentation	Measure 2.9	The AI model is explained, validated, and documented, and AI system output is interpreted within its context to inform responsible use and governance.
MAN-01	No human review queue	Manage 2.4	Mechanisms are in place to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use.
MAN-02	No feedback loop	Measure 3.3	Feedback processes for end users and impacted communities to report problems and appeal system outcomes are established.
MAN-02	No feedback loop	Manage 4.2	Measurable activities for continual improvements are integrated into AI system updates and include regular engagement with relevant AI actors.
MAN-03	No formal SLA	Manage 4.1	Post-deployment AI system monitoring plans are implemented, including mechanisms for incident response, recovery, and change management.
MAN-04	No data retention policy	Manage 1.2	Treatment of documented AI risks is prioritized based on impact, likelihood, and available resources.
MAN-05	Multi-tenancy not implemented	Manage 1.3	Responses to AI risks deemed high priority are developed, planned, and documented.

System Context

System Type: AI-assisted decision support (document extraction)

Risk Tier (NIST): Limited risk — outputs are structured data requiring human review before action

Deployment: Single-tenant, AWS serverless, developer-owned

Data Processed: Business invoices and receipts (no PII beyond vendor/customer names and financial data)

Users: SMB owners, operations managers, bookkeeping staff

Decision Authority: Human — DocFlow extracts and presents; humans act on outputs

---

GOVERN Function

Establish accountability structures, policies, and culture for AI risk management.

Governance starts with accountability — knowing who owns what, who can act on it, and who answers when something goes wrong. For DocFlow, the foundational accountability controls are present: resources are attributed, access is scoped, costs are tracked, and decisions are documented. The gaps in this function are not operational. They are structural. DocFlow has the controls in place. What it lacks is the formal policy layer — the documented acceptable use, the designated risk owner, and the incident response plan — that would make those controls defensible in an enterprise context. Decisions are documented phase-by-phase, traceable to specific dates and reasoning.

What's In Place

Resource Ownership All AWS resources are tagged with `Project`, `CostCenter`, `Environment`, and `ManagedBy`. Weekly automated tag compliance audits (Lambda + EventBridge + SNS) confirm ownership attribution. Every resource has an identifiable owner.

Access Controls

- IAM least-privilege per function — each Lambda has scoped permissions
- API Gateway rate limiting: 5 req/s, burst 10, 100 req/day
- No public write access to any data store
- S3 bucket policies restrict access by function

Cost Controls

- Per-document cost tracked: $0.034
- Budget alerts configured
- Rate limiting as financial guardrail pre-deployment
- CONFIG.SIMULATE pattern — no production API calls during development (simulation mode prevents live service consumption before guardrails are in place)

MAP Function

Identify and categorize AI risks in context of intended use.

Mapping is the due diligence function — understanding what risks the system carries before it touches a production environment. DocFlow's Map posture is stronger than most self-built systems at this stage: the intended use is clearly scoped, known failure modes are documented with root causes and remediation paths, and third-party AI dependencies are identified. What's missing is the formal classification layer — the documented reasoning that places DocFlow in a specific NIST risk tier and defines what that tier means for how the system is deployed and monitored.

What's In Place

DocFlow's intended use is clearly scoped: extract structured data from invoices and receipts. The system is not designed or deployed for decisions with legal, employment, credit, or health consequences.

Known Limitations Register
Phase 4 documents the 20% PDF failure rate with:

- Root cause (deep Textract incompatibility with certain PDF encodings)
- Confidence correlation (failures correlate with document complexity)
- Remediation path (pdf2image + poppler image conversion fallback)
- Deliberate ship decision with documented reasoning

This constitutes a functional known limitations register consistent with NIST AI RMF documentation requirements.

Third-Party AI Dependencies
DocFlow uses two AWS managed AI services:

- AWS Textract — OCR and form extraction
- AWS Comprehend — entity detection, sentiment analysis, key phrases

Both are black-box managed services. Input/output behavior is observable. Internal model architecture and training data are not disclosed by AWS.

MEASURE Function

Analyze and assess AI risks quantitatively and qualitatively.

Measure is where quantitative evidence of system behavior is most visible. DocFlow stores every processed document with a confidence score, a timestamp, and a status — constituting a queryable audit log of all AI system outputs. The gaps in this section are not about missing data. The data exists. What is missing is the analysis layer: the stored outputs have not been interrogated for bias, calibration accuracy, or model drift.

What's In Place

Performance Metrics

Metric	Value	Tracking
Success rate	100% (150-doc batch test)	batch_test.py output
Avg extraction confidence	80.88%	DynamoDB DocFlowRecords
Processing time	~30 seconds	CloudWatch metrics
Known failure rate	~20% (complex PDFs)	Phase 3/4 documentation

Audit Trail

Every processed document is stored in DynamoDB with:

- documentId — unique identifier
- processedAt — ISO timestamp
- status — success/failed
- extractionConfidence — numeric score
- fullText — complete extracted content
- keyValuePairs — structured field extraction
- entities — detected entities
- sentiment — overall sentiment classification

This constitutes a queryable audit log of all AI system outputs.

Confidence Scoring
Textract returns per-block confidence scores. DocFlow calculates and stores average extraction confidence per document. Low-confidence results are surfaced to users at the point of output, enabling human review before action is taken.

MANAGE Function

Prioritize and address identified AI risks.

Manage is the function where DocFlow is strongest. Rate limiting, quota enforcement, graceful failure handling, CloudWatch logging, SNS notifications, and documented remediation paths are all present and operational. The gaps here are enterprise-specific — the controls are sufficient for a single-tenant system but have not been extended to handle multi-tenant data isolation, formal SLAs, or user feedback loops.

What's In Place

Operational Guardrails

- Rate limiting (velocity control): 5 req/s, burst 10
- Daily quota (volume control): 100 req/day
- CONFIG.SIMULATE flag: prevents production API calls during development (simulation mode disables live service consumption until guardrails are in place)
- Graceful failure handling: failed documents return error status, not silent failures

Incident Visibility

- CloudWatch logs all Lambda invocations
- SNS email notification per processed document
- DynamoDB records document status (success/failed)
- CloudTrail captures all API calls

Remediation Paths
All known limitations have documented remediation paths:

- Complex PDF failures → pdf2image + poppler image conversion
- Confidence below threshold → human review queue (architectural recommendation)
- API quota exceeded → alert via AWS Budgets + CloudWatch alarm

Continuous Compliance

- Weekly tag audit Lambda confirms resource ownership and governance metadata compliance across all DocFlow resources.

Enterprise Deployment Requirements

The following requirements represent the delta between DocFlow's current single-tenant, developer-owned state and the minimum viable governance posture for a general enterprise deployment. Items are organized by priority — blockers that must be resolved before any customer deployment, enhancements that reduce ongoing risk, and mature practice items that signal long-term governance commitment.

Must Have

- ☐ MAN-05 — Multi-tenancy — tenant isolation at DynamoDB and S3 layer
- ☐ GOV-02 — Authentication — Cognito or equivalent; no anonymous API access
- ☐ MAN-04 — Data retention policy — define and implement automated deletion
- ☐ MAN-01 — Human review queue — confidence threshold routing
- ☐ MAP-01 — Formal risk classification document (NIST AI RMF tier)
- ☐ GOV-03 — AI use disclosure to end users
- ☐ GOV-04 — Incident response runbook

Should Have

- ☐ MEA-01 — Bias evaluation across document types and demographics
- ☐ MAP-02 — Training data provenance documentation for Textract and Comprehend
- ☐ MEA-02 — Confidence calibration analysis
- ☐ MEA-03a — Periodic regression testing against held-out document set
- ☐ MAN-02 — Feedback mechanism for incorrect extractions
- ☐ MAN-03 — Formal SLA

Nice to Have

- ☐ MEA-05 — Adversarial testing protocol
- ☐ GOV-05 — Third-party audit of AI system outputs
- ☐ MEA-03b — Automated Model drift monitoring
- ☐ MEA-06 — Explainability documentation for Comprehend entity classifications

Remediation Path

Gaps identified in this assessment are prioritized into three horizons based on risk level and effort required for enterprise deployment.

Pre-Deployment Requirements

These items are all deployment blockers. DocFlow cannot be responsibly deployed to a multi-customer or regulated enterprise environment without them.

ID	Item/	Effort	Cost Impact
MAN-05	Multi-tenancy — tenant isolation at S3 and DynamoDB	2–3 days	+$0
GOV-02	Cognito authentication — no anonymous API access	1 day	+$0.0055/MAU
MAN-04	Data retention policy + automated deletion	1 day	Reduces S3/DynamoDB cost
MAN-01	Human review queue for low-confidence extractions	1–2 days	+$0
MAP-01	Formal NIST AI RMF risk classification document	4 hours	$0
GOV-03	AI use disclosure to end users	2 hours	$0
GOV-04	Incident response runbook	4 hours	$0

Total remediation timeline: ~2 weeks with negligible infrastructure cost impact

Short-Term Remediations (30-60 Days)

These items reduce ongoing risk and build the evidence base for model trustworthiness.

ID	Item	Effort	Cost Impact
MEA-01	Bias evaluation across document types	1 week	+$0 (testing cost only)
MAP-02	Training data provenance (Textract + Comprehend)	3–4 days research	$0
MEA-02	Confidence calibration analysis	3–4 days	+$0
MEA-03	Monthly regression testing protocol	2 days setup	+~$5/month
MAN-02	Feedback mechanism for incorrect extractions	2–3 days	+$0
MAN-03	Formal SLA definition	1 day	$0

Total remediation timeline: ~ 3 weeks engineering with additional recurring ~$5/month.

Long-Term Remediations (90+ Days)

These items reflect mature AI governance practice and are appropriate once the system has production usage data to analyze.

ID	Item	Effort	Cost Impact
GOV-05	Third-party independent AI system audit	External engagement	$5,000–$15,000 one-time
MEA-03	Automated model drift monitoring	1 week	+~$10/month (CloudWatch)
MEA-06	Explainability documentation for Comprehend	1–2 weeks research	$0
MEA-05	Adversarial testing protocol	1 week	$0

Total remediation timeline: external audit is the largest investment; infrastructure additions are minimal

---

Assessor Notes

This assessment was conducted by the system's builder, not an independent third party. For enterprise deployment, an independent governance review would be required. This document serves as a self-assessment and interview artifact demonstrating applied knowledge of the NIST AI RMF against a production system.

The most significant governance gap in DocFlow is not technical — it's organizational. The same person who built the system, owns the infrastructure, and interprets the outputs is also the person who conducted this assessment. In any production deployment, those roles must be separated.

That's not a DocFlow-specific problem. It's the central challenge of AI governance at scale: the people closest to the system are least positioned to govern it objectively.

Appendix A: Assessment Scoring Rubric

Alignment status for each NIST AI RMF function is determined by the following criteria:

Status	Definition
🟢 Aligned	All core controls for this function are implemented and documented. Gaps are minor or cosmetic.
🟡 Partial	Core controls are partially implemented. At least one Medium-risk gap exists with no remediation in place.
🔴 Not Aligned	Few or no controls implemented. High-risk gaps present with no remediation path defined.

Companion article: "What FinOps Taught Me About AI Governance" — carlandrainthecloud.substack.com
System repository: github.com/theDovelyDev/theprojectfolder/tree/main/IDR_pipeline