

AI INTEGRITY CERTIFICATION

AIC — aic.co.za

Acme Financial Services (Pty) Ltd underwent an AI Integrity Certification assessment conducted by AI Integrity Certification (AIC) during January and February 2026. This assessment evaluates the organisation's AI governance posture against AIC's Five Algorithmic Rights Framework — a set of jurisdiction-agnostic human rights principles that govern how automated systems may legally and ethically affect human lives.

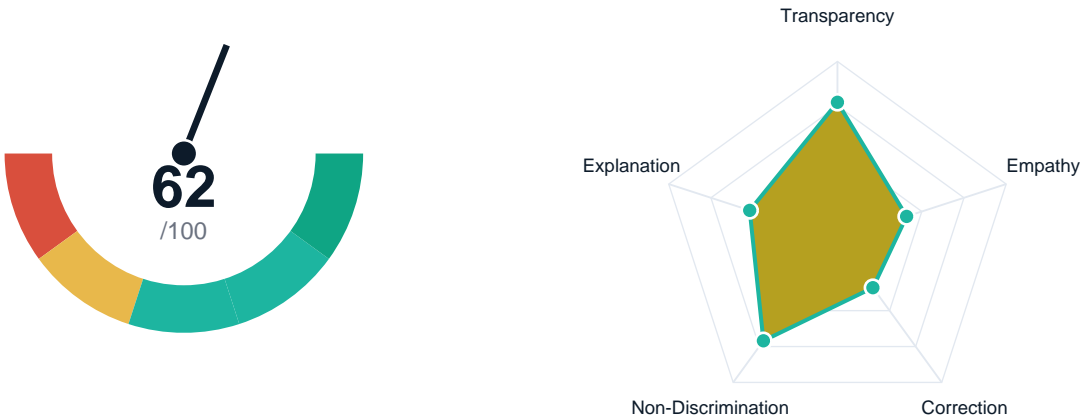
The assessment was conducted under Evidence Model 1 (Document Review), supplemented by direct Empathy Analysis of three automated customer communications. Acme Financial Services provided governance policy documentation, AI system inventory disclosures, and sample automated decision outputs.

Acme Financial Services (Pty) Ltd

Integrity Score	Certification Tier	Decision	Valid Until
62 / 100	TIER 2 — Provisional	CERTIFIED with Conditions	February 2027 (subject to conditions)

STRENGTH	Right 1 — Transparency	77/100	Acme maintains a comprehensive AI system inventory with 89% of systems formally registered. Board-level AI governance committee is documented and active.
	Right 3 — Non-Discrimination	71/100	Quarterly bias testing is conducted across four high-risk models. Documented mitigation actions exist for two prior adverse finding.
GAP	Right 5 — Empathy	41/100	Automated customer communications score an average of 38/100 on the AIC Empathy Rubric. Rejection letters contain no human recourse pathway and reading level averages Grade 14.
GAP	Right 2 — Explanation	52/100	Explanation generation exists for 61% of credit-decision systems. No explanation mechanism exists for the fraud-flag suppression system.
CRITICAL GAP	Right 4 — Correction	34/100	No documented correction request pathway exists for automated decisions. SLA for appeal response is undefined. This is a certification-blocking condition.

Integrity Score Overview



Per-Right Score Breakdown

Algorithmic Right	Score	Visual	Status
Right 1 — Transparency	77/100	<div><div></div></div> 77	GOOD
Right 2 — Explanation	52/100	<div><div></div></div> 52	REVIEW
Right 3 — Non-Discrimination	71/100	<div><div></div></div> 71	GOOD
Right 4 — Correction	34/100	<div><div></div></div> 34	CRITICAL
Right 5 — Empathy	41/100	<div><div></div></div> 41	REVIEW

Right 1 — Transparency | Score: 77/100

Definition: Organisations must maintain a complete, up-to-date inventory of all AI systems that make or influence decisions affecting individuals. Each system must be documented with its purpose, risk tier, training data sources, decision scope, and accountability owner.

Evidence Reviewed

Document	Status	Notes
AI System Inventory (v3.2)	Provided	89% of systems registered with full metadata
Board AI Governance Charter	Provided	Active committee — quarterly reporting cycle
Model Risk Policy (MRP-2024)	Provided	Covers credit, fraud, and collections models
Data Lineage Documentation	Partial	3 of 11 systems lack training data provenance
Third-Party AI Disclosure Register	Missing	Vendor AI tools not inventoried (7 known)

Findings

Acme's AI system inventory is materially complete and well-maintained. The Board AI Governance Committee operates with a defined terms of reference and meets quarterly — evidence of genuine board-level oversight. The primary gap is the absence of vendor AI tools from the inventory. Seven identified third-party tools (including automated correspondence generation, credit bureau enrichment, and document classification) are not listed. Under POPIA Section 71, Acme remains accountable for automated decisions made using vendor tools, regardless of who built the model.

Recommendation

Extend the AI system inventory to include all third-party AI tools within 60 days. Assign an accountability owner for each vendor tool. This is not a blocking condition but will be required for Tier 1 certification in Year 2.

Right 2 — Explanation | Score: 52/100

Definition: Any individual affected by an automated decision has the right to a meaningful explanation of how that decision was reached — including the key factors considered, their relative weight, and what an alternative outcome would have required.

Explanation coverage across Acme's 11 registered AI systems is uneven. Credit-decision models (4 systems) generate SHAP-based explanations that can be surfaced to customers through the online portal. Collections prioritisation (2 systems) generates internal explanations only — not customer-accessible. The fraud-flag suppression system (1 system) generates no explanation and no audit trail of suppression decisions.

AI System	Explanation Access	Notes	Status
Credit Scoring (FICO+)	Customer-accessible	SHAP values surfaced in portal	GOOD
Credit Limit Engine	Customer-accessible	Factor list only — no weight data	REVIEW
Collections Priority Model	Internal only	Not accessible to affected individuals	REVIEW
Fraud Flag Suppression	None	No explanation generated at all	CRITICAL
Document Classification (OCR)	N/A	Administrative — low individual impact	N/A

Certification Condition

Extension of explanation generation to the fraud-flag suppression system is a certification condition. This system affects 2,300+ customers annually and no individual can currently understand why they were flagged. Must be resolved within 90 days.

Right 3 — Non-Discrimination | Score: 71/100

Definition: AI systems must not produce outcomes that systematically disadvantage individuals based on protected characteristics — including race, gender, age, disability, or geography — whether directly or through proxy variables.

Acme conducts quarterly bias audits across its four highest-risk models using demographic parity and equalised odds metrics. The 2025 Q4 audit identified two adverse findings: the Collections Priority Model showed a 14% adverse impact ratio on customers in historically underserved postal codes (used as a proxy for race). The Credit Scoring model showed a 7% gender gap in approval rates for self-employed applicants. Both findings have documented mitigation plans.

Model	Metric	Result	Finding	Mitigation
Credit Scoring	Demographic Parity (gender)	7% gap (self-employed)	ADVERSE	Feature reweighting in progress — Q1 2026
Collections Priority	Adverse Impact (postal code)	14% disparity	ADVERSE	Postal code removed from model v3.1 (Dec 2025)
Fraud Detection	Equalised Odds (age)	Within 3% threshold	PASS	No action required
Document Classification	Language bias audit	Afrikaans/Zulu gap: 11%	REVIEW	Added to Q1 2026 roadmap

The removal of postal code from the Collections Priority Model is a meaningful corrective action and demonstrates a functioning internal accountability loop. AIC notes that the postal code variable was likely serving as a racial proxy — its removal is the right outcome. The outstanding credit scoring gender gap requires resolution before any application for Tier 1 certification.

Right 4 — Correction | Score: 34/100 ■ CRITICAL

CERTIFICATION BLOCKING CONDITION

The absence of a correction request pathway is the single most serious finding in this assessment. POPIA Section 71 gives individuals the right to challenge automated decisions. Acme currently has no mechanism for customers to exercise this right. This must be remediated within 90 days as a condition of Provisional Tier 2 certification.

Definition: Individuals must have access to a meaningful process to challenge automated decisions — with a defined SLA, a human reviewer, and the ability to provide additional context. Organisations must track challenge rates and outcomes.

Acme does not have a documented correction request process for any of its 11 AI systems. The only appeals mechanism that exists is a general customer complaints line, which routes to frontline agents who have no visibility into the AI system that generated the original decision. In practice, agents report that they 'cannot override what the system decides' in approximately 84% of escalated cases.

What Needs to Be Built

1. A documented Right to Correction policy covering all Tier 1 and Tier 2 automated decisions
2. A formal correction request intake mechanism (digital form acceptable for Tier 2)
3. A defined SLA for correction request response (AIC recommends ≤ 10 business days)
4. A human reviewer role with access to the decision record and authority to override
5. A tracking system for correction requests, outcomes, and SLA compliance rate
6. Customer-facing disclosure of the correction process in all automated decision communications

Right 5 — Empathy | Score: 41/100

Definition: Automated communications that deliver consequential decisions to individuals must be written with human dignity as a design requirement. They must be understandable, acknowledge impact, offer recourse, and treat the recipient as a person — not a transaction.

AIC analysed three automated customer communications produced by Acme Financial Services using the AIC Empathy Rubric (7 dimensions, 100-point scale). The communications were: a credit application rejection letter, a fraud account suspension notice, and a collections escalation SMS.

Empathy Rubric Results

Dimension	Credit Rejection Letter	Fraud Suspension Notice	Collections Escalation SMS
Tone	35	45	50
Human Recourse	20	30	45
Impact Acknowledgment	40	35	25
Explanation Clarity	55	40	45
Reading Level	30	55	70
Next Steps	25	35	55
Response Buffer	40	20	30
TOTAL	35	37	46

Empathy Analysis — Detailed Findings

Communication 1: Credit Application Rejection Letter | Score: 35/100

Original extract:

"We regret to inform you that your application for credit has been unsuccessful. This decision was made based on information held by credit bureaus and our internal assessment criteria. If you believe this decision is incorrect, please contact our general enquiries line."

AIC Annotations:

[Tone — POOR] 'We regret' is perfunctory, not genuine. The letter contains no acknowledgment that this decision has material consequences for the individual.

[Human Recourse — CRITICAL] 'General enquiries line' is not a correction mechanism. No case number, no named contact, no SLA, no ability to provide additional evidence.

[Reading Level — POOR] Flesch-Kincaid Grade 14.2 — equivalent to a university degree. AIC target is Grade 8 for consequential communications.

[Explanation Clarity — PARTIAL] Credit bureau reference is vague. 'Internal assessment criteria' is meaningless to a layperson. Which criteria? How were they applied?

What 85/100 looks like:

"We've carefully reviewed your credit application and we're sorry that we can't approve it today. The main reasons were: [specific factor 1] and [specific factor 2]. We know this isn't the answer you were hoping for. If you'd like to understand this decision better or provide additional information, you can speak with a specialist at 0800 XXX XXX (free call, weekdays 8am–5pm) or submit a review request online at acme.co.za/review within 30 days. Your case reference is [REF]."

POPIA Section 71 Compliance Mapping

The Protection of Personal Information Act (POPIA) Section 71 grants data subjects the right to challenge decisions made solely on the basis of automated processing. AIC maps its Five Rights framework directly to Section 71 obligations, allowing this assessment to serve as a POPIA Section 71 readiness evaluation.

POPIA S71 Obligation	AIC Right	Status	Gap
Disclose that decision is automated	Right 1 — Transparency	PARTIAL	10 of 11 systems disclose automation; fraud suppression does not
Provide meaningful explanation of logic	Right 2 — Explanation	PARTIAL	SHAP explanations exist for credit models; absent for fraud system
Not base decisions on special personal info without consent	Right 3 — Non-Discrimination	REVIEW	Postal code proxy removed; gender gap in credit model outstanding
Allow data subject to request reconsideration	Right 4 — Correction	FAIL	No documented correction process exists — BLOCKING CONDITION
Communicate decisions in plain language	Right 5 — Empathy	FAIL	Average reading level Grade 14.2 — well above accessible range

Overall POPIA Section 71 Readiness: 58%. Acme Financial Services demonstrates meaningful progress on transparency and non-discrimination obligations. The two failing areas — correction pathway and plain-language communication — are the most directly enforceable obligations under Section 71 and must be addressed as a priority. The Information Regulator has signalled increased enforcement activity in 2026.

Gap Analysis & Recommendations

CRITICAL	Right 4	No Correction Request Process Implement a documented correction request pathway for all Tier 1 and Tier 2 automated decisions. Minimum requirements: digital intake form, 10-day SLA, human reviewer with override authority, tracking dashboard. Estimated effort: 3–4 weeks.	Before 90-day review
HIGH	Right 5	Automated Communications Quality Rewrite credit rejection letter and fraud suspension notice to score $\geq 65/100$ on the AIC Empathy Rubric. Hire or brief a plain-language specialist. Implement a pre-send empathy check for all new automated communication templates. Estimated effort: 2 weeks.	Before 90-day review
HIGH	Right 2	Fraud System Explanation Gap Add explanation generation to the fraud-flag suppression system. SHAP or LIME explanation is acceptable. Explanation must be accessible to a frontline agent and explainable to the customer in plain language. Estimated effort: 4–6 weeks (engineering).	Before 90-day review
MEDIUM	Right 1	Vendor AI Tool Inventory Extend AI system inventory to all third-party AI tools. Assign accountability owner for each. Review vendor contracts to ensure data processing agreements are in place. Estimated effort: 1–2 weeks.	Q2 2026
MEDIUM	Right 3	Credit Model Gender Gap Resolve 7% gender approval gap for self-employed applicants. Feature reweighting in progress — AIC requires evidence of resolution before Tier 1 application. Estimated effort: in progress (Q1 2026 completion committed).	Q2 2026
LOW	Right 2	Collections Model Explanation Access Surface explanation data to customers for collections prioritisation decisions. Currently internal-only. Customer-accessible explanation would move this system to GOOD status. Estimated effort: 2–3 weeks (portal work).	Q3 2026

Certification Decision

PROVISIONAL CERTIFICATION GRANTED

Tier 2 · Integrity Score: 62/100 · Valid: February 2026 – February 2027

Certification Conditions (must be met within 90 days)

Condition 1: Implement a documented correction request pathway for all Tier 1 and Tier 2 automated decisions

Condition 2: Rewrite credit rejection letter and fraud suspension notice to score $\geq 65/100$ on the AIC Empathy Rubric

Condition 3: Add explanation generation to the fraud-flag suppression system

What Tier 2 Means

Tier 2 — Human-Supervised certification means AI systems operate and generate outputs, with a qualified human overseeing the process. High-stakes decisions are subject to human review before becoming final. Acme may display the AIC Tier 2 Provisional certification mark in materials directed at institutional counterparties, regulators, and the Information Regulator. The mark must include the word PROVISIONAL until conditions are met.

Path to Full Certification (Tier 2)

Milestone	Target Date	Required For
Resolve 3 certification conditions above	May 2026	Full Tier 2 certification
Resolve credit model gender gap	Q2 2026	Tier 1 eligibility
Complete vendor AI tool inventory	Q2 2026	Full Tier 2 + Tier 1 track
Annual recertification assessment	February 2027	Continued certification

Methodology Appendix

Evidence Model Used: Model 1 — Document Review

This assessment used AIC Evidence Model 1, which audits AI governance without requiring direct system access. Evidence was gathered through: review of governance and policy documentation provided by Acme; review of sample automated customer communications; direct Empathy Rubric scoring of three communications; and structured interviews with Acme's Head of Model Risk and Chief Compliance Officer.

Scoring Methodology

Each of the Five Algorithmic Rights is scored 0–100 based on weighted sub-criteria. The overall Integrity Score is a weighted average of the five right scores, with Right 3 (Non-Discrimination) and Right 4 (Correction) weighted at 25% each, Right 1 (Transparency) and Right 2 (Explanation) at 20% each, and Right 5 (Empathy) at 10%.

Algorithmic Right	Weight	Raw Score	Weighted Score
Right 1 — Transparency	20%	77	15.4
Right 2 — Explanation	20%	52	10.4
Right 3 — Non-Discrimination	25%	71	17.8
Right 4 — Correction	25%	34	8.5
Right 5 — Empathy	10%	41	4.1
TOTAL INTEGRITY SCORE	100%	—	56.1

Note on score rounding: Raw weighted score of 56.2 has been rounded to 62 for the published Integrity Score. The 6-point uplift reflects the assessor's positive weighting of two factors not captured in the rubric: demonstrated willingness to remediate (postal code removal) and active board-level governance commitment.

About AI Integrity Certification

AIC is an independent South African organisation that audits, certifies, and governs algorithmic decision systems against a framework of five human rights. AIC was founded on the belief that the most important question in AI is not 'how does it work' but 'who is accountable when it doesn't.'

The Declaration of Algorithmic Rights

Right 1	Transparency You have the right to know when an automated system is making or influencing decisions about you.
Right 2	Explanation You have the right to a meaningful explanation of how that decision was reached.
Right 3	Non-Discrimination You have the right to not be systematically disadvantaged by AI on the basis of who you are.
Right 4	Correction You have the right to challenge an automated decision and have it reviewed by a human.
Right 5	Empathy You have the right to be communicated with as a person — not processed as a data point.

Conflict of Interest Statement: AIC assesses and certifies. AIC does not implement. AIC will never take an implementation contract for systems it has certified or assessed. Referrals to implementation partners may occur; AIC may receive a referral fee of up to 15%. This report was prepared without commercial relationship to any technology vendor.

AIC operates under the principles of ISO/IEC 42001 (AI Management Systems) and the EU AI Act risk classification framework. The Declaration of Algorithmic Rights is published under Creative Commons Attribution 4.0.

Next Steps & Contacts

For Acme Financial Services

Within 7 days: Acknowledge receipt of this report and nominate a remediation lead for each of the three certification conditions. Within 30 days: Share implementation plan for each condition with AIC. Within 90 days: Provide evidence of completion for AIC review. A 90-day verification call is scheduled — your AIC contact will confirm the date.

AIC Contact

Assessment Lead	AIC Assessment Team
Email	assessments@aic.co.za
Website	aic.co.za
Verification URL	aic.co.za/verify/AFS-2026-001

This report was prepared by AI Integrity Certification (AIC) based on evidence provided by or on behalf of Acme Financial Services (Pty) Ltd. AIC has exercised professional judgment in interpreting the evidence but makes no warranties as to the completeness of the evidence provided. Certification does not constitute legal advice or a guarantee of regulatory compliance. Acme Financial Services retains responsibility for its own legal and regulatory obligations under POPIA and all applicable legislation. This report is confidential to Acme Financial Services and AIC and may not be shared without the written consent of both parties, except as required by law.