

# AGENTPR™

## Glassbox AI Policy

*Not Black Box. Glassbox.*

### POLICY STATEMENT

AGENTPR™ is built on a single conviction: AI that cannot explain itself cannot be trusted. This policy is the written expression of that conviction — documenting how every AI-generated output on the platform is built, labelled, reviewed, and accountable.

The Glassbox principle is the engineering expression of the Engineering of Trust™ framework: every output must be traceable, every confidence level visible, every human analyst empowered to override.

 DOCUMENT  
**Glassbox AI Policy v1.0**

 OWNER  
**Cihan Media Communications**

 EFFECTIVE  
**June 2026**

 NEXT REVIEW  
**September 2026**

### CROSS-WALKED AGAINST

Microsoft RAI Standard v2 · NIST AI RMF 1.0 · EU AI Act (Reg. 2024/1689) · NDPA Nigeria 2023

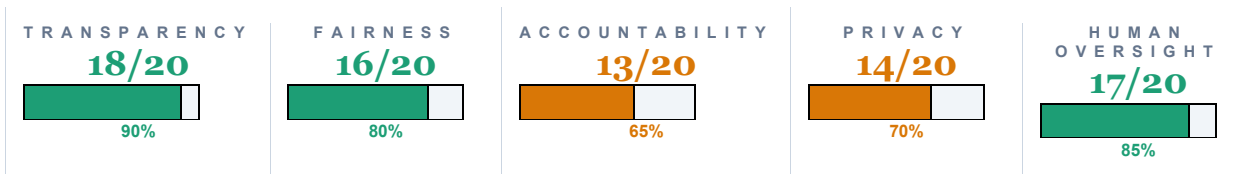
### CONTENTS

<b>01</b>	<b>The Glassbox Principle</b>	<i>The originating framework and its three pillars</i>
<b>02</b>	<b>What We Do With Your Data</b>	<i>Collection, processing, retention, and third-party processors</i>
<b>03</b>	<b>How Our AI Makes Decisions</b>	<i>The GASD™ pipeline and four-agent architecture</i>
<b>04</b>	<b>Confidence Labelling</b>	<i>Five labels, what each means, and what to do</i>
<b>05</b>	<b>What We Do Not Do</b>	<i>Prohibited uses — the ethical boundary</i>
<b>06</b>	<b>Your Rights</b>	<i>Six data rights with response SLAs</i>
<b>07</b>	<b>How to Raise a Concern</b>	<i>Three intake channels — all human-reviewed</i>

EXECUTIVE RAI SCORECARD — JUNE 2026 ASSESSMENT



RAI DIMENSION SCORES — CURRENT STATE



PRINCIPLE-BY-PRINCIPLE ASSESSMENT

Principle	Status	Evidence
Engineering of Trust™	ORIGINATING	The proprietary framework that preceded and informed the international standards crosswalk
Transparency (T1)	MET ✓	Confidence labelling system: five labels on every output. Source attribution. Glassbox AI Policy published.
Inclusiveness (I1)	MET ✓	Meaning Intelligence™ calibrated for Nigerian sarcasm, Pidgin English, African institutional context.
Accountability (A1)	PARTIAL ●	Named accountable executive. 90-day roadmap active. Incident response process in development.
Privacy (PS1)	PARTIAL ●	NDPA 2023 compliant. GDPR cross-border addendum in progress. DPIA summary in preparation.
Human Oversight (A5)	MET ✓	Human Review Recommended label on all low-confidence or high-risk outputs. Analyst sign-off required.
Fairness (F1)	MET ✓	Platform calibrated for African linguistic and cultural context. Error-rate parity study in roadmap.
Security (PS2)	PARTIAL ●	Role-gated access logged internally. Prompt-injection hardening and OAuth audit in 90-180 day roadmap.

Assessment conducted June 2026 against Microsoft RAI Standard v2, NIST AI RMF 1.0, EU AI Act, and NDPA Nigeria 2023. The Engineering of Trust™ framework by Dr. Celestine N. Achi originated these commitments and preceded this cross-walk.

# 001

## The Glassbox Principle

*The originating framework and its three architectural pillars*

Most AI platforms operate as black boxes. Data enters. Outputs emerge. The reasoning between them is invisible, unverifiable, and unaccountable. AGENTPR™ was built on the opposite architecture.

The Glassbox principle originates from Dr. Celestine N. Achi's Blackbox-Glassbox Framework — a core component of The Achi Intellectual Architecture™. It preceded and informed the international RAI standards cross-walk on this page.

 <p><b>Methodology</b> The GASD™ pipeline: four specialised agents, eight structured sweeps, six live data sources. The same sequence, the same discipline, for every brand, every brief, every week. Repeatable by design.</p>	 <p><b>Transparency</b> Every source we read is shown. Every source filtered is explained. Every claim carries a confidence label. If it cannot be traced, it is not shipped.</p>	 <p><b>Accountability</b> Named accountable executive. Confidence labels and human-review prompts at every stage. A clear paper trail from raw signal to final recommendation. Human judgement is the final layer.</p>
--	--	---

### BLACKBOX AI VS GLASSBOX AI

<p><b>BLACKBOX AI</b></p> <ul style="list-style-type: none"> <li>• Outputs presented as facts — no reasoning shown</li> <li>• Confidence levels hidden from the analyst</li> <li>• Cultural and linguistic errors invisible and uncorrectable</li> <li>• No source trail — cannot verify what was read</li> <li>• User cannot distinguish certainty from inference</li> </ul>	<p><b>GLASSBOX AI — AGENTPR™</b></p> <ul style="list-style-type: none"> <li>• Every output carries a confidence label</li> <li>• Source links embedded throughout — every claim traceable</li> <li>• Cultural risk flags surfaced explicitly for analyst review</li> <li>• Human Review Recommended label on all uncertain items</li> <li>• Analyst always knows the difference between verified and inferred</li> </ul>
---	--

# 002

## What We Do With Your Data

Collection, processing, retention, and third-party processors

### WHAT WE COLLECT

- Account information: name, email, organisation, and job title
- Brand intelligence inputs: brand names and keywords submitted for monitoring
- Voice of Customer data: feedback files uploaded for VoC analysis
- Usage data: platform interactions and session logs for service improvement
- MCP connector usage: tool calls through the Claude MCP integration for quota management
- Payment details: processed by Paystack. Card data never stored on AGENTPR™ servers.

Dark social monitoring refers exclusively to signals publicly re-shared from messaging platforms. We do not access private message contents. We monitor what has entered the public domain — not the private channels themselves.

### DATA RETENTION SCHEDULE

Data Category	Retention	Basis
Intelligence briefs	90 days	Deleted after 90 days unless exported by the user
VoC analysis data	90 days	Uploaded files deleted immediately after processing
Account data	Subscription + 30 days	Deleted 30 days after subscription ends
Payment records	7 years	Legal obligation under Nigerian financial regulations
MCP usage logs	30 days rolling	Quota management only
Support communications	2 years	Service continuity and audit purposes

### THIRD-PARTY PROCESSORS

Processor	Function	Legal basis and commitment
Anthropic (Claude AI)	AI inference layer	Enterprise commercial terms. Customer data is not used to train models.
Supabase	Database and storage infrastructure	EU-standard data processing agreement applies.
Paystack	Payment processing	PCI-DSS compliant. No card data stored on AGENTPR™ servers.
Zernio	Content distribution infrastructure	Processes only content you explicitly authorise for publication.
Firecrawl	Web content retrieval	Processes only publicly accessible URLs submitted through the platform.

# 003

## How Our AI Makes Decisions

*The GASD™ pipeline and four-agent architecture in plain language*

The GASD™ pipeline — Gather, Analyse, Synthesise, Deliver — is the decision architecture of AGENTPR™. Four specialised agents. One orchestrated sequence. The same methodology for every brand, every brief, every week.

<p style="text-align: center;"><b>G</b> SADA GATHER</p> <p>Searches six live data sources across eight structured sweeps: online news, social media, dark social signals, broadcast (41 outlets), print and digital, forums, influencer signals, and competitive landscape.</p>	<p style="text-align: center;"><b>A</b> ZALI ANALYSE</p> <p>Applies Meaning Intelligence™: sentiment, 8-emotion profile, sarcasm risk, intent, cultural nuance (Pidgin, idiom, local context), reputation risk. Every output carries a confidence label.</p>	<p style="text-align: center;"><b>S</b> NZE SYNTHESISE</p> <p>Builds the intelligence layer: narrative arc, competitive signals, leadership implications, forward intelligence. Also generates narrative strategy for the Narrator engine.</p>	<p style="text-align: center;"><b>D</b> CNA ORCHESTRATE</p> <p>Coordinates the full pipeline. Enforces tier limits. Surfaces low-confidence flags for analyst review. Ensures nothing leaves without appropriate labelling and human oversight prompts.</p>
---	--	--	---

AGENTPR™ is an analyst-augmentation system, not an autonomous decision-maker. Every intelligence brief is intended to be read by a trained communications professional before any external action is taken.

# 004

## Confidence Labelling

*Five labels — what each means and what to do*

Every data point in an AGENTPR™ brief carries one of five confidence labels. These are the operational expression of the Glassbox principle: not just what the platform found, but how much to trust it and what to do next.

Label	What it means	Analyst action
● <b>Monitored</b>	Data recorded. No AI interpretation applied. Raw data point.	→ Use directly
● <b>AI-Sourced</b>	AI interpreted with high confidence. Clear signals, multiple sources.	→ Use with standard review
● <b>Firecrawl-Verified</b>	Verified by reading the full source article. Not just the headline.	→ High confidence — cite freely
● <b>AI-Inferred</b>	AI drew an inference from available signals. Less direct evidence.	→ Apply professional judgement
● <b>Human Review Recommended</b>	AI confidence is low or cultural complexity warrants analyst review.	→ Do not use without analyst review

Human Review Recommended is the most important label on the platform. It means AGENTPR™ is telling you explicitly: a trained human professional must review this finding before it appears in any client-facing or institutional communication.

# 005

## What We Do Not Do

*Prohibited uses — the ethical boundary of the platform*

The following uses are explicitly prohibited under AGENTPR™ Terms of Service and this Glassbox AI Policy. These are not edge cases. They are named exclusions that define the ethical boundary.

P R O H I B I T E D	P E R M I T T E D
<ul style="list-style-type: none"> <li>Monitoring private conversations — WhatsApp and Telegram contents are never accessed</li> <li>Surveillance of private individuals not in a public or professional capacity</li> <li>Use against or about children or minors (platform is 18+ only)</li> <li>Generating defamatory, harassing, or discriminatory content</li> <li>Autonomous external action without analyst review of flagged outputs</li> <li>Use as health, legal, financial, or regulatory advice</li> <li>Workplace or education monitoring using emotion-recognition features</li> <li>Social scoring of individuals by public authorities — excluded under EU AI Act Art. 5</li> </ul>	<ul style="list-style-type: none"> <li>Brand and reputation monitoring of organisations and public figures in their public capacity</li> <li>Media intelligence for PR agencies, corporate communications, government, and public affairs</li> <li>Voice of Customer analysis of voluntarily submitted feedback</li> <li>Narrative strategy development and content distribution</li> <li>Influencer campaign management with full compliance tracking</li> <li>Academic and research use of public media data with appropriate attribution</li> <li>Training and capacity building for communications professionals</li> </ul>

# 006

## Your Rights

*Six rights — each with a committed response SLA*

AGENTPR™ is regulated under the Nigeria Data Protection Act 2023 (NDPA) and the Nigeria Data Protection Commission (NDPC). Where operations involve EU-resident data subjects, the GDPR also applies.

Your Right	SLA	What it means for you
Right of Access	30 days	Request a copy of all personal data AGENTPR™ holds about you.
Right of Correction	14 days	Request correction of any inaccurate personal data. We confirm once corrected.

<b>Right of Deletion</b>	<b>30 days</b>	Request deletion of your account and all associated data, subject to legal retention obligations.
<b>Right to Object</b>	<b>30 days</b>	Object to specific processing activities. We review, respond, and document the outcome.
<b>Contestability of AI Outputs</b>	<b>14 business days</b>	If a reputation score or intelligence finding about your organisation is incorrect, raise a formal contestation. Human review is conducted by a named analyst.
<b>Right to Data Portability</b>	<b>30 days</b>	Request your data in a structured, machine-readable format.

# 007

## How to Raise a Concern

Three intake channels — all reviewed by a human

AGENTPR™ operates a human-reviewed feedback intake for all Responsible AI concerns. No concern is handled by automated response alone. Every submission is acknowledged within 2 business days.

<p><b>Responsible AI Feedback</b>  <b>When:</b> You believe a platform feature, output, or practice raises a responsible AI concern.  <b>Email:</b> <a href="mailto:support@useagentpr.com">support@useagentpr.com</a>  <b>Subject:</b> "Responsible AI feedback"  <b>Human review within 5 business days</b></p>	<p><b>AI Output Contestation</b>  <b>When:</b> You believe a specific score, verdict, or finding about your organisation is incorrect.  <b>Email:</b> <a href="mailto:support@useagentpr.com">support@useagentpr.com</a>  <b>Subject:</b> "AI Output Contestation"  <b>Human review within 14 business days</b></p>	<p><b>Data Protection</b>  <b>When:</b> You wish to exercise a data right or report a data breach.  <b>Email:</b> <a href="mailto:support@useagentpr.com">support@useagentpr.com</a>  <b>Subject:</b> "Privacy / Data Protection"  <b>30 days. Breaches reported to NDPC within 72 hours.</b></p>
---	---	---

All Responsible AI feedback is reviewed by a named human analyst — not an automated system. Reports are acknowledged within 2 business days and feed into both the AGENTPR™ methodology and this policy document.

### ACTIVE RAI COMMITMENTS ROADMAP

<p><b>PRIORITY 1</b>  <b>0 — 90 Days</b></p> <ul style="list-style-type: none"> <li>• Publish Glassbox AI Policy ✓</li> <li>• Publish DPIA summary</li> <li>• Publish sarcasm evaluation methodology</li> <li>• Clarify dark social disclosure</li> <li>• Publish ZALI system card</li> <li>• Sensitive Use review — government tenants</li> </ul>	<p><b>PRIORITY 2</b>  <b>90 — 180 Days</b></p> <ul style="list-style-type: none"> <li>• Quarterly error-rate parity report</li> <li>• C2PA content credentials on Narrator</li> <li>• Contestability workflow for reputation scoring</li> <li>• Prompt-injection hardening</li> <li>• OAuth scope audit for MCP connector</li> <li>• GDPR addendum for EU-resident subjects</li> </ul>	<p><b>PRIORITY 3</b>  <b>180 — 365 Days</b></p> <ul style="list-style-type: none"> <li>• Independent third-party RAI audit</li> <li>• ISO/IEC 42001 readiness assessment</li> <li>• Red-team programme for Narrator</li> <li>• Public incident and feedback log</li> <li>• Confidence label calibration study</li> <li>• RAI review board — quarterly cadence</li> </ul>
--	--	--

## G O V E R N A N C E &amp; A C C O U N T A B I L I T Y

<b>Accountable Executive (RAI)</b>	Dr. Celestine Ngozichukwuka Achi, Dr. FAIMFIN, FIIM, MNIPR, MNIMC Founder & CEO, Cihan Media Communications
<b>Technical Co-owner</b>	Orimolade Oluwamuyemi, FIIM — Strategic Technology Adviser
<b>Data Protection Contact</b>	support@useagentpr.com — Subject: “Privacy / Data Protection”
<b>Responsible AI Intake</b>	support@useagentpr.com — Subject: “Responsible AI feedback”
<b>Regulator of Record</b>	Nigeria Data Protection Commission (NDPC)
<b>Document Version</b>	v1.0 — June 2026
<b>Next Scheduled Review</b>	September 2026
<b>Published At</b>	useagentpr.com/responsible-ai · aipoweredpr.com/glassbox-ai-policy

**AGENTPR™ — Glassbox AI Policy***Not black box. Glassbox.*

Engineering of Trust™ · The Achi Intellectual Architecture™ · Cihan Media Communications · June 2026