UBIQUITY ARTIFICIAL INTELLIGENCE (PTY) LTD

Parliamentary Policy Submission

Draft South Africa National Artificial Intelligence (AI) Policy

Primary Reference	Government Gazette No. 54477, Notice 3880 of 2026 Draft South Africa National Artificial Intelligence (AI) Policy Published 10 April 2026
Submitted by	Ubiquity Artificial Intelligence (Pty) Ltd
Founder	Kaveer Beharee
Date	14 April 2026
Comment Deadline	10 June 2026
Submission Position	Supportive of the Draft Policy's direction, with implementation-level detail in areas requiring further regulatory specificity

This submission is provided as a constructive, technically grounded contribution to South Africa's national AI governance framework. It is offered in a supportive spirit - not as an alternative to the Draft Policy, but as an implementation-focused extension of it.

Table of Contents

Submission Alignment Matrix

The table below maps each section of this submission to the relevant provisions of the Draft National AI Policy.

Submission Section	Draft Policy Anchor	Value Added by This Submission
Executive Summary	Explanatory Note; Foreword (pp. 4-7)	Positions this submission as implementation-focused and supportive of the Draft Policy's direction.
1. Introduction and Global Context	Section 1 (pp. 8-13)	Retains regulatory-vacuum analysis and continental leadership framing.
2. National Security and AI	Rights and public-interest framing (pp. 8-10, 22-30)	Adds detail on unacceptable-risk use cases, procurement, and electoral integrity.
3. Data Residency and Digital Sovereignty	Policy context and infrastructure references (pp. 9-10, 24-30)	Develops a risk-tiered data-governance and sovereignty model.
4. Privacy, Data Protection and AI	POPIA, PAIA, rights-based approach (pp. 9-10, 22–25, 27-30)	Operationalises privacy, explainability, and biometric safeguards.
5. Consumer Protection in the AI Economy	Human-centred values and vulnerable communities (pp. 22-25)	Preserves the financial AI analysis and identifies a key high-risk sector.
6. Fair Use, IP and AI-Generated Content	Copyright and performers' rights references (pp. 9-10)	Develops practical options for training data, AI-generated works, and synthetic media.
7. Algorithmic Accountability and Transparency	Ethics, fairness, accountability, explainability (pp. 22-30)	Translates principles into impact assessments, classification, and oversight mechanisms.
8. AI in Critical National Infrastructure	Infrastructure, safety, monitoring, cross-sector regulation (pp. 24-30)	Adds resilience, continuity, supply-chain, and incident response detail.
9. International Regulatory Benchmarking	AU, Smart Africa, OECD, UNESCO references (pp. 10, 22-23, 27)	Benchmarks international models calibrated for South Africa's context.
10. Proposed Regulatory Framework	Institutional infrastructure and staged implementation (pp. 24-30, 4-5)	Refines institutional proposals to fit the Draft Policy's phased pathway.
11. Implementation Roadmap	Explanatory Note and staged implementation (pp. 4-5)	Adds sequencing, sandboxing, and capacity detail consistent with the Draft Policy.
12–13. Additional Issues and Conclusion	Vision, objectives and sectoral strategy (pp. 16-31)	Retains broader strategic ideas while closing in a constructive, aligned tone.

Executive Summary

The Draft National Artificial Intelligence Policy establishes a strong foundation for South Africa's approach to artificial intelligence, particularly through its emphasis on inclusive growth, human-centred values, ethical governance, institutional development, and a phased regulatory pathway. Ubiquity AI welcomes that direction and submits this document as a constructive, implementation-focused contribution.

This submission integrates Ubiquity AI's substantive policy analysis with direct alignment to the gazetted Draft Policy, preserving original contributions in several areas where the Draft Policy is intentionally high-level and open-ended. These areas include: a risk-proportionate regulatory regime; national security and critical-infrastructure safeguards; data residency and digital sovereignty; AI-specific privacy protections; consumer protection in the AI economy; intellectual property and synthetic media issues; algorithmic accountability; and a detailed treatment of financial services AI as a high-risk domain.

A central contribution of this submission is the argument that financial services should be treated as a high-risk AI deployment environment in South Africa. In a society characterised by deep inequality, over-indebtedness, irregular incomes, and uneven access to formal financial services, AI systems used in credit, insurance, debt collection, financial guidance, and consumer engagement can either expand financial dignity or entrench exclusion. The Draft Policy's commitment to inclusion, fairness, and human-centred design creates the policy basis for a more detailed regulatory response in this domain.

At several points, the original Ubiquity AI analysis was more prescriptive than the Draft Policy. Those recommendations have been refined here to fit the Draft Policy's staged approach. For example, proposals relating to a future Artificial Intelligence Governance Act, a dedicated regulatory authority, and more specific mandatory standards are now positioned as medium-term or phased implementation options rather than immediate preconditions. This adjustment preserves the substance of the original proposal while aligning it to the Government's stated sequencing of policy finalisation, guidelines, sectoral strategies, institutional design, and later regulatory development.

Ubiquity AI notes and endorses the Draft Policy’s identification of Education, Healthcare, Agriculture, and Public Administration as the four critical priority sectors for AI implementation in South Africa. These sectors reflect genuine developmental urgency and represent the domains in which AI can most immediately improve the lives of the majority of South Africans.

This submission’s detailed treatment of financial services AI, critical infrastructure AI, and national security AI should be read as supplementary to - not in competition with - those priority sectors. This submission is offered in a supportive spirit: not as an alternative to the Draft Policy, but as a technically grounded extension of it. Ubiquity AI respectfully submits that South Africa has an opportunity to define a globally relevant model of AI governance that is constitutionally grounded, economically developmental, technologically credible, and distinctly responsive to the realities of the African continent.

1. Introduction and Global Context

1.1 The AI Inflection Point

We are living through the most consequential technological transition since the Industrial Revolution. Generative AI systems can now produce human-quality text, images, audio, code, and decisions. Large Language Models (LLMs), multimodal foundation models, autonomous agents, and AI-powered decision systems are being deployed across financial services, healthcare, education, criminal justice, social welfare, and national defence - often without adequate legal frameworks, transparency, or accountability mechanisms.

Global private AI investment reached approximately USD 140 billion in 2024, according to the Stanford AI Index - with total estimated economic impact projections substantially higher. The International Monetary Fund estimates that AI could affect 60% of jobs in advanced economies and approximately 40% in emerging economies, including South Africa. The International Monetary Fund estimates that AI could affect 60% of jobs in advanced economies and approximately 40% in emerging economies, including South Africa.

For a country with an unemployment rate exceeding 32%, AI could represent an existential policy challenge. But it doesn't not have to be. Throughout history, from the Industrial Revolution, to emergence of the motor or software industries, the early days were marked with labour disruption. However, every innovation lead to productivity growth has led to explosive job creation. The World Economic Forum, which projects that while 85 million jobs may be displaced globally, 97 million new roles could emerge through technological maturationThe trick is to establish stabilisation policies early on, while disruptive industries mature.

CRITICAL STATISTIC: South Africa's Unemployment Context

South Africa's unemployment rate stood at 31.4% in Q4 2025. The IMF estimates AI could displace up to 40% of jobs in emerging economies. Proactive AI regulation and large-scale reskilling are not merely economic policy, but a matter of social stability.

1.2 South Africa's Current Policy Landscape

South Africa does not yet have a dedicated AI regulatory framework. The following legislation provides only partial, indirect coverage:

Existing Legislation	Relevance and Gaps
Protection of Personal Information Act (POPIA), 2013	Governs personal data processing; provides limited coverage of automated decision-making under Section 71, but does not address AI training data use or model-level harms.
Electronic Communications and Transactions Act (ECTA), 2002	Covers digital transactions but predates generative AI and algorithmic systems.
Cybercrimes Act, 2020	Addresses cybercrime but does not specifically regulate AI-generated malware or deepfakes.

Existing Legislation	Relevance and Gaps
Competition Act (as amended), 2019	Provides limited recourse against algorithmic collusion and data monopolisation.
Broad-Based Black Economic Empowerment Act	Does not address AI-enabled economic exclusion or algorithmic discrimination.
National Development Plan 2030	References the knowledge economy but contains no AI-specific commitments.

This legislative gap represents a significant governance risk. Without a dedicated AI framework, South Africa is exposed to regulatory arbitrage by foreign AI companies, unchecked deployment of harmful AI systems, and the erosion of constitutional rights through algorithmic decision-making.

1.3 Why South Africa Must Lead on the African Continent

South Africa is the most advanced digital economy in Sub-Saharan Africa and a founding member of the African Union. The AU's Continental AI Policy (2024) calls on member states to enact national AI frameworks by 2027. South Africa has both the opportunity and the obligation to lead the continental response - setting regulatory standards that protect African citizens, attract responsible AI investment, and project African values into global AI governance.

SCENARIO: The Regulatory Vacuum Risk

A major US-based AI hiring platform begins operating in South Africa, using algorithmic screening models trained predominantly on North American demographic data. The system consistently ranks candidates from townships and historically disadvantaged backgrounds lower than statistically equivalent candidates from former Model C schools.

Without an AI-specific anti-discrimination provision, the CCMA and Labour Court have no direct legal basis to compel algorithmic transparency or impose remedies. The employer claims proprietary protection over the model.

Under a future Artificial Intelligence Governance Act (AIGA), such systems would be classified as High-Risk AI, requiring pre-deployment bias auditing, algorithmic explainability, and mandatory human review.

2. National Security and AI

2.1 The AI-Enabled Threat Landscape

Artificial Intelligence fundamentally alters the national security calculus. It enables adversaries - state and non-state alike - to conduct operations of unprecedented scale, speed, and sophistication. South Africa, as a G20 member hosting critical continental financial infrastructure, diplomatic missions, and strategic natural resources, faces a material and growing AI-enabled threat environment.

Key National Security Threat Categories

Threat Category	Description
AI-generated disinformation and deepfakes	Targeting electoral processes, public officials, and social cohesion.
Autonomous and AI-assisted cyberattacks	Attacks on critical national infrastructure including Eskom, SARS, SARB, and Transnet.
AI-enabled foreign surveillance	Exploitation of unregulated data collection by foreign intelligence services.
Autonomous weapons procurement	Acquisition by non-state actors of dual-use AI with lethal capability.
AI-powered social engineering	Targeted attacks on government officials and SOE executives.
Backdoor procurement risks	Foreign-owned AI systems embedded in government procurement with undisclosed capabilities.

2.2 Prohibited AI Applications (Absolute Red Lines)

Inspired by Article 5 of the EU AI Act, this submission recommends that the future Artificial Intelligence Governance Act (AIGA) absolutely prohibit the following AI applications in South Africa:

Proposed Schedule of Absolute Prohibitions (AIGA)

Prohibited Application	Rationale
Social scoring systems	No public or private entity may deploy AI systems that score, rank, or classify citizens based on social behaviour, political opinion, or personal characteristics for purposes of differential access to services, opportunities, or rights.
Real-time biometric mass surveillance in public spaces	Prohibited without prior judicial authorisation and Parliamentary oversight, except in bona fide national security operations governed by RICA.
Subliminal AI manipulation techniques	AI systems that exploit subliminal mechanisms bypassing conscious cognition to produce commercial or political outcomes.

Prohibited Application	Rationale
AI targeting minors and cognitively impaired persons	Systems that exploit the vulnerabilities of children or persons with cognitive impairments for commercial or political purposes.
AI-generated non-consensual intimate imagery	Deepfake pornography, to be codified as a criminal offence under the Cybercrimes Act.
Autonomous lethal decision systems	Deployed without a responsible human commander in the decision loop.
AI-powered foreign interference tools	Any AI system deployed by or on behalf of a foreign state to influence South African public opinion, electoral outcomes, or government decision-making.

2.3 AI in Defence and Intelligence

The South African National Defence Force (SANDF) and State Security Agency (SSA) must be permitted to develop and deploy AI within a controlled, classified governance framework. This submission recommends:

Measure	Detail
Classified AI Governance Protocol	A dedicated protocol for SANDF, developed in consultation with the Joint Standing Committee on Intelligence.
Procurement prohibition	No AI defence systems to be procured from sanctioned or adversarially aligned states.
Mandatory security vetting	All AI systems deployed in government, National Key Points, or critical national infrastructure to be security vetted.
Cyber AI Response Unit	Established within SAPS and the SSA to address AI-enabled crimes.
Parliamentary disclosure	Mandatory in camera disclosure to Parliament of all AI systems deployed in intelligence and law enforcement operations.

SCENARIO: Deepfake Electoral Interference

Six weeks before the 2026 Local Government Elections, a highly realistic AI-generated video purporting to show a senior party leader accepting a bribe is widely circulated on WhatsApp and X. The video is indistinguishable from authentic footage. SABC, eNCA, and community radio stations unknowingly broadcast it.

The Electoral Court has no jurisdiction over AI content, and ICASA lacks the technical capability to respond.

Under the AIGA, this would constitute a Category 1 AI Security Offence, triggering mandatory content authentication protocols, a rapid-response AI Threat Assessment, and prosecutorial referral to the NPA with maximum penalties of R50 million or 10 years' imprisonment.

2.4 Foreign AI Systems in Government Procurement

The acquisition of AI systems by organs of state from foreign suppliers presents unique national security risks. This submission recommends the following procurement safeguards:

Safeguard	Requirement
National Security AI Assessment (NSAIA)	All AI systems procured by government must undergo a NSAIA conducted by the AI Regulatory Authority in consultation with the SSA.
Supplier disclosure	Suppliers of government AI systems must disclose nationality of ownership, training data provenance, and any foreign government relationships.
Strategic threat restriction	AI systems classified as critical infrastructure AI may not be owned, operated, or significantly controlled by entities from states designated as strategic threats.
Domestic preference	Government AI procurement must prioritise domestically developed solutions and SADC-origin systems where technically comparable alternatives exist.

3. Data Residency and Digital Sovereignty

3.1 The Stakes of Data Sovereignty

AI feeds on data. The entity that controls the data upon which AI systems are trained and upon which they operate exercises profound power over the inferences, decisions, and predictions those systems produce. Permitting unrestricted cross-border transfer of South African personal and national data to foreign AI companies constitutes a form of digital colonialism - one that extracts value from South African society while concentrating economic power and strategic intelligence in foreign jurisdictions.

South Africa currently has no mandatory data localisation requirement, relying instead on POPIA's Section 72 cross-border transfer conditions. This is insufficient for the AI era, where training data volumes, model weights, and inference logs contain aggregate societal intelligence that transcends individual data-subject rights.

Comparative International Data Sovereignty Approaches

Jurisdiction	Data Sovereignty Approach
China	Mandates localisation of data collected within China for AI systems deployed to Chinese users.
India	Digital Personal Data Protection Act (2023) creates strong localisation requirements for 'significant data fiduciaries'.
European Union	GDPR cross-border transfer regime (adequacy decisions and SCCs) represents a middle path that South Africa should adapt and strengthen.
South Africa (current)	Relies on POPIA Section 72 cross-border transfer conditions only - insufficient for the AI era.

3.2 Proposed Tiered Data Residency Framework

This submission recommends a tiered data residency regime calibrated to the sensitivity of the data category - all figures simply represent a starting point and must be establish after expert review:

Tier	Data Category	Residency Requirement	Consequence of Non-Compliance
Tier 1 - National Security	Data generated by or relating to SANDF, SSA, SAPS, NPA, and designated National Key Points	Absolute localisation; no cross-border transfer permitted	Criminal sanction; R100m penalty
Tier 2 - Critical Infrastructure	Data generated by Eskom, PRASA, SANRAL, SARB, and major financial institutions	Must be stored in South Africa; transfer requires AI Regulatory Authority authorisation	R50m penalty; licence revocation

Tier	Data Category	Residency Requirement	Consequence of Non-Compliance
Tier 3 - Government and Public Sector	All data collected or processed by organs of state	Primary copy in South Africa; secondary copies require SCC equivalent	R20m penalty; contract cancellation
Tier 4 - Personal Data (AI-processed)	Personal data used to train or operate AI systems affecting South African citizens	Transfer permitted with POPIA Section 72 compliance and AI Regulatory Authority notification	POPIA enforcement plus R10m AI penalty
Tier 5 - Anonymised Commercial AI Data	Aggregated or anonymised AI operational and commercial data	Free flow, subject to re-identification risk assessment	Administrative fine up to R5m

3.3 Sovereign AI Infrastructure

South Africa must invest in public AI infrastructure to avoid permanent dependency on foreign hyperscale cloud providers. Key recommendations include:

SCENARIO: The Cloud Dependency Risk

During the 2025 load-shedding crisis, Eskom's AI-based predictive load management system - hosted on AWS servers in Ireland - became inaccessible for 11 hours due to an AWS regional outage unrelated to South Africa. Emergency response was delayed and manual override procedures failed.

The AIGA would mandate that all Critical Infrastructure AI must have primary processing and failover capability on South African soil.

Proposed Initiative	Description
State AI Infrastructure Corporation (SAIIC)	Modelled on Saudi Arabia's SDAIA, to develop and operate nationally owned compute infrastructure.
National AI Processing Zones	At least two government-grade data centres with guaranteed uptime, security certification, and preferential pricing for public sector and SMME use.
Section 12B Tax Incentives	Incentivisation of private investment in South African data centres through the Income Tax Act, extended to AI compute infrastructure.
SADC Regional Data Compact	Negotiated through DIRCO, establishing mutual recognition of data governance standards and regional data processing hubs to serve the continent.

4. Privacy, Data Protection and AI

4.1 POPIA and the AI Gap

The Protection of Personal Information Act (POPIA) represents South Africa's foundational data protection legislation and provides a basis for AI regulation. However, POPIA was designed for the pre-AI era and contains critical gaps when applied to contemporary AI systems:

POPIA Provision	AI-Specific Gap
Section 71 - Automated Decision-Making	Section 71 does allow data subjects to make representations about automated decisions that significantly affect them, and to request human review. However, it does not require proactive explainability, pre-deployment impact assessments, or transparency about training data composition - leaving material gaps specific to AI-era automated decision-making
Consent Framework	POPIA's consent model does not address the use of publicly available data to train AI models, or the re-use of historical data for new AI applications.
Information Regulator	Has limited technical capacity to audit AI systems; no AI-specific investigative powers.
Data Subject Rights	No specific right to explanation of AI decisions; no right to contest automated profiling.
Data Minimisation	AI training often requires massive data volumes; tension with minimisation principles remains unresolved.
Purpose Limitation	Foundation models trained on broad internet data do not have a defined 'purpose' at the time of data collection.

4.2 Proposed AI-Specific Privacy Provisions

This submission recommends that the AIGA introduce the following AI-specific privacy provisions, operating in harmony with POPIA:

4.2.1 Right to Algorithmic Explanation

Any natural person subject to a consequential AI-assisted decision - including credit, insurance, employment, benefits, healthcare, or criminal justice determinations - shall have the right to receive a meaningful, plain-language explanation of:

(a) the fact that AI was used;
(b) the principal factors influencing the outcome; and
(c) the process for human review.

4.2.2 AI Data Protection Impact Assessments (AI-DPIA)

Any organisation deploying a High-Risk or Critical-Risk AI system must conduct and register an AI-DPIA with the Information Regulator prior to deployment. The AI-DPIA must assess: training data provenance and representativeness; risk of discrimination or bias; data minimisation compliance; and retention and deletion protocols for training data.

In addition, and specifically for AI systems that have been fine-tuned through Reinforcement Learning from Human Feedback (RLHF) or equivalent techniques, the AI-DPIA must include a dedicated Proximal Policy Optimisation (PPO) Transparency Disclosure, setting out:

(a) the design and training data of the reward model used to generate PPO training signals;
(b) the demographic and geographic profile of annotators whose preferences shaped the reward model;
(c) the PPO objective function parameters, including any constraints, clipping ratios, or safety penalties applied; and
(d) the material behavioural guardrails installed through PPO and their functional effect on model outputs in South African deployment contexts. The PPO Transparency Disclosure must be reviewed and approved by the Information Regulator as part of the AI-DPIA registration process.

4.2.3 Training Data Consent Framework

The use of personal data to train AI systems constitutes a new processing purpose requiring separate consent or a valid alternative lawful processing ground. Organisations may not rely on original collection consent to justify AI training data use.

This principle extends to the use of personal data or personally derived preference signals in the training of reward models used in Proximal Policy Optimisation (PPO) fine-tuning: where human annotators' ratings of AI outputs are derived from or informed by personal data about South African consumers, that data use constitutes a new processing purpose and must be separately authorised.

AI developers must therefore disclose, within their AI-DPIA submissions, whether personal data was used at any stage of the PPO training pipeline - including reward model training - and the legal basis for that use.

A dedicated Regulatory Code of Practice for AI Training Data - including PPO reward model data practices - should be developed by the Information Regulator within 18 months of AIGA commencement.

4.2.4 Biometric AI Protections

The processing of biometric data - including facial geometry, voiceprint, gait analysis, and emotional state inference - by AI systems shall be treated as a special category requiring explicit opt-in consent, or a judicial order in law enforcement contexts. Biometric AI databases maintained by private entities shall be subject to quarterly audits.

> **COMPARATIVE NOTE: Illinois Biometric Information Privacy Act (BIPA)**
>
> BIPA imposes strict consent requirements for biometric data processing and has resulted in over USD 1.5 billion in class action settlements in the United States. South Africa should adopt equivalent protections before commercial biometric AI becomes widespread.

SCENARIO: Facial Recognition and Racial Bias

A major South African retail chain deploys facial recognition AI across 200 stores to identify persons listed on internal fraud watchlists. The system has a 23% false positive rate for darker-skinned individuals due to training data imbalance - disproportionately flagging Black customers. Store security confronts incorrectly identified persons, resulting in public humiliation and reputational harm.

Under existing law, the retailer faces only a possible POPIA complaint. Under the AIGA, this deployment would constitute an unauthorised High-Risk AI application, require mandatory suspension, trigger an investigation with powers to impose fines of up to R10 million per affected individual, and compel algorithmic bias remediation.

4.3 Children and Vulnerable Persons

AI systems that process data about children, elderly persons, persons with disabilities, or persons in financial distress shall be subject to the highest level of privacy protection:

Measure	Requirement
Advertising to minors	AI-powered targeted advertising to persons under 18 based on profiling shall be prohibited.
Educational AI platforms	Must obtain parental consent and limit data use strictly to educational purposes.
Child welfare AI	AI systems deployed in social welfare, foster care, or criminal justice contexts involving minors must undergo Child Rights AI Impact Assessments.

5. Consumer Protection in the AI Economy

5.1 The AI Consumer Risk Landscape

South African consumers are increasingly exposed to AI systems in contexts where power asymmetries are significant, technical literacy is limited, and redress mechanisms are inadequate. The Consumer Protection Act, 2008 (CPA) provides foundational protections but does not address AI-specific harms including algorithmic deception, AI-generated dark patterns, chatbot impersonation, and dynamic AI pricing.

5.2 Transparency and Disclosure Requirements

Requirement	Description
AI identity disclosure	Mandatory disclosure to consumers when interacting with an AI system rather than a human, with no exception for commercial or competitive sensitivity.
South African AI Content Label (SAICL)	AI-generated content - including marketing materials, news articles, financial advice, and product descriptions - must be clearly labelled.
Decision-time disclosure	AI systems used in customer-facing decision-making (credit, insurance, debt collection) must disclose their use to the affected consumer prior to or at the time of the decision.
Chatbot identification	Virtual assistants deployed by financial services providers, healthcare platforms, and government must identify themselves as AI and disclose their capabilities and limitations.

5.3 Prohibited Consumer AI Practices

Prohibited Practice	Description
AI-powered dark patterns	User interface designs driven by AI that manipulate consumers into unintended purchases, subscriptions, or disclosures.
Dynamic pricing discrimination	AI systems that charge higher prices to consumers based on inferred vulnerability, urgency, or limited alternatives - for example, surge pricing for essential medicines or emergency transport.
Predatory AI lending	Automated credit and lending systems targeting over-indebted consumers with algorithmically optimised high-interest products.
Synthetic impersonation	AI systems impersonating government agencies, banks, or trusted institutions to extract personal information or financial payments.
AI-generated fake reviews	The use of AI to generate fraudulent product or service reviews at scale.

CONSUMER ALERT: Predatory AI Lending

Predatory AI lending apps targeting low-income South Africans via WhatsApp and informal digital channels have been identified in the Western Cape and Gauteng. These apps use AI to assess creditworthiness from mobile phone metadata without explicit consent. Under the AIGA, this constitutes both a privacy violation and a prohibited consumer AI practice.

5.4 AI and Financial Services Consumer Protection

Financial services represent the highest-risk consumer AI deployment environment in South Africa. The consequences of AI failures in this sector are not just an inconvenience - they are financially catastrophic for individuals, structurally destabilising for the market, and constitutionally impermissible where they entrench historical patterns of economic exclusion.

5.4.1 The Structural Bias Problem: Why LLMs Are Not Neutral Financial Advisors

A foundational assumption embedded in most commercial AI financial services products is that the underlying AI model is broadly neutral, objective, and universally applicable. This assumption is demonstrably false, and its falsity has specific, measurable consequences for South African financial consumers. Two compounding bias mechanisms are of particular regulatory concern.

5.4.1(a) The Double-Layer Bias Problem

LLMs acquire their financial knowledge and reasoning patterns from datasets overwhelmingly consisting of English-language internet content, predominantly originating from the United States, United Kingdom, and Western Europe.

This creates a foundational Western financial bias: the models' understanding of creditworthiness, investment logic, insurance risk, debt management, and financial planning reflects the regulatory environments, economic structures, social welfare systems, and consumer behaviours of developed Western economies - not of South Africa's dual economy characterised by high informality, limited banking penetration, irregular income patterns, and township-based economic activity.

This Western bias is compounded by a second layer: the training corpora of LLMs dramatically over-represent non-professional financial content producers relative to qualified financial professionals. Financial content on the internet is overwhelmingly generated by bloggers, social media influencers, and YouTube creators - not by qualified financial advisers, actuaries, economists, or regulators.

Financial-themed content channels on YouTube consistently generate the highest revenue per thousand views (RPM) of any content category, creating powerful economic incentives for non-professional content creators to produce high-volume financial content designed to maximise engagement rather than accuracy.

TECHNICAL BRIEFING: The Double-Layer Bias: Regulatory Significance

The double-layer bias is not additive - it is multiplicative. The Western training bias pre-loads the model with inapplicable financial assumptions. The content ecosystem distortion then selects for the most internet-viral versions of those already-biased assumptions.

A South African consumer receiving AI-generated financial advice from such a system is receiving advice shaped by neither South African law, nor South African economic conditions, nor qualified South African professional judgement - but by what generates the most views on American financial YouTube channels.

This is not a bug that can be patched. It is an architectural characteristic of how these systems learn.

5.4.1(b) RLHF Cultural Misalignment: The Human Annotator Problem

The dominant technique used to fine-tune LLMs - Reinforcement Learning from Human Feedback (RLHF) - relies on human annotators to review model outputs and provide preference ratings.

5.4.1(b)(i) Proximal Policy Optimisation (PPO) and the Guardrail Transparency Gap

Central to how RLHF is implemented in practice is a specific machine learning algorithm known as Proximal Policy Optimisation (PPO). PPO is the reinforcement learning algorithm that translates human annotator preference ratings into actual changes to a model's behaviour.

After human annotators rate model outputs, a separate "reward model" is trained on those ratings; PPO then uses that reward signal to update the AI model's parameters - reinforcing outputs that annotators preferred and suppressing those they did not. PPO functions as the primary guardrail mechanism that shapes how AI models behave: what they say, what they refuse to say, and how they frame responses. In the financial services context, PPO determines whether a model gives conservative or aggressive credit advice, whether it flags risky products or promotes them, and whether it directs consumers toward regulated South African financial pathways or defaults to Western financial assumptions.

The regulatory significance of PPO is that it operates as an opaque optimisation process. Neither consumers nor regulators can observe the reward model's design, the preference data on which it was trained, or the PPO objective function that shaped the model's guardrails.

The same cultural misalignment problems that affect RLHF annotator data are therefore embedded and amplified by PPO into the model's core behavioural constraints - without disclosure, without audit, and without any South African regulatory oversight. This submission recommends that transparency into PPO parameters, reward model design, and PPO training data should be a mandatory disclosure requirement for any High-Risk AI system deployed in South Africa, as further detailed in Sections 4.2.2, 5.4.2(c), 7.3, and 10.1.

The structural problem is this: RLHF annotators are overwhelmingly not South African. They are predominantly employed in the United States, India, and Kenya, reflect the cultural, economic, and value frameworks of those countries, and have no formal financial qualifications, no

understanding of South African financial law, and no lived experience of the South African economic environment they are being asked to evaluate.

SCENARIO: RLHF Bias in Practice: Credit Counselling

A South African consumer earning R8,500 per month through a combination of part-time retail work, informal spaza shop income, and irregular piece-work queries an LLM-powered financial wellness app about managing a R45,000 debt across three micro-lenders.

The model recommends consolidating the debt into a single personal loan and cutting discretionary spending - advice calibrated for a salaried employee with a formal credit record and access to mainstream banking products. The model makes no reference to the National Credit Act's debt relief provisions, the National Credit Regulator's debt counselling pathway, or the consumer's rights under Section 86 of the NCA.

The consumer consolidates into a fourth micro-lender at 29.8% interest and deteriorates further into over-indebtedness. Under the proposed AIGA, this AI system would constitute an unlicensed financial advisory service operating in violation of the Financial Advisory and Intermediary Services Act.

5.4.2 Regulatory Framework for AI in Financial Services

Section 5.4.2 is likely to be the most contested provision in this proposal. The Portfolio Committee can expect strong pushback from the world's largest multinational AI companies. Technology blackboxes represent the greatest AI risk to our country. These proposals require careful expert consultation.

5.4.2(a) Licensing and Regulatory Classification

No AI system may be deployed to South African consumers for the purpose of financial advice, credit assessment, insurance underwriting, investment recommendation, or related financial services without prior authorisation from both the FSCA and the AI Regulatory Authority.

AI Application Type	Regulatory Requirements	Oversight Authority
Robo-Advisory / AI Financial Planning	FAIS Category I/II licence + High-Risk AI registration + Bias Audit Certificate	FSCA + AI Regulatory Authority
AI Credit Scoring / Decisioning	NCA compliance certification + High-Risk AI registration + NCR notification	FSCA + AI Regulatory Authority + NCR
AI Insurance Underwriting and Pricing	High-Risk AI registration + Actuarial Fairness Certificate + PEPUDA proxy audit	FSCA + AI Regulatory Authority
AI-Powered Financial Chatbots	Limited Risk registration	AI Regulatory Authority
LLM-Based Financial Content Tools	AI Regulatory Authority notification + Content Source Disclosure + Western Corpus Warning Label	AI Regulatory Authority

5.4.2(b) Mandatory South African Financial Corpus Requirement (SAFC)

Any AI system authorised to provide financial services to South African consumers must demonstrate that its training corpus incorporates a defined minimum weighting of South African-specific financial knowledge from authoritative sources. Systems that cannot demonstrate a minimum 35% weighting of South African authoritative financial content shall not be authorised for consumer-facing financial service deployment.

Mandatory South African Financial Corpus (SAFC) Requirements

These specific proposals are starting points for expert engagement rather than demands.

Source Category	Minimum Corpus Requirement
National Credit Act (NCA) and all subordinate regulations	Full statutory and regulatory corpus
FSCA and NCR regulatory guidance, circulars, and determinations	Complete official publications
SARB Monetary Policy Committee statements, Governor speeches, and prudential standards	Complete official publications
JSE Listing Requirements, market conduct rules, and investor protection guidance	Complete official publications
South African consumer financial education materials (accredited by FSCA)	Minimum 50,000 document corpus
SARS tax guidance relevant to personal and SMME financial planning	Full SARS public guidance library
Ombudsman determinations (Banking Services, FAIS, Short-Term Insurance)	Minimum 5 years of published decisions
Peer-reviewed South African financial and economic research	Minimum 10,000 academic document corpus
South African informal economy and stokvels research	Minimum 5,000 document corpus

5.4.2(c) RLHF Cultural Alignment Requirements

Requirement	Detail
South African Annotator Minimum	Minimum 30% of total annotator hours for the South African financial services fine-tuning stage must be South African citizens or permanent residents with demonstrable knowledge of the South African financial regulatory environment.
Annotator Qualification Standards	RLHF annotators must demonstrate working knowledge of the NCA, FAIS, the FSCA's Treating Customers Fairly (TCF) framework, and relevant South African consumer protection law.
RLHF Audit Trail Obligation	Developers must maintain complete records of RLHF training sessions used for South African financial services fine-tuning, including annotator demographics, qualification assessments, inter-annotator agreement rates, and preference decision logs.

Requirement	Detail
Cultural Scenario Testing	Prior to authorisation, AI financial systems must pass a South African Cultural Financial Scenario Test (SACFST) covering informal income verification, stokvels and burial societies, migrant remittances, township-based SMMEs, over-indebtedness under NCA Section 86, and rural community financial access.
RLHF Refresh Cycle	South African financial AI authorisations shall be subject to an 18-month renewal cycle requiring evidence of updated RLHF fine-tuning incorporating South African annotator feedback.
PPO Parameter Disclosure	The deploying entity must submit, as part of its authorisation application and each 18-month renewal, a PPO Parameter Disclosure Statement prepared by an accredited AI auditor, describing: the reward model architecture and training data; the PPO clipping parameter (epsilon) and its calibration rationale; KL-divergence penalties applied to constrain model drift; and all safety-critical reward shaping applied to South African deployment contexts.
PPO Reward Model Cultural Audit	The reward model used in PPO fine-tuning for South African financial services deployment must be audited to confirm that South African annotator preference data was materially weighted in reward model training. The audit must also assess whether the PPO guardrails direct consumers toward NCA Section 86 debt counselling, FSCA-regulated products, and NCR-compliant financial processes, rather than Western-default financial pathways.
PPO Behavioural Constraint Register	SAIRA shall maintain a non-public PPO Behavioural Constraint Register documenting the substantive guardrails installed through PPO in each authorised AI financial services system. This Register shall be available to the FSCA SupTech Unit under formal cooperation arrangements and to the Information Regulator for AI-DPIA verification.

INTERNATIONAL PRECEDENT: A Global First

No jurisdiction has yet introduced annotator demographic requirements as a regulatory condition for financial AI deployment. South Africa would be the first globally to codify this requirement - a pioneering step that addresses a structural training deficiency the global AI governance community has identified but not yet regulated. This represents an opportunity for South Africa to set an international standard that protects its own consumers while providing a model for other emerging economies facing the same structural bias challenge.

5.4.2(d) Bias, Fairness and Actuarial Auditing

AI Financial Services Bias Audit Schedule

Audit Type	Timing	Conducted By	Scope
Pre-Authorisation Audit	Before deployment	Accredited independent AI auditor	Full demographic bias test across South African profiles
Post-Authorisation Audit	12 months after deployment	Accredited independent AI auditor	Full re-audit with updated South African population sample
Quarterly Monitoring Report	Every 3 months	Internal AI governance function + AI Regulatory Authority submission	Statistical distribution analysis across demographic proxies
Event-Triggered Audit	Within 30 days of complaint, significant output change, or model update	AI Regulatory Authority-directed or independent	Targeted audit of flagged bias dimension
RLHF Refresh Validation	At 18-month authorisation renewal	Accredited auditor + South African annotator verification	Confirmation of South African cultural alignment maintenance

5.4.2(e) Enforcement, Penalties and Remediation

Specific figures - penalty levels, percentage thresholds, FTE requirements - are presented as proposals for consultation rather than fixed positions. All figures are starting points for expert engagement rather than demands.

Violation	Penalty	Remediation Obligation
Deployment without authorisation	Criminal offence + R50m administrative penalty per AI system	Mandatory suspension; consumer notification
Failure to meet SAFC Requirement (35% threshold)	R20m per financial year of non-compliance	Mandatory system suspension pending compliance; Western Corpus Warning Label mandatory
Failure to meet RLHF annotator requirements	R15m per authorisation cycle	Revocation of authorisation until re-training demonstrated
Discriminatory AI output affecting 100+ consumers	R5m per affected consumer category + mandatory audit	Recalculation and remediation for all affected consumers at deployer's cost
Failure to provide algorithmic explanation on request	R500,000 per incident + R5,000 statutory damages per consumer	Explanation to be provided within 5 business days; FSCA referral

Violation	Penalty	Remediation Obligation
AI system failure causing material financial harm	R25m + civil liability for actual harm caused	Mandatory consumer restitution fund; public reporting
Non-submission of quarterly bias monitoring report	R1m per month of non-compliance	FSCA licence condition suspension after 3 months
Failure to disclose PPO parameters and reward model design	R10m administrative penalty per authorisation cycle + mandatory suspension of authorisation pending full PPO Transparency Disclosure; SAIRA may publish summary of non-disclosed guardrails in Annual AI State of the Nation Report.	Mandatory system suspension pending compliance

SCENARIO: Insurance AI Discrimination

A major South African insurer deploys an AI underwriting model that uses telematics, home address data, and social media signals to price motor insurance. Analysis reveals that residents of townships receive premium quotes 34% higher than residents of comparable income in suburban areas - a proxy for racial discrimination. The insurer claims the model uses no racial data directly.

Under the AIGA's AI fairness provisions and PEPUDA, the use of racially correlated proxies constitutes indirect algorithmic discrimination. The AI Regulatory Authority may order an algorithmic audit, mandatory recalculation of affected premiums, and publication of the findings.

5.4.3 Special Protections for Vulnerable Financial Consumers

Vulnerable Group	Required Protections
Over-indebted consumers	AI systems deployed in debt counselling must comply with NCA Section 86 debt counselling protections and NCR requirements.
Informal economy workers	AI credit and insurance systems must be capable of assessing applications from consumers with irregular, multi-source, or informally documented income. Systems that systematically decline applications from consumers without formal payslips must demonstrate the non-discriminatory basis for this approach.
Rural and digitally excluded consumers	AI financial systems deployed in rural areas or on USSD/feature phone platforms must be specifically tested for performance parity with urban/smartphone deployments.
Elderly consumers (60+)	Enhanced suitability safeguards apply. Recommendations for complex financial products to elderly consumers must trigger mandatory human review.

Vulnerable Group	Required Protections
First-time credit applicants and youth	AI credit systems must not use absence of credit history as a negative scoring variable without offsetting alternative data sources.

5.5 Access and Digital Inclusion

Measure	Description
Universal Service Obligation extended to AI	ICASA shall require all major AI platform operators to provide accessible, multilingual (including all 11 official languages), and data-light AI services.
AI Accessibility Standards	All AI systems deployed by government must comply with accessibility standards for persons with disabilities under the UN Convention on the Rights of Persons with Disabilities.
Rural AI Access Fund	A levy on AI platform operators with annual South African revenue exceeding R500 million shall fund AI access infrastructure in underserved communities.

6. Fair Use, Intellectual Property and AI-Generated Content

6.1 The AI Copyright Challenge

Generative AI systems - including large language models, image generators, and music composition tools - are trained on vast corpora of human-created works, frequently without the consent of rights holders, without compensation, and without attribution. This presents a fundamental challenge to South Africa's creative economy. The SACO 2020 mapping study found that creative economy employment accounted for 7% of all jobs in South Africa in 2017, about 1.1 million jobs. More critically, the SACO 2022 mapping study found that the total contribution of the cultural and creative industries to South Africa's GDP was R161 billion in 2020 - representing just under 3% of South Africa's total economic production

Current South African copyright law - principally the Copyright Act 98 of 1978 and the Performers' Protection Act 11 of 1967 - does not adequately address AI-specific IP issues. The Copyright Amendment Bill, which has been before Parliament since 2017, must be updated to incorporate AI-specific provisions urgently.

6.2 Copyright in AI Training Data

Issue	Recommended Position
Text and Data Mining (TDM) Exception	South Africa should introduce a limited, non-commercial TDM exception permitting AI training on publicly available data for research and educational purposes - modelled on EU CDSMD Article 4, but with an opt-out mechanism for rights holders.
Commercial AI Training	Commercial AI developers must obtain licences or pay equitable remuneration when training AI systems on copyrighted South African works. The Copyright Tribunal shall have jurisdiction to determine fair remuneration in disputed cases.
Transparency Obligation	AI developers must disclose the categories of copyrighted works used in training upon request by a rights holder or by the AI Regulatory Authority.

GLOBAL PRECEDENT: International Developments

Significant litigation on AI training data copyright is actively progressing in both the United Kingdom and the United States - including major cases involving generative AI developers - though definitive judicial precedents have not yet been fully established in either jurisdiction. The EU AI Act, which is in force, requires high-risk AI providers to publish training data summaries. South Africa should monitor these developments closely and proactively codify clear domestic positions before foreign precedents become the de facto standard applied to AI operating in South Africa.

6.3 Ownership of AI-Generated Works

Principle	Detail
AI systems cannot hold copyright	This principle shall be codified in the Copyright Amendment Act.
Human authorship requirement	Copyright in AI-assisted works shall vest in the human author(s) who made sufficiently original creative contributions. Pure AI output with no meaningful human creative input shall enter the public domain.
AI-generated commercial works	Where an AI system generates commercially valuable output in the course of a commercial enterprise, the employer or commissioning party shall own copyright in the output, subject to rights of the human operators who provided significant creative direction.
Digital Content Creator Rights Charter	To be developed in consultation with CIPRO, SAMRO, and the creative industries sector.

6.4 Deepfakes, Synthetic Media and Personality Rights

Proposed Measure	Description
Performers' Protection Act amendment	Extend performers' rights to their AI-synthesised likeness and voice, prohibiting commercial exploitation without consent.
Tort of AI-based Identity Misappropriation	Enable natural persons to sue for damages when their likeness, voice, or persona is reproduced by AI without consent for commercial or defamatory purposes.
Mandatory provenance metadata	C2PA standards to be embedded in all AI-generated media distributed on South African platforms with more than 100,000 monthly active users.
Non-consensual deepfake intimate imagery	Criminalised as a Category 1 AI Offence under the AIGA with a minimum sentence of five years' imprisonment.

SCENARIO: Musical Artist Voice Cloning

A South African music producer uses a commercially available AI voice cloning service to produce and commercially release 12 tracks in the voice of a nationally recognised artist - without consent and without payment. The artist's record label has no legal remedy under the current Performers' Protection Act, which covers only recorded performances, not synthesised AI voices.

Under the proposed AIGA amendments, the producer commits a civil and criminal offence, the platforms distributing the content are required to take it down within 24 hours of notification, and the artist is entitled to damages equal to three times the commercial value of the releases.

7. Algorithmic Accountability and Transparency

7.1 The Black Box Problem

Modern AI systems - particularly deep neural networks and large language models - operate as statistical systems whose decision logic is not readily interpretable by humans. This “black box” characteristic poses fundamental challenges to accountability, due process, and the rule of law.

The opacity operates at multiple levels: the base model’s pre-training on billions of text tokens is not auditable by design, or more accurately, computationally infeasible to audit. The data is there, but the sheer scale (trillions of parameters) makes it impossible for a human to trace a specific output back to a specific training token.

The RLHF fine-tuning process is implemented through reinforcement learning algorithms - of which Proximal Policy Optimisation (PPO) is the most historically prominent and widely documented, though newer approaches including Direct Preference Optimisation (DPO) are increasingly used. Regardless of the specific algorithm, the parameters governing this process - including reward model design, clipping ratios, and KL-divergence penalties - are typically treated as proprietary by AI developers, yet substantially influence what an AI system will and will not say, what advice it generates, and what risks it surfaces or suppresses

For regulatory purposes, PPO transparency is therefore as important as RLHF annotator disclosure - it is the mechanism through which human preferences (themselves culturally and demographically biased) are algorithmically embedded as permanent behavioural constraints. When an AI system denies a person a mortgage, flags a taxpayer for audit, scores a welfare applicant, or informs a sentencing recommendation, the absence of explainability potentially violates private sector AI decisions under the Consumer Protection Act, the NCA, or POPIA; or in public, foundational principles of administrative justice enshrined in the Promotion of Administrative Justice Act (PAJA) and the Constitution.

7.2 Risk-Based Regulatory Classification

Proposed Four-Tier AI Risk Classification Framework

Risk Tier	Description	Regulatory Requirements	Examples
UNACCEPTABLE RISK (Prohibited)	AI systems that pose fundamental threats to constitutional rights and democratic values	Absolute prohibition; criminal sanction	Social scoring, mass biometric surveillance, manipulation systems, autonomous weapons
HIGH RISK (Regulated)	AI systems with significant consequences for persons' rights, safety, or livelihood	Prior authorisation, mandatory auditing, human oversight, explainability, registration with AI Regulatory Authority	Credit scoring, employment screening, benefits determination, criminal justice, healthcare diagnosis, educational assessment, immigration
LIMITED RISK (Transparency Obligations)	AI systems interacting with users in consequential ways	Disclosure obligation, labelling, user rights notification	Chatbots, AI content generation tools, recommendation engines, sentiment analysis

Risk Tier	Description	Regulatory Requirements	Examples
MINIMAL RISK (Voluntary Codes)	AI systems with low potential for harm	Voluntary industry codes of practice; AI Regulatory Authority guidance	Spam filters, AI chess, simple automation, AI editing tools

7.3 Algorithmic Impact Assessments (AIA)

All High-Risk AI systems must undergo a mandatory Algorithmic Impact Assessment (AIA) prior to deployment and at regular intervals thereafter. The AIA framework shall require:

AIA Component	Requirement
Technical documentation	Design, capabilities, limitations, and training data of the AI system. For systems fine-tuned using Reinforcement Learning from Human Feedback (RLHF), technical documentation must include a Proximal Policy Optimisation (PPO) Transparency Statement disclosing: (i) the reward model architecture and its training data provenance; (ii) the PPO algorithm parameters used in fine-tuning, including the clipping ratio (epsilon), entropy bonus, and value function architecture; (iii) the substantive behavioural guardrails installed through PPO and the policy rationale for each; and (iv) the KL-divergence constraint applied to limit drift from the reference model. For High-Risk AI systems deployed in South Africa, the PPO Transparency Statement must be certified by an accredited independent AI auditor.
Fairness analysis	Statistical evidence of fairness across protected characteristics under the Equality Act.
Risk assessment	Foreseeable misuse, unintended consequences, and failure modes.
Human oversight protocol	Specification of when and how human judgement overrides AI recommendations.
Monitoring and incident reporting plan	Ongoing monitoring framework and incident escalation procedures.
Public summary	Public disclosure of AIA findings, except for validly classified national security systems.

7.4 AI in Government Administration

The use of AI in administrative decision-making by organs of state raises the most acute rule-of-law concerns. This submission proposes a Government AI Charter, enacted as a Schedule to the AIGA, incorporating the following principles:

Charter Provision	Requirement
Government AI Principles	Legality, Fairness, Reasonableness, Procedural Propriety, and Proportionality shall govern all AI deployments in public administration.
No fully automated government decisions	No organ of state may make a final binding administrative decision affecting a citizen's rights using a fully automated AI system without human review and sign-off.
AI Procurement Register	All AI systems deployed by organs of state shall be listed in a publicly accessible AI Procurement Register maintained by the OCPO.
AI in Social Grants	SASSA may use AI for fraud detection and eligibility screening, but any decision to suspend, reduce, or cancel a grant must be reviewed by a human official and communicated with reasons.
AI in Criminal Justice	NPA, SAPS, and the Judiciary may use AI as advisory tools only. AI risk assessment tools in bail applications, sentencing, and parole decisions must be disclosed to defendants and subject to challenge.

SCENARIO: AI in Social Grant Administration

SASSA deploys an AI system to screen for fraud in social grant applications, trained on historical fraud patterns. The system flags approximately 340,000 legitimate pensioners for potential fraud - disproportionately targeting rural elderly beneficiaries who have irregular banking histories (a known but legitimate pattern).

Under the AIGA's Government AI Charter, this scenario would be unlawful. SASSA would be required to notify affected persons with reasons, restore suspended grants pending human review, and submit a Systemic AI Incident Report to the AI Regulatory Authority.

8. AI in Critical National Infrastructure

8.1 The Infrastructure AI Risk

South Africa's critical national infrastructure - energy, water, transport, telecommunications, financial systems, and healthcare - is increasingly dependent on AI for operational efficiency, predictive maintenance, and demand management. This dependency creates single points of failure that adversarial actors, natural disruptions, or AI system errors could exploit with catastrophic national consequences.

8.2 Critical Infrastructure AI Sectors

Sector	Covered Entities
Energy	Eskom, NERSA-regulated independent power producers, and municipal electricity distributors.
Water	DWS-operated and municipal water infrastructure, including AI-managed treatment and distribution systems.
Financial	SARB payment infrastructure, JSE trading systems, and systemically important financial institutions.
Transport	Air Traffic and Navigation Services (ATNS), PRASA, SANRAL intelligent transport systems.
Telecommunications	Licensed electronic communications network service providers operating national backbone infrastructure.
Healthcare	National Health Laboratory Service (NHLS), district health information systems, and emergency medical services AI.
Government	SITA-managed government IT systems, SARS eFiling infrastructure, HANIS, and Home Affairs databases.

8.3 Tiered NKP-AI Classification System

Tier	Description	Examples
Tier Alpha - Existential Infrastructure AI	AI systems whose compromise or failure could cause loss of life, national economic collapse, or irreversible damage to state function.	Eskom grid management; ATNS air traffic control; SARB payment settlement; water treatment process control; SAPS/NPA criminal justice AI with real-time decision authority.
Tier Bravo - Strategic Infrastructure AI	AI systems whose compromise would cause severe but recoverable national disruption.	PRASA and SANRAL transport management AI; systemically important financial institution AI; SITA government network management AI; NHLS diagnostic AI; municipal bulk utility management AI.

Tier	Description	Examples
Tier Charlie - Essential Services AI	AI systems whose compromise would cause significant localised disruption.	District health AI; provincial government administrative AI; SASSA grant disbursement AI; major port and logistics AI.

8.4 Security Architecture Requirements

Requirement	Standard
Air-Gap and Network Segmentation	All Tier Alpha NKP-AI systems must operate on physically isolated networks. Data exchange with external systems must occur only through controlled, one-directional data diodes with cryptographic integrity verification.
Domestic Processing - Absolute	All NKP-AI inference, training, and model updates must be processed exclusively on infrastructure physically located within South African territory. No exception, waiver, or emergency deviation is permitted for Tier Alpha systems. This requirement extends explicitly to PPO fine-tuning: the Proximal Policy Optimisation algorithms and reward models used to install behavioural guardrails in NKP-AI systems must be trained and executed exclusively on South African sovereign infrastructure. The PPO reward model - which determines what an AI system will and will not do in operational contexts - must be designed, trained, and validated by personnel holding appropriate security clearances. No foreign-designed, foreign-trained, or foreign-operated PPO reward model may be used to install guardrails in a Tier Alpha NKP-AI system. The PPO parameters, reward model weights, and training data for all NKP-AI systems must be classified as Tier 1 national security data under the data residency framework in Section 3.
Sovereign Key Management	Cryptographic keys governing NKP-AI system access must be generated, stored, and managed exclusively within South African sovereign infrastructure using SABS-certified hardware security modules (HSMs).
Multi-Party Authorisation	Any NKP-AI function capable of causing national-scale impact must require cryptographically verified authorisation from a minimum of three independent, geographically separated human officials before execution.
Mandatory Redundancy Architecture	Every NKP-AI system must have a fully operational, independently powered, geographically separated backup system capable of assuming full operational load within 90 seconds of primary system failure.

8.5 Enforcement and Sanctions

Sanctions below are simply indicative, and a starting point for discussion. All penalties must be evaluated by an expert panel.

Violation	Sanction
Failure to maintain air-gap or domestic processing requirements	Criminal offence; R100 million administrative penalty; mandatory system shutdown pending remediation; personal criminal liability for the Chief Executive and Chief Information Officer.
Supply chain security breach resulting in compromised NKP-AI system	Criminal offence classified as sabotage of a National Key Point; prosecution under both the AIGA and the National Key Points Act; minimum 15-year imprisonment for deliberate introduction of compromised components.
Failure to report a NKP-AI security incident within 15 minutes	R10 million per hour of delay beyond the notification window; personal liability for the designated NKP-AI Security Officer.
Failure to maintain 21-day manual continuity capability	Licence revocation; mandatory third-party operational takeover at operator's cost until compliance is demonstrated.

9. International Regulatory Benchmarking

South Africa's AI regulatory framework must be informed by, but not derivative of, global best practice. The following comparative analysis identifies the most relevant international models.

International AI Regulatory Approaches: Comparative Summary

Jurisdiction	Approach	Key Strengths	Considerations for South Africa
European Union - EU AI Act (2024)	Comprehensive, risk-based mandatory regulation	Strong rights protection; extraterritorial effect; sets the global regulatory standard	Compliance burden may chill innovation in developing-economy contexts
United States - NIST AI RMF + Executive Order (2023)	Voluntary standards plus sector regulation	Flexible, innovation-friendly	Fragmented; limited enforcement; relies on industry self-regulation
United Kingdom - Pro-Innovation Approach	Principles-based, sector-led regulation	Agile; avoids over-regulation; empowers existing sector regulators	Uncertainty for businesses; slow to address systemic risks
China - State-Directed AI Governance	Security-first, state-aligned AI development	Strong sovereignty protections; fast deployment	Limited civil liberties protections; not transferable to South Africa's constitutional framework
Singapore - Model AI Governance Framework	Voluntary framework plus regulatory sandbox	Business-friendly; test-and-learn culture	Voluntary adoption is limited without enforcement
South Africa - Proposed Risk-Based Hybrid	Mandatory for High-Risk AI; innovation sandbox; sovereignty provisions	Constitutionally grounded; development-sensitive; sovereign	Requires significant institutional capacity investment

10. Proposed South African AI Regulatory Framework

The Draft Policy proposes an institutional architecture centred on a National AI Commission or Office, an AI Ethics Board, an AI Regulatory Authority, an AI Ombudsperson Office, an AI Safety Institute, and broader inter-regulatory coordination, coordinated through a National AI Regulatory Forum convened by the DCDT and bringing together ICASA, the Information Regulator, the Competition Commission, SARB, FSCA, CSIR, and DTIC.

Critically, the Draft Policy explicitly declined to create a single AI super-regulator, opting instead for a whole-of-government, sector-coordinated model. While we would prefer a single 'super regulator' for which we have prepared a separate submission to support, if required, Ubiquity AI's position is to support and help operationalise that distributed architecture.

References to a future Artificial Intelligence Governance Act should be understood as a medium-term legislative pathway consistent with the Draft Policy's staged model. The institutional proposals below are best read as implementation detail for the multi-regulator framework the Draft Policy contemplates.

10.1 Artificial Intelligence Governance Act (AIGA)

Consistent with the Draft Policy's staged implementation approach, the AIGA is positioned as a medium-term legislative consolidation option. When introduced, it should:

AIGA Element	Description
Constitutional grounding	Be grounded in the Bill of Rights (Chapter 2 of the Constitution), including rights to dignity, equality, privacy, freedom of expression, administrative justice, and access to information.
AI Regulatory Authority establishment	Establish the South African AI Regulatory Authority (SAIRA) as an independent, adequately resourced regulatory body, consistent with the institutional configuration contemplated by the Draft Policy.
Definitions	Define AI, AI systems, AI providers, AI deployers, and other key terms in a technology-neutral manner. Definitions should expressly encompass key fine-tuning methodologies including Reinforcement Learning from Human Feedback (RLHF) and the Proximal Policy Optimisation (PPO) algorithms used to implement RLHF - recognising that PPO is the primary mechanism through which AI guardrails are installed and that transparency into PPO parameters is a necessary condition for meaningful AI accountability.
Risk classification	Establish the four-tier risk classification system and corresponding obligations.
Penalty regime	Create proportionate civil and criminal penalty regimes.
Transition periods	Provide for transition periods calibrated to system risk level.
Regular review	Require statutory review every three years to keep pace with technological change.

10.2 South African AI Regulatory Authority (SAIRA) - Proposed Structure

Specific figures - penalty levels, percentage thresholds, FTE requirements - are presented as proposals for consultation rather than fixed positions. All figures are starting points for expert engagement rather than demands. All penalties must be justified by the Parliamentary Committee based on detailed modelling of harm scenarios.

Element	Detail
Board of Directors	11 members appointed by the Minister, comprising legal, technical, ethics, and civil society expertise, with gender parity and racial representivity requirements.
Operational Divisions	AI Safety and Compliance; Data and Privacy (liaising with the Information Regulator); National Security AI (liaising with SSA and SANDF); Innovation and Standards.
Regional presence	Offices in all nine provinces within 5 years of establishment.
Advisory Panel	A multi-disciplinary AI Technical Advisory Panel including academic, industry, and civil society representatives.
Licensing powers	Licensing and registration of High-Risk and Critical-Risk AI systems.
Investigation powers	Investigations, inspections, and audits of AI systems — including access to source code and training data under judicial warrant.
Penalty powers	Administrative penalties up to R100 million per violation for legal entities; R5 million per violation for natural persons.
Enforcement powers	Prohibition orders and mandatory suspension of non-compliant AI systems; criminal referrals to the NPA.
Annual reporting	Publication of an Annual AI State of the Nation Report.

10.3 National AI Innovation Sandbox

To ensure that the regulatory framework does not stifle innovation - particularly by SMMEs, startups, and researchers - the AIGA shall establish a National AI Innovation Sandbox operated by SAIRA in partnership with the DTIC.

Qualifying organisations may develop and test novel AI applications in a controlled regulatory environment with relaxed compliance obligations for defined periods of up to 24 months. An Africa AI Sandbox Corridor shall be developed with partner countries to enable cross-border AI innovation within a governance framework.

10.4 National AI Skills and Transformation Agenda (NAISTA)

Initiative	Description
AI Literacy Curriculum	Integration of foundational AI literacy into the CAPS curriculum from Grade 7, with dedicated AI modules in Information Technology and Computer Applications Technology.

Initiative	Description
TVET AI Pathway	Development of accredited AI practitioner qualifications at NQF Levels 4–6 at all 50 TVET campuses, focused on practical AI application rather than academic AI theory.
AI Innovation Bursary Fund	500 new bursaries per year at postgraduate level for AI research, with preference for historically disadvantaged students.
Corporate AI Transformation Obligation	Companies with annual revenue exceeding R1 billion that deploy AI systems must submit an AI Transformation Plan demonstrating how AI deployment will create rather than simply destroy employment.
AI Township Economy Programme	Co-funded AI business development support for township entrepreneurs in a dedicated AI SMME Incubation Programme.

11. Implementation Roadmap

The Draft Policy's Explanatory Note sets out a clear three-phase staged implementation model:

Year-1
(2025/26) now substantially complete) covering policy finalisation, identification of unacceptable-risk regulatory requirements, and initiation of National AI Policy Guidelines;

Year-2
(2026/27) covering publication of guidelines, implementation of high-risk use case requirements, development of sectoral AI strategies, and commencement of institutional framework design; and

Year-3
(2027/28) covering full implementation of outstanding policy interventions. Ubiquity AI strongly supports that sequencing.

The roadmap below adds implementation detail consistent with the Draft Policy's own timetable, while ensuring that work streams on unacceptable-risk use cases, financial AI, public-sector AI, and critical infrastructure are not delayed.

Proposed 48-Month Implementation Phasing

Phase	Title	Key Milestones	Lead Actors
Phase 1 (Months 1–12)	Foundation	Draft and introduce AIGA to Parliament; establish AI Regulatory Authority Interim Board; publish POPIA AI Code of Practice; launch public consultation; develop National Security AI Assessment framework.	DTIC, Parliament, Information Regulator
Phase 2 (Months 13–24)	Establishment	Enact AIGA; operationalise AI Regulatory Authority; launch AI Procurement Register; open AI Innovation Sandbox; publish AI Risk Classification Guidelines.	AI Regulatory Authority, OCPO, DHET
Phase 3 (Months 25–36)	Enforcement	Begin High-Risk AI registration programme; commence AI audits; NAISTA curriculum rollout; negotiate SADC AI Compact; publish first Annual AI State of Nation Report.	AI Regulatory Authority, SADC, DHET, TVET
Phase 4 (Months 37–48)	Maturity and Review	First AIGA statutory review; Critical Infrastructure AI Standard fully implemented; AI Sandbox second cohort; South Africa leads African Union AI Charter negotiations.	All stakeholders

11.1 Resourcing the AI Regulatory Authority

Specific figures are presented as proposals for consultation rather than fixed positions. All figures are starting points for expert engagement rather than demands.

Resource Item	Detail
Staff complement (Year 1)	120 FTEs including lawyers, AI engineers, data scientists, economists, policy analysts, and administrative staff.
Establishment costs	Office fit-out, IT infrastructure, and regulatory technology systems: estimated R175 million (year-one). • Governance & Policy: R15 - R25 million • AI Monitoring systems: R35 - R55 million • Regulatory sandbox: R20 - R40 million • Personal/experts (AI engineers, data scientists etc): R25 - R40 million (set-up team) • Operational admin: R10 - R20 million
Annual operating budget (Year 1)	Estimated R270 million, to be funded from the fiscus with a phased levy on AI operators from Year 3. (About twice the National Credit Regulators 's annual budget) • Personnel (120 FTEs): R169.3 million - Includes 30% overhead for benefits. Reflects high market rates for AI Developers (Avg. R763k) and Senior Leads (Up to R1.5M) • ICT & High-Performance Computing: R45 million - Cloud credits, secure auditing tools, and specialised server clusters for model testing • Regulatory Sandbox Ops: R25 million - Dedicated technical vetting, data sanitisation, and security for testing third-party models. • Professional & Legal Fees: R15 million - Specialised AI audits, external legal counsel for high-risk disputes, and ethics advisory. • Admin & Infrastructure: R15 million Office space, utilities, and general corporate services.
AI Operator Levy	Annual levy of 0.1% of South African AI-derived revenue for entities with revenue exceeding R100 million per annum.
International Technical Assistance	The AI Regulatory Authority may accept technical cooperation assistance from the EU AI Office, NIST, and Singapore IMDA to build capacity.
Skills pipeline	10 AI Specialist Bursaries per year at postgraduate level, bonded for 5 years of service.

12. Additional Critical Policy Issues

12.1 AI and Employment - Just Transition

EMPLOYMENT ALERT: Scale of AI-Driven Job Displacement

A 2025 World Economic Forum report estimates that South Africa could lose up to 1.4 million jobs to AI automation by 2030, concentrated in administrative, transport, and basic service sectors. Without a Just Transition policy framework, this represents a structural social crisis.

Measure	Description
AI Employment Impact Assessments	Mandatory for any organisation deploying AI that will result in the retrenchment of 50 or more employees, filed with the Department of Employment and Labour and the AI Regulatory Authority simultaneously.
Extended Section 189 consultation	The existing retrenchment consultation process shall be extended to cover AI-driven job losses, with a minimum 90-day consultation period and obligation to explore redeployment and reskilling alternatives.
National AI Just Transition Fund	Capitalised by a 2% levy on AI operator revenues exceeding R500 million per annum, to fund retraining and income support for workers displaced by AI.
Labour representation on AI Regulatory Authority Board	Organised labour (COSATU, FEDUSA, NACTU) shall have permanent representation on the Board and in the AI Transformation Plan review process.

12.2 AI and Healthcare

Measure	Description
High-Risk AI classification for clinical AI	All AI systems used in clinical diagnosis, treatment recommendation, drug dispensing, or surgical assistance shall be classified as High-Risk AI, requiring pre-market authorisation in addition to SAHPRA medical device registration.
National Health AI Ethics Protocol	To be developed by the NDOH, addressing consent, data ownership, and community benefit requirements for health AI data use.
Demographic performance parity	AI healthcare systems deployed in public hospitals must demonstrate equivalent or superior performance across all racial and demographic groups in the South African population.
Community Health Worker AI Tools	AI-assisted tools for community health workers in rural and peri-urban settings shall receive prioritised regulatory fast-track, given the national shortage of healthcare professionals.

12.3 AI and Cultural Preservation and Indigenous Languages

The Draft Policy includes a specific strategic objective (Objective 'e') that is absent from most international AI frameworks but is of profound importance for South Africa: the use of AI to digitise and preserve indigenous languages, arts, music, and literature, and to enable real-time translation across all 12 official languages. Ubiquity AI strongly endorses this objective and recommends the following implementation measures:

Firstly, a National AI Language and Culture Fund should be established, resourced through the DSAC and DTIC, to co-fund AI-driven digitisation projects for Nguni, Sotho, Venda, Tsonga, and other language groups currently underrepresented in global AI training corpora.

Secondly, SAIRA (the AI Regulatory Authority) should develop minimum standards for South African language representation in AI systems deployed in public education, public health, and government service delivery - ensuring that AI tools function with equivalent quality across all 12 official languages within a decade, not only English and Afrikaans.

Thirdly, AI systems used for cultural content generation - including music, visual art, oral history digitisation, and performing arts - should be subject to the intellectual property protections recommended in Section 6 of this submission, with additional protections for community-owned intangible cultural heritage.

This is particularly important given the risk that generative AI systems trained predominantly on Western content corpora could appropriate, commodify, or distort indigenous South African cultural expressions.

12.4 AI and the Environment

Measure	Description
AI Environmental Impact Standard	SAIRA, in consultation with DFFE, shall develop an AI Environmental Impact Standard requiring large AI operators to report energy consumption, water usage, and carbon emissions of their South African operations annually.
Renewable energy incentives	Tax incentives for AI data centres powered by renewable energy, introduced under the Revenue Laws Amendment Act.
International advocacy	South Africa shall advocate for AI environmental disclosure standards at the AU and G20 level.

12.4 AI and the Rule of Law

Issue	Recommended Approach
AI-generated legal documents and evidence	Courts shall develop rules of evidence for AI-generated materials, requiring provenance disclosure, authentication, and human legal responsibility for AI-generated filings.
AI in the Judiciary	No court may make a binding judicial decision using AI as the primary decision instrument. AI may be used as a research, drafting, or case management tool only.

Issue	Recommended Approach
Legal profession guidelines	The Department of Justice and the LSSA shall develop guidelines for responsible AI use in legal practice, including ethical obligations around AI-generated advice.
AI access to justice tools	AI-powered legal access tools for self-represented litigants - particularly in the small claims court - shall be actively promoted, subject to quality and bias standards.

12.5 AI Governance and Democratic Accountability

Measure	Description
Parliamentary oversight	SAIRA shall be subject to full Parliamentary oversight, reporting annually to the Portfolio Committee on Trade, Industry and Competition and to the Standing Committee on the Auditor-General.
Civil society participation	Civil society organisations shall have formal standing to file complaints with SAIRA and to participate in regulatory consultations.
National AI Public Dialogue	An annual national AI Public Dialogue programme shall be conducted in all nine provinces, in all official languages, to ensure citizens' voices shape AI governance.
Right of petition	The right to petition SAIRA for investigation of any AI system shall be enshrined in the AIGA.
Whistleblower protections	Protections under the Protected Disclosures Act shall be explicitly extended to cover disclosures about AI system harms or non-compliance.

13. Conclusion and Summary of Recommendations

South Africa's opportunity is not merely to adopt AI. It is to shape an African model of AI governance that takes innovation seriously while refusing to treat social harm, exclusion, opacity, or sovereignty loss as acceptable collateral costs. The recommendations below are submitted in that spirit - not to displace the Government's framework, but to demonstrate how that framework can be strengthened through practically grounded, technically specific proposals.

13.1 The Urgency Imperative

The global AI regulatory landscape is crystallising rapidly. The EU AI Act is already in effect. The United States, United Kingdom, Singapore, China, India, Brazil, and dozens of other jurisdictions have enacted or are enacting AI-specific regulatory measures.

Every month that South Africa delays comprehensive AI governance, more AI systems are deployed on South African citizens without adequate safeguards, more data leaves South African borders without sovereign protection, and more legislative ground cedes to foreign regulatory frameworks as the de facto standard.

This submission is not about creating bureaucratic obstacles to innovation. It is about building the foundations of a trustworthy, sovereign, and human-centred AI economy - one that harnesses AI's extraordinary potential for development, education, healthcare, and economic inclusion, while ensuring that no South African citizen is left behind, harmed, or exploited by algorithmic systems beyond democratic accountability. We need to think about the opportunity of AI akin to the early days of the industrial revolution and the internet. This is a transformative technology which will touch every aspect of life and society.

13.2 Priority Recommendations

#	Recommendation
1	Enact the Artificial Intelligence Governance Act (AIGA) as primary legislation in the 2026 Parliamentary cycle.
2	Establish the South African AI Regulatory Authority (SAIRA) as a fully resourced, independent Schedule 3A public entity.
3	Adopt a four-tier, risk-proportionate AI classification system aligned with constitutional values.
4	Introduce absolute prohibitions on social scoring AI, AI mass surveillance, and AI electoral manipulation.
5	Enact tiered data residency requirements for AI systems, including mandatory localisation for Critical Infrastructure AI.
6	Amend POPIA to introduce AI-specific rights: algorithmic explanation, AI-DPIA, and training data consent.
7	Introduce AI-specific consumer protection provisions supplementing the Consumer Protection Act.
8	Amend the Copyright Act to address AI training data use, AI-generated works, and personality rights protection.

#	Recommendation
9	Mandate Algorithmic Impact Assessments for all High-Risk AI systems, with public disclosure of findings.
10	Adopt a Government AI Charter prohibiting fully automated administrative decisions without human review.
11	Establish a National AI Innovation Sandbox for responsible experimentation by SMMEs and researchers.
12	Launch the National AI Skills and Transformation Agenda (NAISTA) with dedicated TVET qualifications and bursaries.
13	Introduce an AI Just Transition Fund and mandate AI Employment Impact Assessments for large-scale AI-driven job displacement.
14	Lead African Union AI Charter negotiations as a key element of South Africa's continental leadership agenda.
15	Negotiate AI Sovereignty Clauses in bilateral trade and investment agreements through DIRCO.

"Umuntu ngumuntu ngabantu"

A person is a person through other persons.

The governance of AI must embody this principle: technology that serves humanity, accountable to communities, shaped by democratic values. South Africa has the constitutional foundation, the legal tradition, the technical talent, and the moral authority to build an AI governance regime that is a model not only for Africa, but for the world.

Kaveer Beharee | Founder, Ubiquity AI | April 2026

Appendix: Glossary of Key Terms

Term	Definition
Artificial Intelligence (AI)	A machine-based system that, for a given set of human-defined objectives, makes predictions, recommendations, decisions, or generates content influencing real or virtual environments.
AI System	Any AI-powered software, algorithm, model, platform, or service deployed for a specific purpose.
AI Provider	Any natural or legal person who develops, places on the market, or puts into service an AI system.
AI Deployer	Any natural or legal person using an AI system in the course of professional activities.
Algorithmic Decision	Any determination materially affecting a person's rights, interests, or access to services, made wholly or partly on the basis of automated AI processing.
Biometric Data	Personal data resulting from technical processing relating to physical, physiological, or behavioural characteristics, including facial recognition, voice recognition, gait analysis, and emotional inference.
Foundation Model	An AI model trained on broad data at scale and adaptable to a wide range of tasks, including large language models and multimodal models.
Generative AI	AI systems capable of generating novel text, images, audio, video, code, or other content in response to inputs.
High-Risk AI	AI systems deployed in contexts with significant potential for harm to rights, safety, livelihood, or access to essential opportunities or services.
AI-DPIA	Artificial Intelligence Data Protection Impact Assessment - a pre-deployment assessment for higher-risk AI systems involving personal data.
Deepfake	AI-generated synthetic media - video, audio, or image - that realistically depicts a person saying or doing something they did not say or do.
LLM (Large Language Model)	A type of generative AI model trained on large text corpora and capable of generating human-quality text.
Robo-advisory	An automated AI-powered platform that provides financial planning or investment guidance without a human financial adviser.
Corpus	The body of text and data used to train an AI model - effectively the model's 'reading list'.

Term	Definition
Reinforcement Learning from Human Feedback (RLHF)	A training technique where human annotators rate outputs as better or worse, and the model is repeatedly adjusted to produce outputs those humans prefer.
Double-Layer Bias	A compounding bias problem in which the training corpus skews heavily toward one context - here, predominantly Western financial contexts - and then over-represents high-volume non-professional content relative to qualified professional sources, thereby amplifying distortion.
SAIRA	South African AI Regulatory Authority - the proposed independent regulatory body for AI governance in South Africa.
AIGA	Artificial Intelligence Governance Act - the proposed primary AI legislation for South Africa.
SAFC Requirement	Mandatory South African Financial Corpus Requirement - the requirement that AI financial services systems incorporate a minimum 35% weighting of South African authoritative financial content in their training corpus.
SACFST	South African Cultural Financial Scenario Test - a pre-authorisation test developed by SAIRA to assess the cultural and regulatory alignment of AI financial systems for South African deployment.
NKP-AI	National Key Points AI - AI systems deployed in or affecting designated critical national infrastructure, subject to the highest level of regulatory oversight.
Proximal Policy Optimisation (PPO)	A reinforcement learning algorithm used as the primary mechanism for fine-tuning AI models through Reinforcement Learning from Human Feedback (RLHF). PPO converts human annotator preference ratings - via a trained reward model - into actual changes to an AI model's parameters, installing behavioural guardrails that determine what the model will and will not say. PPO parameters (including clipping ratios, entropy bonuses, and KL-divergence constraints) directly govern AI model behaviour but are rarely disclosed. Transparency into PPO parameters is a necessary condition for meaningful AI accountability under this submission's proposed regulatory framework.