

R&D WHITE PAPER • PUBLIC EDITION

# SanKam: Africa AI Trade Interpretation Platform

An end-to-end strategy for data collection, model specialisation and production-grade deployment for African trade languages

ORIGINAL PAPER

March 2026

PREPARED BY

Bridge

PUBLIC EDITION

October 2026

PREPARED FOR

SanKam Limited



---

# Contents

|                                                                              |    |
|------------------------------------------------------------------------------|----|
| 1. Executive Summary .....                                                   | 2  |
| 2. The Problem SanKam Is Solving .....                                       | 4  |
| 3. Strategic Thesis: Product Utility as the Engine of Proprietary Data ..... | 6  |
| 4. What Phase 0 Has Demonstrated .....                                       | 7  |
| 5. Product Architecture: Utility First, Data Native .....                    | 10 |
| 6. End-to-End Data Pipeline .....                                            | 12 |
| 7. Model Strategy: From Managed APIs to Self-Hosted Specialisation .....     | 14 |
| 8. Evaluation Framework .....                                                | 17 |
| 9. Technical Infrastructure and Deployment .....                             | 18 |
| 10. Roadmap and Phasing .....                                                | 20 |
| 11. Risks and Mitigation Strategy .....                                      | 21 |
| 12. Conclusion .....                                                         | 22 |
| Appendix A: Target Language Coverage .....                                   | 23 |
| Appendix B: Glossary .....                                                   | 24 |
| References .....                                                             | 26 |

## About this edition

This public edition of SanKam's R&D white paper is shared for partners, researchers and others interested in the thinking behind SanKam. It sets out the problem, what our proof of concept has shown, and how we plan to build speech-to-speech interpretation for African trade languages.

Commercially sensitive material, including financial planning, has been removed. The paper reflects SanKam's R&D position in March 2026. Since then our MVP delivery plan has grown to twelve languages, with each language pair tested separately and released as it meets quality thresholds. Figures, plans and timelines are indicative and may change.

# 1. Executive Summary

SanKam addresses a clear and underserved problem: language barriers continue to limit trade efficiency, increase transaction risk, and exclude traders from market access across African borders and regions. Existing digital tools are not designed for the linguistic reality of African trade, where interactions are verbal, informal, code-switched, dialect-sensitive, and rich in domain-specific terminology. Professional human interpretation remains too expensive for most traders, while general-purpose AI systems are insufficiently adapted to the complexity of African languages in real trade contexts.

SanKam's proposition is to build a purpose-designed speech-to-speech interpretation platform for African trade languages, beginning with a utility-first product that solves a real problem from day one and progressively evolving that product into a defensible, proprietary language AI asset. The core logic: a useful product drives adoption; adoption generates in-domain speech and conversation data; that data enables model specialisation; and the resulting quality improvements compound SanKam's competitive position over time. This data flywheel is central to both the business case and the technical roadmap.

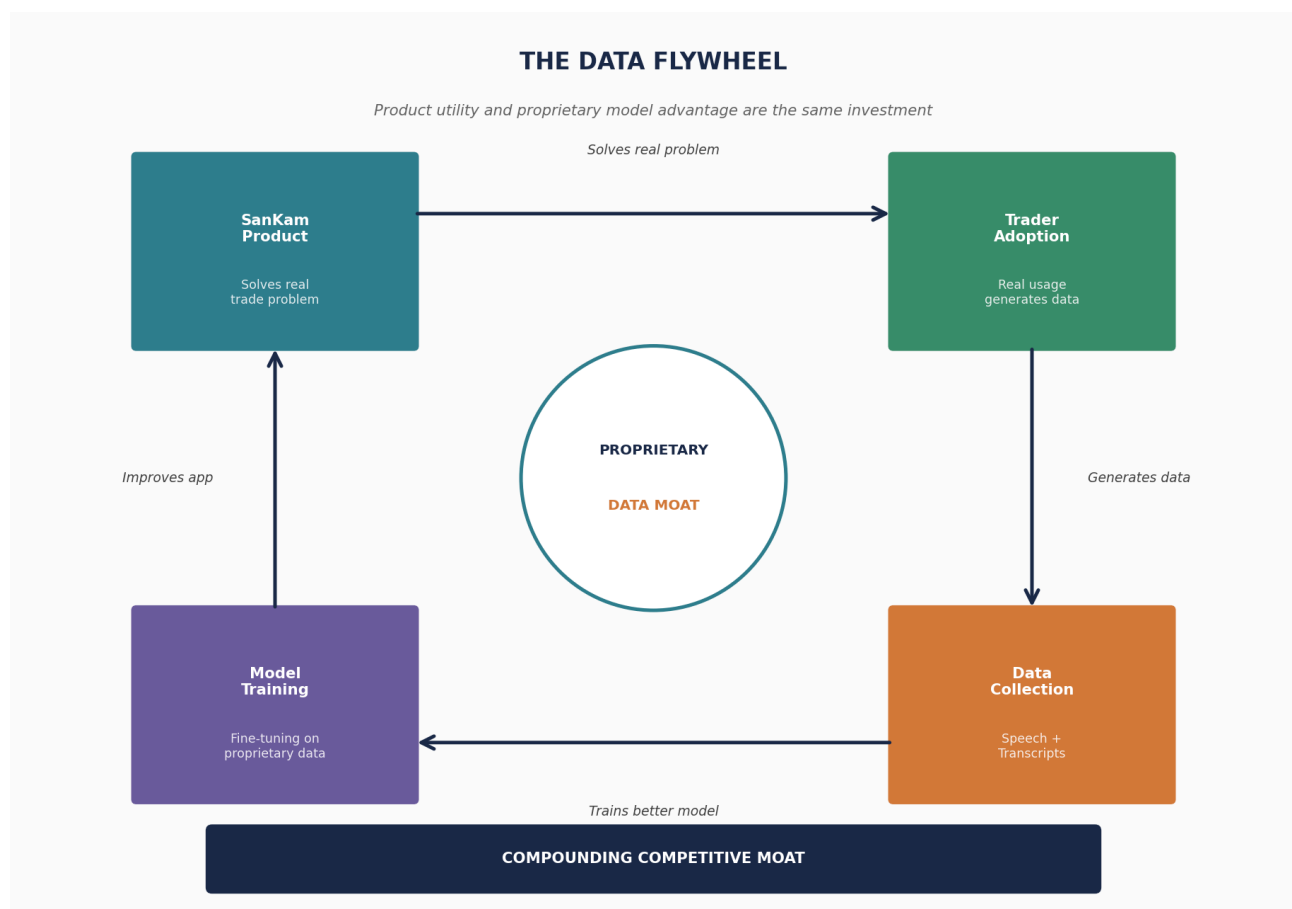


Figure 1: The data flywheel. Product utility and proprietary model advantage are the same investment.

Phase 0 has demonstrated early technical feasibility. A working proof-of-concept pipeline was built and tested across 8 real trade scenarios. The proof of concept confirmed that African-specialised translation models outperform general-purpose alternatives on culturally nuanced language pairs, and that an end-to-end speech pipeline is technically viable in a real trade context. It also surfaced the limitations that now define the production roadmap: offline capability is absent, current APIs do not allow fine-tuning, code-switching remains a major challenge, and tonal, dialectal and trade-specific vocabulary gaps require proprietary data collection and model adaptation.

This white paper sets out the R&D and product development path from proof of concept to production-grade platform, covering: the problem structure; what Phase 0 has proven and what remains unproven; the data strategy required; the model architecture and training pathway; the technical infrastructure needed for real African trade environments; and the phasing and risk management required to reach production scale.

The roadmap is a three-phase build over 12 to 18 months, transitioning SanKam from an API-based proof of concept to a self-hosted, fine-tunable, offline-capable production platform with a proprietary dataset and model stack tailored to African trade communication.

---

## 2. The Problem SanKam Is Solving

### 2.1 Africa's language reality remains structurally underserved by modern AI

Africa contains approximately 2,144 living languages and roughly 30% of global linguistic diversity. Yet the overwhelming majority of these languages remain underrepresented in modern NLP, speech and multimodal AI systems. Even where coverage exists, it is usually shallow and biased toward formal written text rather than spoken, colloquial or regionally specific usage. Languages such as Yoruba, Hausa, Swahili, Amharic and Wolof may appear in multilingual model coverage lists, but support is often weak in dialectal variation, oral fluency, code-switching and trade-specific usage. Independent benchmarks document significant quality gaps, with leading models achieving substantially lower performance on African languages than European ones across translation, comprehension and named entity recognition tasks (Adelani et al., 2022; Ojo et al., 2025).

This matters because trade communication is not conducted in neat, formal, standardised text. It is verbal, fast, negotiated, context-heavy and embedded in mixed-language speech. A trader in Kano may use Hausa shaped by local vocabulary, intonation and code-switching patterns that differ materially from another Hausa speaker elsewhere. These variations are precisely the features least well captured by current general-purpose systems.

### 2.2 Trade communication is a compelling and structurally tractable entry point

Intra-African trade reached \$220.3 billion in 2024, growing 12.4% year on year (African Development Bank, African Economic Outlook 2024). 85.8% of employment in Africa is informal, making informal and cross-border trade the primary economic reality for the majority of the continent's working population (ILO, 2018). The African Continental Free Trade Area aims to deepen regional integration, yet the digital tools supporting everyday commercial communication remain anchored disproportionately to global languages. This creates a gap between strategic integration ambitions and the actual language reality of everyday market actors.

The commercial pain point is meaningful. Language barriers increase transaction friction, raise fraud risk, slow negotiation and limit expansion. Language barriers can add significant costs to cross-border trade: research at high-volume corridors, including Chirundu (Zimbabwe and Zambia), estimates their trade cost equivalent at around 5 to 12% of transaction value for informal traders (Brenton and Isik, *De-Fragmenting Africa*, World Bank, 2012). Professional human interpreters cost \$70 to \$200 per hour, which is economically unviable for traders managing consignments under \$2,000.

Trade communication is also technically valuable because it exhibits patterns that make it suitable for focused model development: domain vocabulary (quantities, pricing, customs, logistics), repeated intent structures (offer, counter-offer, acceptance, rejection), high rates of code-switching, and concentrated use cases that can be mapped, annotated and learned progressively.

## **2.3 Why this is difficult, and why that creates opportunity**

Three structural constraints make this problem hard. First, the data does not exist at useful scale: African trade speech datasets are limited or non-existent for many target languages and dialects. Second, relevant trade conversations include characteristics generic systems handle poorly: mixed languages, tonal sensitivity, regional variation, informal speech and context-dependent meaning. Third, the deployment environment is unforgiving: inconsistent bandwidth, mid-range Android devices, and users who require practical reliability, not novelty.

These constraints are precisely why there is opportunity. A platform that solves this well does not merely build a feature. It accumulates a proprietary dataset, a domain-adapted model stack, operational knowledge of African trade corridors, and a trust position with real users. That combination is difficult to replicate and creates the basis for durable defensibility.

---

### 3. Strategic Thesis: Product Utility as the Engine of Proprietary Data

SanKam's core insight is that data collection should not be treated as a separate research exercise detached from user value. Instead, the product itself should become the mechanism through which valuable data is generated. This is more scalable and more defensible than relying on static public datasets or artificial collection exercises.

The reinforcing loop is as follows: traders use SanKam because it solves an immediate communication need; usage produces real speech, transcripts, corrections and trade-specific context; that data is structured and annotated into model-ready training assets; models improve on African trade communication; improved performance increases usefulness and adoption; adoption generates more and better data.

The stronger version of this narrative is not just that "data improves the model", but that SanKam is designed to create exactly the kind of proprietary, in-domain dataset that current global models do not possess. That is the source of both technical advantage and enterprise value. This approach also aligns product strategy with R&D efficiency: enter through a narrow set of high-value use cases, generate real usage, and progressively expand capability with evidence rather than assumption.

## 4. What Phase 0 Has Demonstrated

### 4.1 Objective and scope

The purpose of Phase 0 (proof of concept) was not to prove production readiness. It was to test whether the core speech-to-speech interpretation workflow could be assembled using available components, and whether African-specialised tools would show a meaningful advantage in the target use case. That objective has been met.

Phase 0 delivered a working end-to-end pipeline: speech-to-text via OpenAI Whisper, translation via Lelapa AI Vulavula (primary) with GPT-4o fallback, and text-to-speech via OpenAI gpt-4o-mini-tts. The system supports 10 languages across all major African trade blocs and includes transparent routing metadata for every response.

### 4.2 Pipeline architecture

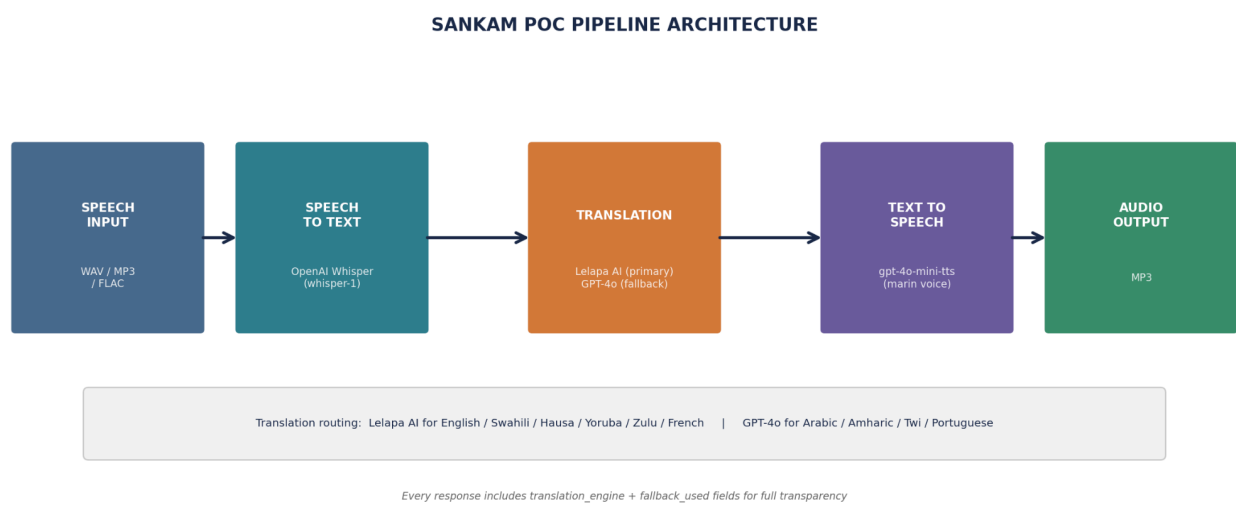


Figure 2: Proof-of-concept pipeline architecture. A modular, model-agnostic design means any component can be upgraded without architectural change.

### 4.3 Test results

Testing was conducted across 8 real trade scenarios, covering grain price negotiation, bulk discount offers, market logistics, cross-border payment terms, export documentation, and a full end-to-end speech-to-speech pipeline test on real FLAC audio. Pass criteria were defined as: accurate conveyance of the full commercial intent of the source utterance; correct script rendering; and latency within an acceptable range for real-time conversation. A partial pass indicates the core intent was conveyed but with content omission or imprecision.

| Scenario |                         | Language pair      | Engine    | Result |
|----------|-------------------------|--------------------|-----------|--------|
| 1        | Grain price negotiation | English to Swahili | Lelapa AI | Pass   |

|   | Scenario                                      | Language pair         | Engine    | Result       |
|---|-----------------------------------------------|-----------------------|-----------|--------------|
| 2 | Bulk discount offer                           | Hausa to English      | Lelapa AI | Partial pass |
| 3 | Market announcement                           | French to Yoruba      | Lelapa AI | Pass         |
| 4 | Price rejection and negotiation               | Zulu to English       | Lelapa AI | Pass         |
| 5 | Cross-regional grain trade (no English pivot) | Hausa to Swahili      | Lelapa AI | Partial pass |
| 6 | Trade account and payment terms               | Swahili to Arabic     | GPT-4o    | Pass         |
| 7 | Export delivery and certificate of origin     | Portuguese to Amharic | GPT-4o    | Pass         |
| 8 | Full end-to-end pipeline (real FLAC audio)    | English to Swahili    | Lelapa AI | Pass         |

The strongest result was Test 3 (French to Yoruba), where Lelapa produced correct diacritics and the respectful plural address form appropriate for a public announcement. This level of cultural nuance is a direct product of African-language specialisation. Test 8 confirmed the full end-to-end speech-to-speech pipeline on real FLAC audio: Whisper transcription was accurate and clean MP3 output was returned successfully. The two partial passes reflect known Lelapa limitations on multi-sentence Hausa inputs and direct cross-African language pairs without a bridge language.

#### 4.4 What Phase 0 has not yet demonstrated

The following critical capabilities remain unproven and define the Phase 1 agenda:

- Robust offline or low-connectivity operation
- Production-grade performance on code-switched speech
- Fine-tuning on proprietary trade datasets
- High-quality tonal and dialect-sensitive text-to-speech for key African languages
- Sustained cost-efficient inference at scale
- Voice-controlled navigation for low-literacy users

That distinction strengthens rather than weakens the paper. It signals discipline and gives the reader confidence that the roadmap is built on evidence rather than aspiration.

#### 4.5 Proof-of-concept limitations informing the production roadmap

| Limitation                                                                         | Impact                                                          | Production solution                                                                    |
|------------------------------------------------------------------------------------|-----------------------------------------------------------------|----------------------------------------------------------------------------------------|
| Text-to-speech approximates African phonemes rather than reproducing them natively | Reduced comprehension on tonal languages like Yoruba            | Meta MMS: dedicated text-to-speech models for 1,100+ languages with native prosody     |
| No offline capability: full API dependency throughout                              | Unsuitable for border environments with unreliable connectivity | On-device model quantisation (TFLite/ONNX); offline-first architecture (see Section 9) |

| Limitation                                                                    | Impact                                                             | Production solution                                                                   |
|-------------------------------------------------------------------------------|--------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| No fine-tuning path on managed API services                                   | Cannot adapt to proprietary trade vocabulary or domain speech      | Self-hosted open models: Whisper, NLLB-200, MMS, SeamlessM4T                          |
| Code-switching not handled: each utterance treated as monolingual             | Fails on mixed-language input common in real trade conversations   | Multi-task fine-tuning with language ID tokens; annotated training data               |
| Direct cross-African pairs without a bridge language (e.g. Hausa and Swahili) | Imprecise output at current Lelapa scale (0.4B parameters)         | Explicit GPT-4o routing for non-bridge pairs; routing-layer configuration change only |
| No voice-controlled UX: assumes touch literacy                                | Excludes traders unfamiliar with smartphones or touchscreens       | Voice-first interaction model throughout (see Section 5.2)                            |
| Pay-per-use API: no self-hosting                                              | Not viable at production scale on cost or data sovereignty grounds | Self-hosted models; proprietary data assets reduce API dependency over time           |

## 5. Product Architecture: Utility First, Data Native

SanKam should be designed not simply as a translation tool, but as a trade utility platform whose features each generate meaningful training signals. The product features are valuable in their own right, and the key strategic insight is that every user interaction can contribute to the dataset.

| Feature                    | User value                                                                       | Training signal generated                                                        |
|----------------------------|----------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| Voice interpretation       | Traders speak, receive interpreted output and continue a live conversation       | Paired audio and transcript; language and dialect labels; latency metrics        |
| Text chat with translation | Direct messaging with real-time translation across language boundaries           | Parallel corpora; source and target language labels; user corrections            |
| Trade listings             | Post goods in a native language; buyers across language boundaries can read them | Domain vocabulary; product taxonomy; pricing language; regional naming           |
| Correction feedback        | Users flag poor outputs and suggest corrections                                  | Error pairs; preference data for alignment; quality signal per language pair     |
| Profile and onboarding     | Voice-guided setup in a native language; no text literacy required               | Demographic metadata; dialect mapping; trade sector vocabulary priors            |
| Conversation transcripts   | A record of the conversation that both traders can refer back to                 | Full conversational datasets preserving negotiation context and intent structure |

A strong product design positions SanKam as “data native”: the operating model, user experience and consent flow are aligned from the outset to generate long-term model advantage.

### 5.1 Consent and data governance

Because SanKam’s strategy depends on product-generated data, trust and governance cannot be retrofitted later. The consent framework includes multilingual informed consent, anonymisation at source, withdrawal rights, and compliance with the NDPA (Nigeria), DPA (Kenya) and POPIA (South Africa). For a platform serving traders in linguistically diverse settings, transparent governance is part of adoption strategy.

### 5.2 Voice-first accessibility and automatic language detection

A core design principle that distinguishes SanKam from general-purpose translation tools is that the platform must be fully operable by users who have never used a smartphone application before. A significant proportion of the target user base includes traders with limited formal education, low text literacy, or no prior experience with touchscreen interfaces. Designing for this population is not an edge case: it is the primary use case.

The application is designed around a voice-first interaction model throughout. This means:

- **Voice-guided onboarding:** users set up the app entirely by speaking, including selecting their country, language, dialect and trade sector. No typing is required at any stage of onboarding. Prompts are delivered in the user's detected or declared language.
- **Automatic language detection:** the system detects the user's language automatically from the first few seconds of speech, using a combination of on-device fastText classification and Whisper's multilingual identification. Users do not need to select a language manually if they do not wish to.
- **Automatic partner language detection:** when a second trader speaks, the system detects their language independently and routes the translation accordingly. The app identifies two-speaker conversations without any configuration, enabling natural back-and-forth interpretation without the user managing any settings mid-conversation.
- **Voice navigation throughout:** core app functions, including starting a new interpretation, reviewing a previous transcript, accessing help and adjusting settings, can all be triggered by voice command. The interface supports both tap and voice input at every point.
- **Speaker separation and diarisation:** the pipeline identifies distinct speakers within a recording, labels their utterances separately, and routes each to the appropriate translation pathway. This is essential for real two-way trade conversations where both parties speak into a shared device.

From a technical standpoint, automatic language detection uses a lightweight on-device classifier for near-instant identification, with Whisper's language probability distribution as a secondary signal. Speaker diarisation uses a pyannote-based pipeline adapted for noisy market environments. Both components run locally where connectivity permits, falling back to server-side processing otherwise.

The voice-first approach is also a data strategy. Every onboarding interaction, every voice command and every corrected output generates labelled data about how real traders speak, what vocabulary they use, and how they interact with AI-generated audio. This data is qualitatively different from studio-recorded corpora and more representative of real usage conditions.

## 6. End-to-End Data Pipeline

To convert raw product interactions into model-ready training assets, SanKam requires a disciplined data pipeline. The six-stage pipeline below transforms trader interactions into structured, reusable assets that support speech recognition, translation, speech synthesis, confidence scoring and future analytics products.

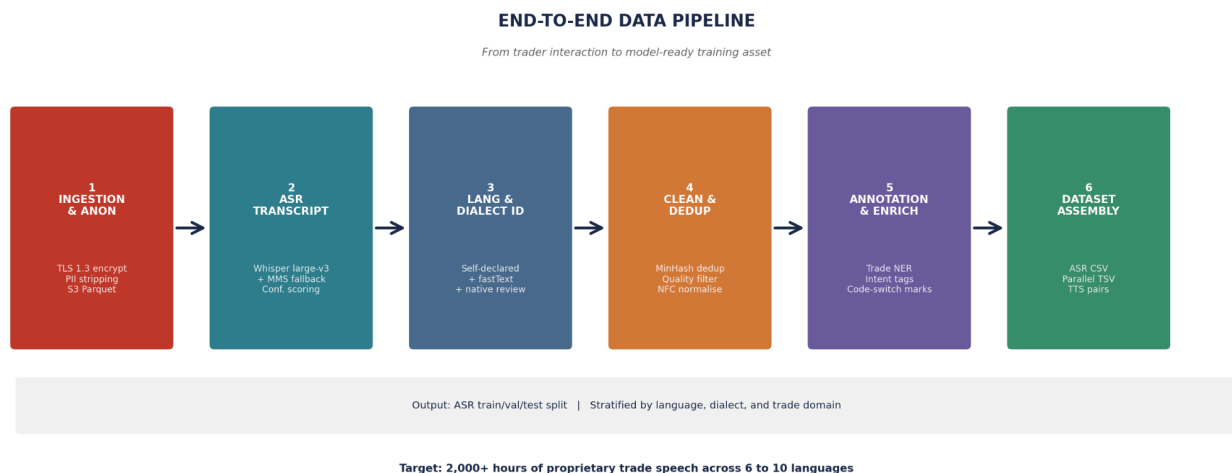


Figure 3: End-to-end data pipeline. The mechanism through which product usage is transformed into proprietary model advantage.

### Stage 1: Ingestion and anonymisation

All data enters through a secure ingestion layer, with TLS 1.3 encryption in transit and encryption at rest. A PII-stripping service runs named entity recognition to replace personal identifiers with anonymised tokens before data enters the training pipeline. Records are written to S3 in Parquet format, partitioned by language, dialect and date (AWS af-south-1, Cape Town).

### Stage 2: ASR and transcription

Voice messages are processed through a dual-model ASR pipeline. Primary: OpenAI Whisper large-v3, supporting 14+ African languages. Fallback: Meta MMS, covering 1,100+ languages. Word-level confidence scoring flags segments below 0.6 for review. Speaker diarisation labels multi-party conversations, enabling per-speaker language identification and separate translation routing.

### Stage 3: Language and dialect identification

Accurate labelling combines three signals: user self-declaration (strongest), a fine-tuned fastText classifier trained on AfroLID and Glot500, and community native-speaker validation. Where automatic classification disagrees with self-declaration, data is flagged. This disagreement data is valuable for mapping dialect boundaries.

## **Stage 4: Cleaning and deduplication**

MinHash-based deduplication, quality filtering (heuristic and trained classifier), Unicode NFC normalisation, and multilingual toxicity filtering. Expected retention rate: 70 to 80% of raw data.

## **Stage 5: Annotation and enrichment**

Clean data is enriched by community annotators: paid native speakers from the same trading communities. Tasks include code-switch span marking, trade named entity recognition (product, price, quantity, location, brand), intent classification (inquiry, offer, counter-offer, acceptance, rejection, logistics) and translation validation.

## **Stage 6: Dataset assembly**

Three model-ready datasets are produced: an ASR fine-tuning dataset (audio and transcript pairs), a parallel translation corpus (source and target sentence pairs), and a text-to-speech fine-tuning dataset (audio and text pairs). All are split 90/5/5 into train, validation and test sets, stratified by language, dialect and trade domain. Test sets are curated by human reviewers.

## 7. Model Strategy: From Managed APIs to Self-Hosted Specialisation

### 7.1 Strategic choice: do not train from scratch

The most credible path is to begin with strong multilingual base models and specialise them through continued pre-training, supervised fine-tuning and alignment using proprietary SanKam data. This is more capital-efficient, more realistic given available resources, and better aligned to the current state of open multilingual models.

### 7.2 Base model selection

| Model               | Parameters                    | African coverage                | Licence        | Assessment                                                                |
|---------------------|-------------------------------|---------------------------------|----------------|---------------------------------------------------------------------------|
| BLOOM / BLOOMZ      | 176B (7B distilled available) | 20+ African languages           | Open RAIL-M    | Strong African coverage; distilled variants viable for fine-tuning        |
| Llama 3.x           | 8B / 70B                      | Limited native; strong transfer | Meta Community | Strong base; requires significant domain adaptation for African languages |
| Aya (Cohere for AI) | 8B / 35B                      | 101 languages including African | Apache 2.0     | Best licence for commercial deployment; strong multilingual baseline      |
| AfroLM / Serengeti  | <1B                           | Purpose-built for African NLP   | Research       | Purpose-built; smaller scale limits general capability                    |

### 7.3 Three-phase training strategy

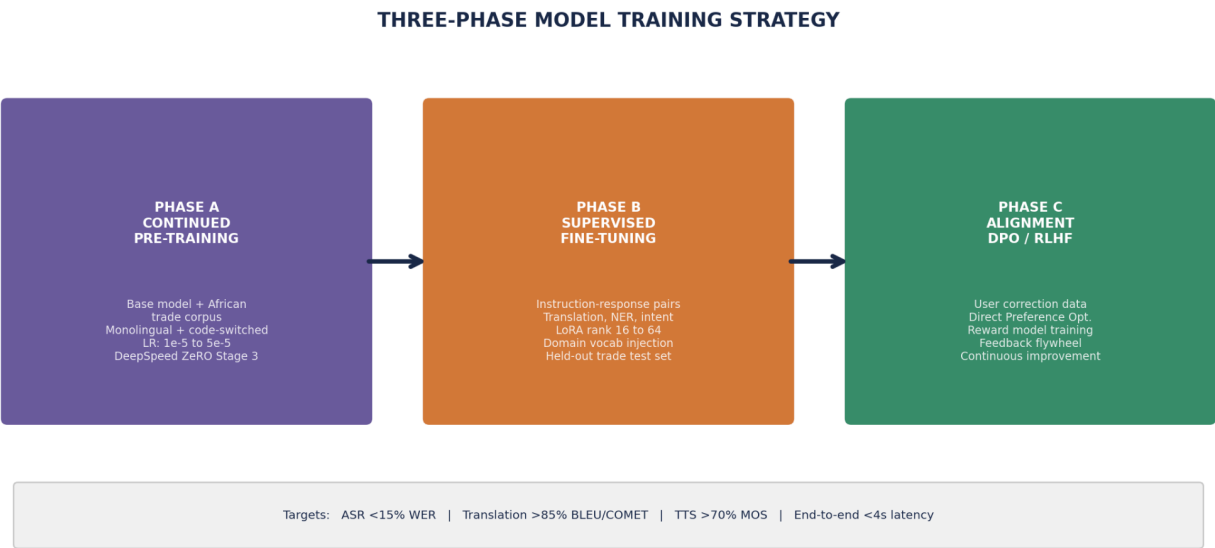


Figure 4: Three-phase model training strategy. Each phase builds on the last, with user correction data becoming the most powerful input in Phase C.

#### Phase A: Continued pre-training

The base model is trained further on a monolingual and code-switched corpus, with a learning rate of 1e-5 to 5e-5 and cosine decay, using DeepSpeed ZeRO Stage 3 on 4 to 8 A100 GPUs. This phase teaches the linguistic patterns, vocabulary and dialectal variations of African trade communication.

#### Phase B: Supervised fine-tuning

The pre-trained model is fine-tuned on instruction and response pairs using LoRA (rank 16 to 64), covering translation, named entity recognition, intent classification and trade communication generation. Domain-specific vocabulary injection is applied, and a held-out trade test set is used for evaluation.

#### Phase C: Alignment (DPO and RLHF)

Using user correction data from the application, a reward model is trained and Direct Preference Optimisation applied. User correction is not merely a UX feature: it is an input into compounding model quality. Every correction a trader makes improves the next model iteration. This is where the data flywheel is most powerful and most defensible as a competitive advantage.

### 7.4 Component-level targets

| Component | Base model                             | Approach                                                                         | Target metric                                                    |
|-----------|----------------------------------------|----------------------------------------------------------------------------------|------------------------------------------------------------------|
| ASR       | Whisper large-v3 / SeamlessM4T encoder | Supervised fine-tuning on trade ASR dataset; LoRA/PEFT; noise-augmented training | 20 to 30% WER reduction over baseline on trade dialogue test set |

| Component      | Base model                               | Approach                                                                            | Target metric                                         |
|----------------|------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------|
| Translation    | NLLB-200-distilled / SeamlessM4T decoder | Domain fine-tuning on parallel trade corpus; LoRA adapters; trade-weighted BLEU     | >85% BLEU/COMET on domain held-out pairs              |
| Text-to-speech | Meta MMS / SeamlessM4T unit decoder      | Adapter tuning; tonal accuracy optimisation for Yoruba; prosody reconstruction loss | >70% MOS naturalness from native speaker panel        |
| End to end     | SeamlessM4T unified model                | Joint fine-tuning; benchmarked against cascaded pipeline                            | Latency <4s; quality parity or better versus cascaded |

## 7.5 Tokeniser adaptation

Standard tokenisers fragment African language words excessively. A Hausa word may split into 5 to 8 sub-tokens, inflating sequence lengths and degrading model performance. A SentencePiece tokeniser trained on the proprietary corpus, merged with the base tokeniser, targets reducing average tokens per word from approximately 4.5 to 1.8 for core target languages. This reduces computational waste and improves model understanding of African language structure.

## 8. Evaluation Framework

A credible language AI programme must define quality in operational terms, not merely benchmark terms. No single metric captures user value. A strong BLEU score does not guarantee comprehension in a noisy market, and a low WER does not guarantee commercial meaning is preserved. The evaluation framework therefore spans four dimensions across three levels: model quality (does the component work in isolation?), user quality (can a trader complete a conversation?) and business quality (does the system perform at a sustainable cost and reliability threshold?).

| Dimension                  | Automated metrics                                                                              | Human evaluation                                                                       | Cadence                                                  |
|----------------------------|------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|----------------------------------------------------------|
| ASR quality                | Word Error Rate (WER) per language and dialect; Character Error Rate (CER) for tonal languages | Native speaker accuracy review on held-out trade dialogue set                          | After each fine-tuning run                               |
| Translation quality        | BLEU, chrF++, COMET; domain-weighted variant for trade terminology                             | Bilingual trade community evaluation: adequacy, fluency, cultural appropriateness      | After each fine-tuning run; monthly on production sample |
| Text-to-speech naturalness | Mel cepstral distortion; pitch trajectory accuracy for tonal languages                         | Mean Opinion Score (MOS) from native speaker panel; comprehension in market conditions | After each text-to-speech fine-tuning run                |
| End to end                 | End-to-end latency; pipeline error rates; fallback frequency                                   | Trader satisfaction; conversation completion rate; intent accuracy                     | Continuous in production; formal quarterly review        |

Quality thresholds are applied language pair by language pair, so each pair is released to traders only once it meets them.

---

## 9. Technical Infrastructure and Deployment

A production architecture for African trade cannot assume stable connectivity or high-end devices. Infrastructure design decisions in Phase 1 have direct consequences for cost, latency, data sovereignty and offline readiness later, so they must be made with production deployment in mind, not prototype convenience.

### 9.1 Offline-first architecture

Offline capability is not a stretch goal for SanKam. It is a baseline requirement. African border posts, open-air markets and rural trade corridors frequently operate with intermittent 2G/3G connectivity or no connectivity at all. A system that fails in these conditions is not production-ready, whatever its quality in controlled settings.

The architecture supports three deployment modes, with automatic switching between them based on available connectivity:

- **Full online mode:** all three pipeline stages (ASR, translation, text-to-speech) run server-side using full-size models. Maximum quality, full language coverage and the lowest device compute requirement. Used when stable connectivity is available.
- **Hybrid mode:** ASR runs on-device using a quantised Whisper Tiny or Small model; translation and text-to-speech run server-side. This reduces bandwidth by eliminating raw audio upload while retaining higher-quality translation and synthesis. Used under limited connectivity.
- **Fully offline mode:** all three stages run on-device using quantised models (TFLite or ONNX Runtime). Quality is reduced versus server-side models, but the system is fully functional with no network dependency. Used when connectivity is unavailable.

On-device model targets are under 500MB in total for the core ASR and translation pipeline, compatible with mid-range Android smartphones (3GB RAM minimum, Snapdragon 600 series or equivalent). This covers the majority of the target device class in African trade markets.

Expected quality trade-offs in fully offline mode are 5 to 10% WER degradation on ASR versus the server-side Whisper large-v3, and an 8 to 15% BLEU reduction on translation versus the distilled NLLB server model. These trade-offs are acceptable given the alternative: no interpretation at all when there is no connectivity.

The switching logic is transparent to the user. A small indicator shows the current mode (online, hybrid or offline). When connectivity is restored, the system automatically upgrades to the higher-quality online pipeline for subsequent utterances, and queues any corrections or feedback generated offline for upload.

Phase 1 includes dedicated engineering time for quantisation research and offline pipeline validation. Specific milestones include: on-device Whisper Tiny validated for Swahili and Hausa by Month 4; the NLLB distilled model quantised to under 200MB in ONNX by Month 5; and the end-to-end offline pipeline tested on the target device class by Month 6.

## 9.2 Cloud infrastructure

The cloud infrastructure spans four layers: the mobile application; an API gateway and routing layer; processing and pipeline services; and a storage and training cluster hosted in AWS af-south-1 (Cape Town) for data residency compliance.

# 10. Roadmap and Phasing

The build is phased so that each stage is gated by evidence from the one before. Timelines are indicative.

| Phase                       | Timeline        | Key milestones                                                                                                                                        |
|-----------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| Phase 1: Foundations        | Months 1 to 6   | Data collection infrastructure deployed; 500 hours recorded across 3 languages; initial Whisper and NLLB fine-tuning; offline pipeline milestones met |
| Phase 2: MVP launch         | Months 7 to 12  | Production MVP; 1 to 2 trade corridor pilots; voice-first onboarding live                                                                             |
| Phase 3: Scale and optimise | Months 13 to 18 | Fine-tuned models on proprietary data; more languages; fully offline inference deployed; enterprise API                                               |

## 11. Risks and Mitigation Strategy

The major risks are treated as design constraints to be actively managed across the programme, not surprises to be handled reactively.

| Risk                                                 | Impact                                                   | Mitigation                                                                                                                             |
|------------------------------------------------------|----------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| Data scarcity in low-resource languages and dialects | Insufficient volume; underperformance on rarer languages | Open African language datasets and community data initiatives; synthetic augmentation; phased rollout by data availability             |
| Low early user adoption; flywheel does not start     | Insufficient data; programme stalls                      | Trade association partnerships for initial acquisition; on-the-ground ambassadors at high-volume border posts; subsidised early access |
| Model bias across dialects, gender and communities   | Platform underserves specific trader communities         | Regular audits across gender, age, dialect and region; community feedback loops built into beta                                        |
| Offline inference infeasible on mid-range devices    | System fails in low-connectivity market conditions       | Dedicated quantisation engineering in Phase 1; progressive enhancement; server-side fallback; target device testing from Month 3       |
| Regulatory fragmentation across markets              | Data protection compliance complexity                    | Design for the strictest standard (POPIA); local legal review per market before launch                                                 |
| Licensing restrictions on model families at scale    | Commercial licensing risk for SeamlessM4T, NLLB and MMS  | Commercial discussions initiated in Phase 1; open alternatives (Aya, BLOOMZ) maintained as fallback                                    |

---

## 12. Conclusion

SanKam addresses a real, documented and economically meaningful problem. Research estimates the trade cost equivalent of language barriers at around 5 to 12% of transaction value for informal traders, and those barriers structurally exclude a significant proportion of the continent's population from the benefits of intra-African trade. No production-grade solution currently exists for this market at an accessible price point.

Phase 0 has confirmed early technical feasibility. African-focused translation models outperform generic alternatives on culturally nuanced language pairs; an end-to-end speech pipeline runs across 10 languages covering all major African trade blocs; and the resulting findings define the Phase 1 agenda with specificity. The path to production is substantial but well defined.

The distinctive characteristic of SanKam's development model is that the product and the R&D programme are not separate activities. The platform generates the data that trains the model, and the model improves the platform. Every trade conversation is both a commercial exchange and a training opportunity. That alignment between product utility and model development is what makes the competitive position compound over time and difficult to replicate.

The next steps are to build the data collection infrastructure and to progress offline pipeline research in parallel.

# Appendix A: Target Language Coverage

The table below sets out the languages targeted in the original R&D plan (March 2026), with ISO codes, key dialect variants, estimated speaker populations and primary trade regions. Language prioritisation is based on speaker population, trade corridor volume and the availability of existing training data. SanKam’s current MVP delivery plan covers twelve languages; see About this edition.

| Language       | ISO 639-3 | Key dialects                         | Speakers      | Trade regions                      | Phase   |
|----------------|-----------|--------------------------------------|---------------|------------------------------------|---------|
| English        | eng       | Standard, Nigerian, Kenyan           | Global        | All corridors (bridge language)    | Phase 1 |
| Swahili        | swh       | Standard, Coastal, Congolese, Kenyan | 100M+         | Kenya, Tanzania, DRC, Uganda       | Phase 1 |
| Hausa          | hau       | Kano, Sokoto, Niamey, Ghanaian       | 80M+          | Nigeria, Niger, Ghana              | Phase 1 |
| Yoruba         | yor       | Standard, Ijebu, Ekiti, Ondo         | 47M+          | Nigeria, Benin, Togo               | Phase 1 |
| French         | fra       | Standard, West African               | Widely spoken | ECOWAS Francophone, CEMAC          | Phase 1 |
| Zulu           | zul       | Standard, Urban Johannesburg         | 12M+          | South Africa, Lesotho              | Phase 1 |
| Arabic         | arb       | Modern Standard, Gulf, Sudanese      | 350M+         | North Africa, Arab trade corridors | Phase 1 |
| Amharic        | amh       | Standard, Gojjam, Gondar, Shewa      | 57M+          | Ethiopia, Eritrea                  | Phase 1 |
| Twi            | twi       | Asante, Fante, Akuapem               | 9M+           | Ghana                              | Phase 1 |
| Portuguese     | por       | Standard, Mozambican, Angolan        | 270M+         | Mozambique, Angola, Cape Verde     | Phase 1 |
| Igbo           | ibo       | Central, Onitsha, Owerri, Ngwa       | 45M+          | Nigeria (Southeast), Cameroon      | Phase 2 |
| Wolof          | wol       | Standard, Urban Dakar, Gambian       | 12M+          | Senegal, Gambia                    | Phase 2 |
| Pidgin English | pcm       | Nigerian, Cameroonian, Ghanaian      | 75M+          | West Africa broadly                | Phase 2 |

## Appendix B: Glossary

Definitions for readers less familiar with machine learning or African trade contexts.

|                       |                                                                                                                                                                                        |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>AfCFTA</b>         | African Continental Free Trade Area: the framework agreement creating a single African market covering 55 countries and 1.3 billion people.                                            |
| <b>ASR</b>            | Automatic Speech Recognition: converting spoken audio into text. Also called speech-to-text.                                                                                           |
| <b>BLEU</b>           | Bilingual Evaluation Understudy: an automated translation quality metric measuring overlap between machine and human reference translations. Higher is better.                         |
| <b>Code-switching</b> | Alternating between two or more languages or dialects within a single conversation or sentence. Common in African trade contexts.                                                      |
| <b>COMET</b>          | Crosslingual Optimised Metric for Evaluation of Translation: a neural translation quality metric that correlates more closely with human judgement than BLEU.                          |
| <b>CPT</b>            | Continued Pre-Training: adapting an existing base language model on a new domain corpus without training from scratch.                                                                 |
| <b>DPO</b>            | Direct Preference Optimisation: an alignment technique that uses human preference data (for example user corrections) to improve model outputs without explicit reward model training. |
| <b>Diarisation</b>    | Speaker diarisation: segmenting audio by speaker identity, enabling multi-speaker transcription and per-speaker language identification.                                               |
| <b>LoRA</b>           | Low-Rank Adaptation: a parameter-efficient fine-tuning method that inserts small trainable matrices into a frozen model, significantly reducing compute requirements.                  |
| <b>MOS</b>            | Mean Opinion Score: a human evaluation metric for audio quality, typically scored 1 to 5 by native speaker panels.                                                                     |
| <b>NER</b>            | Named Entity Recognition: identifying and classifying named entities in text such as products, prices, quantities, locations and brands.                                               |
| <b>NLLB</b>           | No Language Left Behind: Meta's multilingual translation model covering 200+ languages, including 55+ African languages.                                                               |
| <b>ONNX</b>           | Open Neural Network Exchange: a format for machine learning models that enables deployment across different hardware and runtimes.                                                     |
| <b>PII</b>            | Personally Identifiable Information: data that can identify an individual, such as a name, phone number or location. Stripped from training data before processing.                    |
| <b>RLHF</b>           | Reinforcement Learning from Human Feedback: a training approach using human preference signals to align model outputs with desired behaviour.                                          |
| <b>SeamlessM4T</b>    | Meta's unified speech-to-speech and speech-to-text model supporting approximately 100 languages. A candidate for end-to-end pipeline unification in Phase 2 and beyond.                |

|               |                                                                                                                                                       |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>SFT</b>    | Supervised Fine-Tuning: training a model on labelled instruction and response pairs for a specific task.                                              |
| <b>TFLite</b> | TensorFlow Lite: a lightweight inference framework for running machine learning models on mobile and edge devices.                                    |
| <b>TTS</b>    | Text-to-Speech: converting written text into spoken audio. Also called speech synthesis.                                                              |
| <b>WER</b>    | Word Error Rate: an ASR quality metric measuring the percentage of words incorrectly transcribed relative to a reference transcript. Lower is better. |

---

## References

Adelani, D. I., Neubig, G., Ruder, S., Rijhwani, S., Beukman, M., Palen-Michel, C., Lignos, C., Alabi, J. O., Muhammad, S. H., Nabende, P., and Klakow, D. (2022). MasakhaNER 2.0: Africa-centric transfer learning for named entity recognition. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics.

African Development Bank. (2024). *African Economic Outlook 2024*. African Development Bank Group. [afdb.org/en/knowledge/publications/african-economic-outlook](https://afdb.org/en/knowledge/publications/african-economic-outlook)

Brenton, P., and Isik, G. (Eds.). (2012). *De-fragmenting Africa: Deepening regional trade integration in goods and services*. World Bank. [documents.worldbank.org](https://documents.worldbank.org)

International Labour Organization. (2018). *Women and Men in the Informal Economy: A Statistical Picture* (3rd ed.). ILO.

Ojo, J., Ogundepo, O., Oladipo, A., Ogueji, K., Lin, J., Stenertorp, P., and Adelani, D. I. (2025). AfroBench: How good are large language models on African languages? *Findings of the Association for Computational Linguistics: ACL 2025*. [arxiv.org/abs/2311.07978](https://arxiv.org/abs/2311.07978)

OpenAI. (2024). *GPT-4o system card*. OpenAI. [cdn.openai.com/gpt-4o-system-card.pdf](https://cdn.openai.com/gpt-4o-system-card.pdf)

---

© 2026 SanKam Limited. Public edition of the SanKam R&D white paper, prepared by Brdge for SanKam Limited. Figures, plans and timelines are indicative and reflect the R&D position at March 2026. [san-kam.com](https://san-kam.com) · [info@san-kam.com](mailto:info@san-kam.com)