

Textify Voice: Product & Functional Specification

“ Status: draft for implementation · Owner: Joshua Saimon Sakweli (QBIT SPARK)
Companion doc: `ARCHITECTURE.md` (components, data model, APIs, infra). This file is written so a new engineer, or a new AI coding session, can pick up the work with no prior context.

1. Background

Textify Africa (textify.africa) is an existing bulk SMS + WhatsApp platform, **written in Go**, live in Tanzania, Uganda and Nigeria. Its main model is **white-label**: agencies and resellers ("partners") brand the platform as their own, set their own prices, and onboard their own customers. Customers top up a **prepaid wallet** with mobile money or card.

We are adding a third channel: a **call center / business phone line** product, similar in spirit to SemaCall (semacall.com): *"One business. One number. Every team."* Customers call one number; the call reaches the right person (or an AI answers) in the customer's language.

The voice capability is built as a **separate service ("Voice Service") in Java Spring Boot**. Textify consumes it as a client. The Voice Service must not know about Textify's users, auth or money.

2. Goals

1. A Textify customer can get a Tanzanian phone number and have every incoming call answered, by an **AI receptionist** (Swahili first, English supported) or by **human agents in the browser**.
2. AI can **hand the caller over to a human** when needed.
3. Every call is **logged, optionally recorded, transcribed and summarized**.
4. **Missed calls trigger an automatic SMS/WhatsApp** follow-up (Textify already does messaging).
5. Usage is **billed per minute from the existing Textify wallet**.

6. The Voice Service is **reusable** by other QBIT SPARK products later (JikoXpress, Mauzodukani, NexGate) without changes.

3. Non-goals (for now)

- No video calls over phone lines (LiveKit SIP does not support video over SIP anyway).
- No user accounts, roles, plans or pricing inside the Voice Service.
- No approval workflows anywhere. Keep everything self-service and simple.
- Countries other than Tanzania (Uganda/Nigeria later, using the same design).
- Building our own telco connection. We use Africa's Talking for numbers and the PSTN.

4. Key terms (plain language)

Term	Meaning
PSTN / telco network	The normal phone network (Vodacom, Airtel, Yas, Halotel, TTCL).
SIP	The protocol internet phones use to set up calls (ring, answer, hang up). Like HTTP but for calls. Voice audio travels separately over RTP .
SIP account	One login for one phone/softphone (like an email account but for calls).
SIP trunk	A connection carrying many calls at once between two systems over SIP. Here: Africa's Talking ≠ our LiveKit .
Human agent	A staff member who answers calls (in the browser app).
AI agent	A bot that talks on the phone: listens → thinks → speaks.
STT / LLM / TTS	Speech-to-text, the language model (the "brain"), and text-to-speech (the voice).
Realtime / speech-to-speech model	One model that hears and speaks directly (e.g. OpenAI Realtime).
LiveKit room	A live audio session. Each phone call = one room. The caller joins as a "SIP participant"; AI or humans join the same room.
Dispatch rule	LiveKit config that decides which room an incoming call goes to and which AI agent joins.
DTMF	Keypad presses during a call ("press 1").
Cold / warm transfer	Cold: just forward the call. Warm: AI briefs the human first, then hands over.

5. Providers we use and why

Provider	Role	Why
Africa's Talking (AT)	Tanzanian phone numbers + connection to all TZ telcos + SIP trunk	Local numbers and licensing. LiveKit's own phone numbers are US-only.
LiveKit (self-hosted, open source)	Media server: rooms, SIP, browser audio, recording, transfers	Already used in production by us for NexGate (video, spaces, audio calls). Self-hosted = no per-minute LiveKit fee.
OpenAI	AI brain + voice (Realtime model), transcription, summaries	Best overall quality. Swahili quality must be tested; TTS can be swapped.

6. Costs (Africa's Talking Tanzania, as published on africastalking.com/pricing, Oct 2026)

Item	Price
Regular number (Airtel)	Setup TZS 100,000 + 18% VAT; TZS 35,000/month + 18% VAT
Toll-free number	Same as regular, plus TZS 170,000 deposit
Regular incoming	TZS 17 / min
Regular outgoing	TZS 60 / min
Toll-free incoming	TZS 200 / min
SIP agent leg	TZS 4.30 / min
SIP trunk	TZS 8.61 / min
Conference	TZS 8.61 / min
TTS (Google, via AT)	per character

6.1 How the costs combine

Every call has **leg 1** (customer ⇌ our AT number) and, if someone answers, **leg 2** (our number ⇌ whoever answers).

Customer —leg 1—> AT number —leg 2—> answerer
17 in / 60 out SIP trunk to LiveKit: 8.61

forward to a mobile: 60 (avoid)

Scenario	Approx. cost per minute
Inbound call into LiveKit (AI or browser agent)	17 + 8.61 $\approx$ TZS 25.61 + AI cost if AI answers
Outbound call from LiveKit to a customer	60 + 8.61 $\approx$ TZS 68.61 (to confirm)
Inbound forwarded to a staff mobile (fallback only)	17 + 60 $\approx$ TZS 77

Key facts already agreed:

- **Cost follows call minutes, not the number of agents.** Idle agents cost nothing, so customers can add unlimited agents.
- **Avoid forwarding to mobiles.** Agents answer in the browser/app.
- **Design decision:** money is not the constraint, so **all calls go through the SIP trunk into LiveKit** (one call path, simpler code). We don't split IVR onto AT XML to save the 8.61.

6.2 AI cost (on top of telco)

An AI-answered minute adds OpenAI cost (Realtime, or STT + LLM + TTS). This is several times the telco cost per minute. **Exact per-minute figures must be measured during the PoC** from real calls (token/audio usage logged per call).

6.3 Indicative selling prices (suggestion, to be finalized after PoC)

Item	Suggested price
Number rental	TZS 60-75k / month
Inbound minute, human-answered	TZS 40-50 / min
Outbound minute	TZS 90-100 / min
AI-answered minute	TZS 150-250 / min
AI summary add-on	+ TZS X / min (set after measuring)

Pricing lives **in Textify only**. The Voice Service only reports usage.

7. Users & personas

- **Business owner / admin** (Textify customer): gets a number, sets the AI greeting and business info, adds agents, sees calls, recordings and summaries.

- **Human agent:** logs into the Textify web app, sets themselves available, answers calls in the browser, makes click-to-call outbound calls.
- **Caller:** a member of the public calling the business number from any TZ network.
- **Partner (reseller):** later phase. Resells voice under their own brand (same white-label model as SMS).

8. Functional requirements

8.1 Numbers

- FR-N1: Textify admin (QBIT/Textify ops) obtains numbers from AT and registers them in the Voice Service.
- FR-N2: A number is assigned to exactly one **account** (Textify's tenant id, opaque to the Voice Service).
- FR-N3: Each number has an **answer mode**: `AI_FIRST` (AI answers, can hand over), `HUMAN_ONLY` (ring agents), `AI_ONLY`.
- FR-N4: Reassigning or releasing a number is a single API call. No approval.

8.2 Inbound calls

- FR-I1: Caller dials the number → call arrives over the SIP trunk → LiveKit creates a room → Voice Service records a call `RINGING`.
- FR-I2: Before answering, the Voice Service asks Textify "**can this account take a call?**" (prepaid balance). If Textify is unreachable, **allow the call** (fail-open) and bill later.
- FR-I3: If the account can't pay: play a short polite message and end the call (no AI cost).
- FR-I4: Play a **recording consent notice** when recording is on ("Simu hii inaweza kurekodiwa kwa ajili ya ubora wa huduma"; in English: "This call may be recorded for quality of service").
- FR-I5: `AI_FIRST` / `AI_ONLY` → AI receptionist joins and greets with the account's greeting.
- FR-I6: `HUMAN_ONLY` → ring the account's available agents (ring group). If nobody answers within N seconds (default 30) → voicemail **or** AI fallback (configurable) → mark `MISSED` if nobody handled it.
- FR-I7: Caller keypresses (DTMF) are supported for simple menus ("press 1 for sales").

8.3 AI receptionist

- FR-A1: Speaks **Swahili first**, switches to English if the caller does. Handles code-switching (Swahili and English mixed).
- FR-A2: Uses the account's **instructions** (business info, FAQs, hours, tone) supplied by Textify.
- FR-A3: Short, spoken-style answers. Never invents prices or policies. If unsure → offer a human.

- FR-A4: Tools the AI can use (v1): `transfer_to_human`, `take_message`, `end_call`. Later: account-defined tools (check order, book appointment) via webhooks to the client app.
- FR-A5: Handover: AI calls `transfer_to_human` → Voice Service rings available agents → first to accept joins the room → AI briefly introduces the context (warm transfer), then leaves.
- FR-A6: If no agent is available: AI takes a message (name, number, reason) → stored on the call → event to Textify.
- FR-A7: Barge-in/interruptions and turn detection are handled by LiveKit Agents.

8.4 Human agents (browser)

- FR-H1: Agents are identified only by an opaque `agentRef` from Textify. Textify owns the user, login and role.
- FR-H2: Agent sets status `AVAILABLE` / `BUSY` / `OFFLINE` through Textify, which forwards it to the Voice Service.
- FR-H3: Incoming call → available agents in the group get a **ring** in the browser (data message over LiveKit). Accept → join the call room. Reject or timeout → next.
- FR-H4: Ring strategy: `ALL_AT_ONCE` (default) or `ROUND_ROBIN`.
- FR-H5: Agent can mute, hold, hang up, and transfer to another agent.
- FR-H6: **Click-to-call** outbound from the dashboard: agent clicks a number → Voice Service dials out via the trunk → agent and customer in one room.
- FR-H7: Fallback transfer to an external mobile is allowed (costs about TZS 60/min). Off by default.

8.5 Recording, transcription, summary

- FR-R1: Recording on/off per number (default **on**, with consent notice).
- FR-R2: After the call: audio file stored in our object storage (never rely on provider links).
- FR-R3: Transcription (Swahili/English) → **summary** containing:
 - 2-3 line summary
 - customer intent, resolved yes/no
 - follow-up actions
 - sentiment (positive / neutral / negative)
 - topic tag (sales, support, complaint, inquiry, other)
- FR-R4: Transcript + summary attached to the call record; event `summary.ready` sent to the client app.
- FR-R5: Retention: recordings deleted after **90 days** by default (configurable per account). Tanzania Personal Data Protection Act applies.
- FR-R6: Summary is a billable add-on (Textify decides price; Voice Service reports `summaryMinutes`).

8.6 Missed calls & follow-ups

- FR-M1: A call is `MISSED` if no human or AI handled it (e.g. HUMAN_ONLY with nobody answering and no fallback, or the caller hung up while ringing).
- FR-M2: Voice Service emits `call.missed` with the caller number. **Textify** sends the SMS/WhatsApp follow-up ("Samahani tumekukosa, tutakupigia tena"; in English: "Sorry we missed you, we'll call you back") using its existing messaging.

8.7 Usage & billing

- FR-B1: On every call end, the Voice Service emits `call.ended` with usage broken down by leg (seconds): `inboundSeconds`, `outboundSeconds`, `trunkSeconds`, `aiSeconds`, `externalTransferSeconds`, `recorded` (bool), `summarySeconds`.
- FR-B2: Textify prices and deducts from the wallet. The Voice Service never deals with money.
- FR-B3: Events are **idempotent** (event id) and retried until Textify returns 2xx.

8.8 Dashboard (built in Textify, data from the Voice Service)

- Call log: time, direction, caller, number, handled by (AI / agent), duration, status, cost (Textify-computed).
- Call detail: audio player, transcript, summary, AI message taken.
- Number settings: answer mode, greeting, AI instructions, recording on/off, ring group, timeout, fallback.
- Agents: availability, calls handled.
- Simple stats: calls per day, answered vs missed, AI-handled vs human, average duration.

9. Non-functional requirements

Area	Requirement
Latency	AI response start < ~1 s after the caller stops speaking (target). Call connect < 3 s.
Availability	Voice Service must keep taking calls if Textify is down (fail-open balance check, queued webhooks).
Isolation	Dedicated LiveKit box for Textify Voice. Not shared with NexGate.
Security	SIP/RTP ports open only to Africa's Talking IPs. Service tokens between apps. Signed webhooks. No user PII beyond phone numbers.
Privacy	Consent notice, retention policy, recordings in our storage, access only via the client app.

Area	Requirement
Language	Swahili first-class. Test quality with real Tanzanian callers before selling AI.
Simplicity	No approval workflows, no internal OAuth, minimal moving parts.

10. Phases

Phase 0: Proof of concept (goal: one real call end-to-end)

1. Confirm with AT (see §11). Get one TZ number + SIP trunk.
2. Dedicated box: LiveKit server + LiveKit SIP + Redis. Firewall restricted to AT.
3. Inbound trunk + dispatch rule → Spring Boot dials OpenAI Realtime into the room over SIP → AI greets in Swahili.
4. Measure: latency, audio quality, OpenAI cost per minute.
5. **Swahili test**: 10-20 real people (Swahili, English, mixed). Judge understanding and voice. Swap TTS if the voice is weak.

Phase 1: MVP (sellable)

- Spring Boot Voice Service: accounts, numbers, config, calls, events, LiveKit integration.
- AI receptionist with handover to browser agents.
- Browser agent client (in Textify's web app) + ring groups + click-to-call.
- Recording + storage + retention.
- Balance check + `call.ended` usage events → wallet deduction in Textify.
- Missed-call SMS/WhatsApp.

Phase 2

- Transcripts + AI summaries (if not already in MVP).
- Account-defined AI tools (webhooks to the client app: check order, booking).
- Voicemail, business hours / after-hours behaviour.
- Stats dashboard.

Phase 3

- **Voice broadcast** (recorded message to a contact list: "bulk voice", fits Textify's bulk DNA). Check TZ calls-per-second limits.

- **AI outbound campaigns** (reminders that talk back).
- **White-label voice for partners.**
- Other countries (Uganda/Nigeria) through AT.
- Other QBIT products as clients of the Voice Service.

11. Open questions for Africa's Talking (must answer before Phase 1)

1. Can the TZ SIP trunk **route inbound calls for our number to an external SIP URI** (our LiveKit SIP endpoint)? Note: LiveKit SIP does **not support SIP REGISTER**, so AT must send calls to us directly.
2. Can it **accept outbound calls from our server IP** (and what auth: IP allowlist or username/password)?
3. What are **AT's SIP signalling and media IP addresses** (for our firewall allowlist)?
4. Exact billing for **inbound to trunk**: 17 + 8.61? Any extra?
5. Exact billing for **outbound via trunk**: 60 + 8.61? Any extra?
6. Is the **SIP agent TZS 4.30** charged on inbound calls too? Any setup/monthly fee for SIP agents?
7. Are **unanswered rings** to SIP endpoints billed? 7b. The listed regular number is **Airtel**. How do callers on Vodacom, Yas, Halotel and TTCL reach it, and at what cost to them and to us?
8. **Calls-per-second (CPS)** limits and pricing in Tanzania (needed for broadcast).
9. Supported **codecs** (G.711 PCMU/PCMA expected) and DTMF method (RFC 2833/4733).
10. Does reselling voice to our customers need any **TCRA licence** on our side, or is AT's licence enough? Get it in writing.
11. Number porting / getting several numbers quickly; toll-free availability.

AT is **not** on LiveKit's list of tested SIP providers (Twilio, Telnyx, Exotel, Plivo, Wavix, Sinch, didlogic). LiveKit is designed to work with any SIP provider, so the PoC must prove compatibility.

12. Risks

Risk	Mitigation
AT trunk can't deliver to an external SIP URI	Verify first (\$11 Q1). Fallback: AT Voice XML + AT SIP agents for humans; AI limited.
Swahili STT/TTS quality poor	Early real-user test; swap TTS provider; keep human fallback prominent.
AI cost higher than expected	Measure in PoC; short answers; pre-recorded greetings; cheap model for simple turns; price AI minutes higher.

Risk	Mitigation
SIP fraud/attacks	Firewall to AT IPs only; LiveKit inbound trunk <code>allowed_addresses</code> + numbers.
Textify outage blocks calls	Fail-open balance check; queued + retried webhooks.
Regulatory (TCRA, PDPA)	Written confirmation from AT; consent notice; retention policy.
Single box failure	Phase 1 single box is OK; add a second node (LiveKit supports multi-node via Redis) when volume grows.

13. Decisions log

#	Decision
D1	Numbers and telco via Africa's Talking ; media via self-hosted LiveKit ; AI via OpenAI .
D2	All calls go through the SIP trunk into LiveKit (no split IVR on AT XML), for one call path.
D3	Agents answer in the browser (no forwarding to mobiles by default).
D4	Dedicated LiveKit box for Textify Voice, separate from NexGate's LiveKit.
D5	Separate Voice Service in Java Spring Boot . Knows no users/auth/money; owns numbers, routing, calls, recordings, usage events.
D6	Textify (Go) owns users, tenants, partners, wallet, pricing, dashboard, messaging follow-ups.
D7	Preferred: no Python. OpenAI Realtime joins the LiveKit room as a SIP participant, controlled from Spring Boot via OpenAI's call-control REST + WebSocket (ARCHITECTURE §2.5). Fallback: thin Python LiveKit Agents worker (§2.2) if OpenAI SIP interop or Swahili voice quality fails the PoC.
D8	Billing is per minute from the Textify wallet; Voice Service emits usage events only.
D9	Recording default on with consent notice; 90-day retention default.
D10	No approval workflows anywhere.

Revision #1

Created 21 October 2026 13:05:36 by Admin Qbit

Updated 21 October 2026 13:06:07 by Admin Qbit