TRUSTAM DATA - DATA CARD

Market conversation, Nigerian Pidgin

Summary

Natural two-speaker conversations between traders and regular customers, recorded on clip-on mics.

Consented conversations in Lagos markets: pricing, haggling, small talk. Each speaker consents separately; recorded on two clip-on mics. Transcribed, time-aligned layers available.

At a glance

Modality: audio, text Source: task capture

Hours available now: 0 Collectable on request: 200 Status: collect to order

Trades: market trading

Languages: pcm, en, yo

Labels: raw, transcribed Label layers: transcripts

Formats: WAV 48 kHz / 16 bit, JSON metadata per clip, Transcripts (JSON, time aligned) when labelled

Price: on request (sales@trustamai.com).

How it is collected

Recorded in Nigeria by paid, ID-checked adults (BVN or NIN) on the TrustAm Recorder app: 1080p H.264, with motion (IMU) and coarse GPS sidecars.

Every recording passes automatic checks on every minute (file, black/frozen, duplicates, hands, sensors); people review sampled and flagged minutes and grade them A / B.

Recorders are paid per approved minute; nobody pays to join.

Consent and privacy

Each recorder agrees on video to licensing for AI training, in English or Pidgin; the consent version is in every delivery manifest.

Bystanders' faces and number plates are blurred before delivery; children are never the subject; ID numbers and private screens are removed.

A recorder can withdraw: future licensing stops and buyers are asked to delete their copies (licence clause). Locations are coarse (about 1 km) unless a workplace agreed.

Metadata fields per clip

clip_id (string) - Stable clip ID

duration_s (number) - Clip length in seconds

trade (string) - Trade slug

contributor_id (string) - Pseudonymous contributor ID

consent_id (string) - Consent record ID

consent_version (string) - Consent text version

capture_date (date) - Capture date (WAT)

device_model (string) - Phone model

qa_score (number) - 0-1 review score

language (string) - Language code (en, pcm, yo, ig, ha)

speakers (number) - Number of consented speakers

Licence

Non-exclusive licence for AI training. Every contributor is ID-checked and consented on video in their own language; bystander faces are blurred; every minute is reviewed by a person. Full licence: /data/legal/licence (draft until our lawyer delivers the final text).