

Method and Data Notes

Digital Personas

Decodis' Final Analysis and Insights Deck for:

- Urban Kenya
- Rural Kenya
- Lagos
- Urban Northern Nigeria
- Rural Northern Nigeria
- Urban Senegal
- Rural Senegal

JMay 2026



Digital Personas: Data Collection Methodology Summary

- **Data Collection:** Voice-led surveys where voice actors pre-record questions in local language in a friendly casual voice.
- **Respondent requirements:** Respondents needed to know how to answer a phone, speak into it and press keypads. This enabled participation regardless of literacy levels, critical for reaching women with limited formal education
- **Geographic coverage:** Kenya (Nairobi, Kwale, Nakuru), Nigeria (Northern Nigeria: Kano & Kaduna; Lagos), Senegal (Dakar, Diourbel, Kolda, Saint Louis, Thies)
- **Local language adaptation:** Surveys conducted in respondents' preferred languages (Swahili, Hausa, Yoruba, Wolof) with professional translation for analysis
- **Sample frame:** Recruited women with existing phone access (focus on understanding digital lives, not measuring basic phone penetration)
- **Sampling procedures:** Sampling across urban/rural and vulnerability levels (Pathways framework)
- **Modular survey structure:** M1 (Demographics & Device Journey), M2 (Usage & Learning), M3 (Skits on Ownership, Safety, Wellbeing)
- **Two response types:** Keypad questions capture quantitative prevalence; open-ended voice responses and qualitative interviews capture contextual nuance
- **Skit-based exploration:** M3 audio vignettes (e.g., Wambui/Zara scenarios) explore sensitive topics (ownership, harassment, control) indirectly, reducing social desirability bias

Data collection: Response rates

	Urban Kenya	Rural Kenya
Recruited	429	1563
Response rate	94%	55%

	Lagos	Urban NN	Rural NN
Recruited	602	490	695
Response rate	75%	82%	80%

	Urban Senegal	Rural Senegal	Thies
Recruited	659	720	311
Responded	76%	81%	89%

Note: Recruited refers to the number of people recruited to the study. These have been split by rurality across the regions. In Senegal, however, the recruitment data does not include rurality. While there is a one-to-one relationship of region to rurality across most of the sample (all respondents from Dakar, for example, are considered urban), the region of Thies is not consistently rural or urban. Therefore, the region of Thies has been split out from the rest of the sample in this report.

Recruitment

- Field officers were deployed across all regions to recruit respondents in line with the predefined regional sample sizes. Recruitment was conducted in person, with field officers explaining the purpose of the study, confirming eligibility and getting consent.
- During recruitment, respondents selected a preferred date and time for survey participation to ensure availability and reduce non-response.

Nudges and Follow-ups

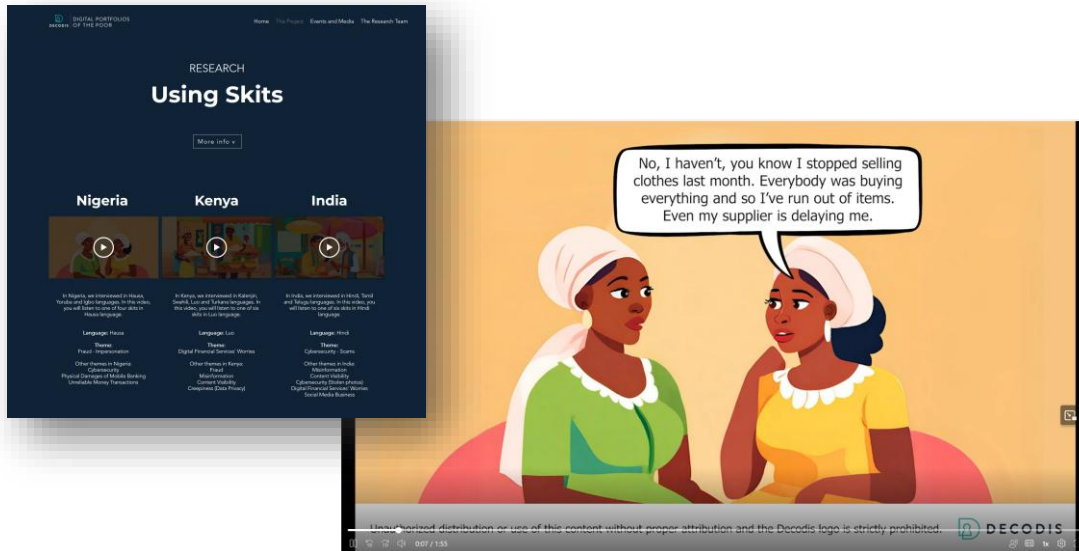
- Prior to survey execution, field officers sent reminders to respondents via phone calls. These reminders served to confirm availability, reinforce consent, and prepare respondents for the voice-based survey.
- Respondents who had not started or completed the survey received follow-up nudges and reminders from the field officers.
- In some cases, field officers conducted in-person follow-ups to resolve access, technical, or comprehension challenges.
- Incentives paid after each module

Survey Calls and Voice Data Collection

Surveys were administered using two ways:

- IVR (Interactive Voice Response):** Respondents received an automated phone call and answered questions using their voice. The IVR system allowed respondents to provide or withdraw consent, pause and resume the survey, or complete it at a later time. Automated calls were scheduled and sent based on the preferred hours indicated by respondents during recruitment. This is our typical methodology and was used in Nigeria and Kenya.
- Survey link:** Respondents received a survey link via WhatsApp or SMS, enabling them to record voice responses at their convenience, with the flexibility to pause and complete the survey later. Phone call costs were prohibitive in Senegal and

Data collection: Using audio skits to capture qualitative responses at scale



CTRL+click on the image to see examples from Digital Portfolios of the Poor, another study on digital trust funded by the Gates Foundation.

Audio skits can explore different aspects of normative frameworks.

Decodis uses between 4-6 skits in studies to capture **very rich perspectives from respondents.**



Responses to questions after audio skits were **three times as long** as other responses.

Note: Videos are for illustration and translation purposes. Respondents were only exposed to audio skits via phone call.

Data processing: Minimal unintelligible response recordings

	Urban Kenya	Rural Kenya
Total # of audio clips	53,460	113,388
% unintelligible clips	2%	1%

	Lagos	Urban NN	Rural NN
Total # of audio clips	59,020	52,520	72,280
% unintelligible clips	5%	9%	9%

How we analyzed the voice clips for these results

Other than for Swahili, where we used Decodis’ own automated translation model, all audios were manually translated to English by local translators.

For the tables to the left, we tagged any response that the translators marks as "Bad quality", "Unclear", or "Noise."

To extract this information for Senegal, we would need to follow a different procedure due to how the data was collected and recorded using a different system.

Data analysis: Using AI to code qualitative data at scale

1. Random selection of approximately 30 responses for each question to manually code to saturation.
2. Write a prompt telling the model to find the manually coded categories within the whole sample.
3. Iterate fine-tuning codes in the way that matches the data most robustly.

Prompt-writing after saturation

****Context****

*The following texts are responses to questions about the risks and benefits of Whatsapp for business, online banking, POS transactions, as well as questions about whose responsibility it is to ensure the safety, security, and access to accurate information about navigating digital platforms.
*These questions were asked to respondents in a survey about digital concerns, behaviours, practices and digital trust to respondents in Nigeria.

****Task****

Based on the context, tag the response to the appropriate category based on what the respondent says about the risks of using online banking, POS transactions or platforms like Whatsapp for business.

****Categorization Scheme**** (A LIMITED LIST IS SHOWN HERE)

UnauthorizedPlatformAccess – Hacking of WhatsApp or bank accounts due to lack of 2FA, malware, or SIM swap.

CyberFraud – Fears of hackers, phishing, impersonation calls, and information theft through digital channels. *Identity&ProfileTheft* – Impersonation on platforms like WhatsApp, with fake profiles used to scam others. *ConnectivityFailures* – Frequent loss of signal, network downtime, or poor internet disrupting transactions, causes anxiety.

FakeDelayedAlerts – Payment confirmations are delayed, forged, or never arrive, leading to confusion or fraud.

FailedDuplicate Transactions – Errors like double debits, failed reversals, and incorrect recipient transfers requiring stressful bank follow-ups.

****Output Instructions****

*Label the response with the relevant category name as listed in the categorization scheme

*If a response falls under multiple classes, string each with an underscore delimiter.

*Each response should receive only one full output.

*Do Not include any other information than what is in the output.

Prompts specified to tag each response

Resp ID	Gender	Age	Region	Income	Transcription of response	Tags
GN_hau_u0840	Male	25-34	Kano	Below ₦50,000	Because the account is registered with the person's main phone number and his identification and it's good to use 2factor Authenticator to avoid hackers. The best way to spread the knowledge is through social media platforms like WhatsApp. It gives you room for making your business to all the people you interact with and through referrals or reposting of your goods. Hacking is the risk of it. They just guess numbers if someone is not lucky will scam him. There's less or no charges and no delay. Someone has to be very careful while making online transactions or filling of forms. Transfer. It's a risky thing as it's also dangerous to them because they won't like it on a transaction between your bank and their banks. POS transactions is exactly as bank's transaction. The only disadvantage is that don't ever make transactions without collecting the transaction slip.	"Personal Responsibility" "Telecom Provider Accountability"

Data reporting: Funnel Note

Along with each chart, a "funnel table" is provided. These values reflect the "funneling" of participants through successive layers of filtering. At the top layer is the Total Sample, i.e. the number of participants recruited and included in the survey.

The next layer indicates the number of participants who were asked a particular question. This number varies across the survey, as some participants skip follow-up questions depending on their previous answers.

The third layer reflects the number of responses to the question participants who responded to the question — not everyone who was asked chose to provide a response.

Finally, when analyzing each question, respondents are categorized according to their answers. Some responses may be irrelevant, unintelligible, or outside the analytical frame and are therefore excluded. The final N captures this last stage of filtering by indicating the number of categorized responses.

