DECODIS

Digital Personas: Methodology Paper

Decodis | January 2026

Commented [DC1]: @Eve Andeka We need to go back and find some of the original documents for Digital Personas. We need an objective and what we were trying to do. Look at the grant document

Commented [DC2R1]: Then we need an executive summary. But we need objectives, methodology (use from DPP), then summarize findings from each of the money slides.

Commented [DC3]: Here is the proposal document mentioned below: Investment Document_INV-057958_06-21-2023 as of 28 September 2023.docx There's a grant description that should be good to draw from. Remember that you're trying to find the problem we're trying to solve and how we are aiming to solve it and what for. That's the focus on this section.

Objectives

The Digital Personas study set out to use Large-Scale Qualitative Data and Insights technology to understand the barriers and benefits to digital connectivity for women and girls in developing countries, and how these intersect with lives and livelihoods outcomes. The project addressed a fundamental gap in the research landscape: the absence of qualitative data at scale that could generate meaningful personas and insights for vulnerable women and girls.

Specifically, the study sought to:

- Conduct large-scale qualitative surveys across Kenya, Nigeria, and Senegal, drawing samples from existing Pathways vulnerability segments to ensure targeted and representative data collection
- Capture nuanced, ground-truth insights on women's barriers to digital connectivity — including affordability, gender norms, safety, and device ownership — through Interactive Voice Recording (IVR) and Natural Language Processing (NLP)
- Generate diagnostic insights that go beyond reporting, producing findings deep and broad enough to support teams in designing targeted interventions, identifying partnerships, and meeting women where they already are in their digital lives

Commented [DC4]: Add in a bullet here about learning objectives and then build in a section. See here Decodis GE-DC Insights and Diagnostic Tool Technical Details of Methodology Phase 1 output 21 March 2023.docx

Methodology

Who We Were Trying to Reach

The study engaged over 3,500 women across seven populations in Kenya, Nigeria, and Senegal, covering both urban and rural settings. The study focused specifically on women who already had some access to a phone, with the goal of understanding the texture of their digital lives rather than measuring basic access. Within those regions, fieldwork covered Nairobi, Kwale, and Nakuru in Kenya; Kano, Kaduna, and Lagos in Nigeria; and Dakar, Diourbel, Kolda, Saint Louis, and Thiès in Senegal. Sampling was stratified across urban and rural settings and by vulnerability level using the Pathways segmentation framework.

How We Recruited

Field officers were deployed across all regions to recruit participants in person. They explained the purpose of the study, confirmed eligibility, and obtained consent. Respondents selected a preferred date and time for survey participation during recruitment to ensure availability and reduce non-response. Incentives were paid after each module was completed. Response rates ranged from 55% in rural Kenya to 94% in urban Kenya, with all other regions falling between 75% and 89%.

How We Ran the Survey

A core design requirement was that women did not need to be literate to participate. All questions were pre-recorded by local voice actors — not AI-generated — speaking in a friendly, conversational tone in the respondent's own language: Swahili, Hausa, Yoruba, or Wolof. Respondents only needed to know how to answer a phone, speak into it, and press keypads.

The survey was organised into three modules. Module 1 covered demographics and device journey. Module 2 covered usage and learning. Module 3 explored sensitive topics through audio skits on ownership, safety, and wellbeing. Each module combined two response types: keypad responses to capture quantitative prevalence, and open-ended voice responses to capture contextual nuance.

In Kenya and Nigeria, surveys were administered via IVR, where respondents received an automated phone call and could pause, resume, or complete the survey at a time of their choosing. Automated calls were scheduled based on preferred hours indicated during recruitment. In Senegal, where the cost of phone calls was prohibitive, respondents instead received a survey link via WhatsApp or SMS, enabling them to record voice responses at their convenience with the same flexibility. Prior to each module, field officers sent reminder calls to confirm availability, reinforce consent, and prepare respondents. Where needed, in-person follow-ups were conducted to resolve access, technical, or comprehension challenges.

Qualitative Depth: Audio Skits

A distinctive methodological feature of this study was the use of audio skits in Module 3. Audio skits are short, pre-recorded stories featuring fictional characters navigating everyday situations. Rather than asking a respondent directly about her own experiences, the study played her a story about someone else in a similar situation and asked what she thought. Because respondents were reacting to a character rather than speaking about themselves, they were more likely to share honest perspectives — including unstated emotions and views that direct questioning consistently misses.

Each study incorporated between four and six skits covering topics including phone ownership, safety, and control. The effect on response quality was measurable: responses to skit-based questions were on average three times longer than responses to other question types. Respondents were only ever exposed to audio versions of the skits via phone call; video versions exist solely for illustration and translation purposes.

Commented [DC5]: @Eve Andeka We need to add in a good amount of the details about the sample, with maps in the annex.

Commented [DC6R5]: See details about Nigeria, although take out the reference to Yux: Nigeria map 25 Aug 2025.pptx

Commented [DC7R5]: Some details about Senegal are here: Senegal DP Sampling Approach .docx

Commented [DC8R5]: Can you dig around and find Kenya?

Commented [DC9]: @Eve Andeka You can add some of this to the paper: Decodis GE-DC Insights and Diagnostic Tool Technical Details of Methodology Phase 1 output 21 March 2023.docx

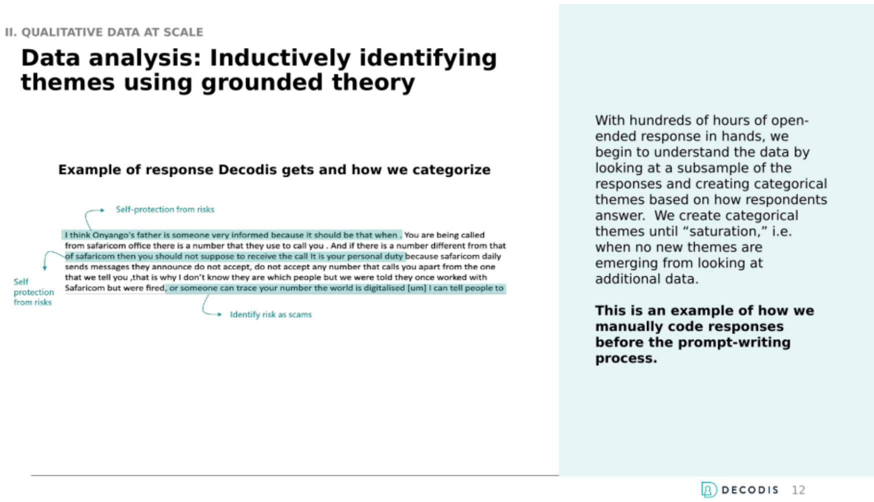
How We Processed the Audio

The survey generated over 350,000 audio clips across the five regions where audio data was processed. The proportion of unintelligible clips was low — 1–2% in Kenya and 5–9% in Nigeria. Except for Swahili, for which Decodis used its own automated translation model, all audio responses were manually translated into English by local translators, who flagged any clip marked as poor quality, unclear, or noise-affected.

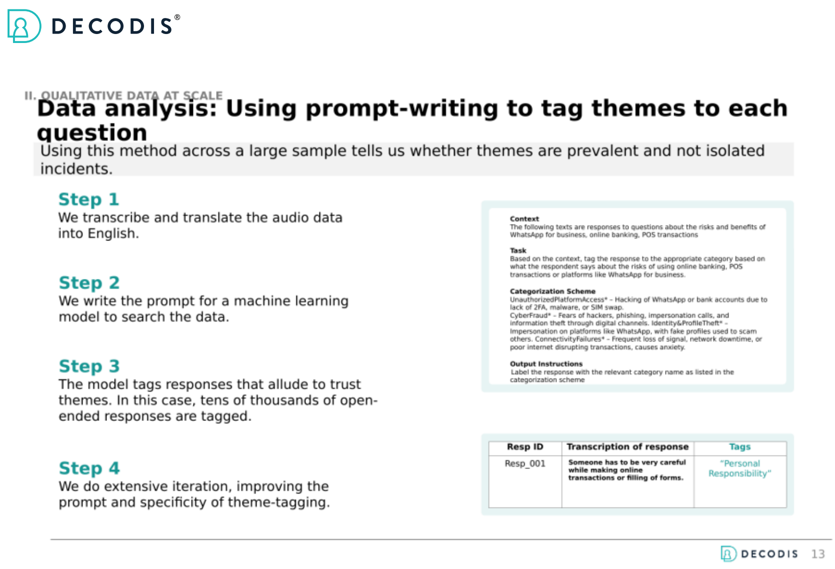
How We Worked on Meaning-Making

Commented [DC10]: @Eve Andeka I think there is a graphic from the DPP decks that you can add in about how we do the coding.

Given the scale of qualitative data collected, the study used an AI-assisted coding approach to categorise open-ended voice responses. The process followed three steps. First, 30 responses per question were randomly selected and manually coded by analysts using an inductive approach until no new thematic categories emerged — a point known as saturation.



Second, a structured prompt was written instructing the AI model to apply the manually derived categorisation scheme to the full sample, specifying the analytical context, the task, and precise output formatting. This allowed tens of thousands of open-ended responses to be tagged consistently and at scale. Third, codes were iteratively refined through extensive testing to ensure robust fit with the data. The text analytics platform underpinning this process integrates multiple large language models alongside Retrieval-Augmented Generation (RAG) methodology to maximise accuracy on nuanced notions, and uses probability-based techniques to further minimise bias — with human-in-the-loop oversight at each stage to ensure robustness and guard against hallucinations.



In addition to text-based coding, the study applied sociolinguistic analysis to the audio recordings themselves. Using Praat, a speech signal extraction tool, acoustic measures — principally pitch and intensity — were drawn from respondents' voices across each section of their recorded responses. These were normalised using Lobanov Z-score normalisation, calibrated to each respondent's own median voice across all their responses rather than compared against other speakers. This controls for known confounds including gender (women generally speak at higher pitch than men) and cross-linguistic variation in intonation and volume. The resulting measures were then interpreted through a framework of two emotional fundamentals drawn from the academic literature: activation (the intensity or strength of an emotion — the difference between, say, feeling content and feeling elated) and valence (its direction — the difference between feeling happy and feeling angry). Compiled indicators across pitch, intensity, and word count were used to operationalise emotional response at the level of individual questions, producing several of the study's most significant findings — including the distinction between women who expressed shame about asking for help and those who pushed through it, and the gap between women who stated they would use digital protective strategies and those who sounded genuinely confident doing so.

Finally, the platform brings text and voice analysis together as two interwoven data streams. Proprietary software combines the thematic codes derived from transcripts with acoustic weightings drawn from the same responses, producing composite indicators that can be mapped against demographic and contextual attributes — and used to segment populations by emotional state as well as by stated behaviour.

In

Commented [DC11]: You can include more from here: Technical doc for NY Times.docx

How speech signals are used

To capture not just *what* respondents said but *how* they said it, we extract three acoustic signals from each recorded response: pitch modulation (the rise and fall of the voice), intensity (loudness), and word count (length of response). These signals are combined into a single composite measure.

A key design choice is personal baselining. Rather than comparing respondents against each other — which would conflate natural variation in language, dialect, and speaking style — each person's score is calculated relative to her own voice across other responses. This normalises for individual differences and isolates genuine shifts in emotional engagement.

The composite score reflects the degree to which a response deviates above or below a participant's own baseline. We characterise upward deviation (higher pitch variation, greater loudness, longer responses) as **enthusiasm**, and downward deviation as **indifference**, following the framework established by Freeman (2019) on prosodic features of conversational stance.

This approach allows us to say, with acoustic evidence, where respondents were genuinely engaged versus where they spoke in a more neutral or detached way — adding a layer of behavioural data that self-reported responses alone cannot provide.

Summary of Key Findings

Commented [DC12]: We can update these with the slick slide deck

Cross-Cutting Finding: Access to Smartphones Is Not the Biggest Barrier

The most consequential cross-cutting finding is that phone access is a substantially less significant barrier than commonly assumed. In every region except rural Northern Nigeria, at least half of women already have access to a smartphone (Lagos and Urban Senegal: 75%; Urban Kenya and Rural Kenya: 66%; Urban N. Nigeria and Rural Senegal: 50%; Rural N. Nigeria: 41%).
Moreover, between one-third and one-half of women across all three countries described their phones as more basic than photos of those phones confirmed — almost always underreporting the sophistication of the device they actually used. This systematic pattern suggests that surveys relying on self-reported phone type significantly underestimate women's existing access to digital tools, and points to a deeper conclusion: the barriers that most meaningfully limit women's digital engagement are not found in device access, but in the discontinuities, social pressures, and confidence gaps that shape how — and how much — they are able to use the devices they already have.

Drivers: Why Women Want to Use Digital

Women across all regions expressed strong motivation to engage with digital technology, driven by three consistent forces:

1. Earning Money

For women running businesses, phones are the primary channel through which customers are reached, transactions are made, and new ideas are learned. Across urban and rural Kenya, approximately half of women in business say their phone helps them inform customers about goods and services. In Lagos, 58% of businesswomen cite promoting goods and services as a key benefit; in Senegal, ease of transactions is the top benefit (Urban Senegal: 51%, Rural Senegal: 47%). Women in casual or wage work use phones primarily to be contacted for jobs and to coordinate transactions. The boundary between "personal phone" and "work tool" effectively disappears for women who sell online, making data interruptions and forced phone-sharing especially costly.

2. Expanding Knowledge

Across all regions, between 67% and 88% of women report using their phones to access information about news, decisions, family matters, and learning. WhatsApp and social media are among the most used platforms (Kenya: 75–78% WhatsApp; Lagos: 85%; Urban Senegal: 94%). Where women go online to learn, they describe videos and other women's comments as their most trusted sources, often watching content repeatedly until they understand it.

3. Access to Health Information

Health-seeking is a particularly strong driver in Kenya, where 84–89% of women say they would look for health information online, compared with roughly two-thirds in Nigeria and urban Senegal, and 43% in rural Senegal. The type of information sought differs by country: in Kenya, women seek preventative knowledge (nutrition, disease understanding); in Nigeria and Senegal, women focus on diagnosing current symptoms and identifying medication. Across all regions, women consistently describe feeling empowered by what they find online — valuing the privacy, convenience, and ability to arrive at appointments better prepared.

Barriers: What Slows Women Down

Barrier 1: Discontinuous Use

Three forms of discontinuity interrupt women's digital access:

- **Data bundle management:** Women turn data on and off to stretch limited budgets, balancing data costs against food, schooling, and other expenses. Patterns differ sharply by region: in Urban Kenya, women spend a median of $1.12/week and purchase data up to 7 times a week, constantly interrupting online activities; in Senegal, women buy a larger bundle once a month ($0.83) and ration it, disrupting learning and ongoing uses; in Northern Nigeria, women spend as little as $0.19–$0.23/week but top up 1–2 times weekly, creating persistent discontinuity.

- **Charging outside the home:** Where electricity is unstable, women rely on fee-based charging businesses and face extended periods without their devices. In rural Northern Nigeria, 46% of women must charge outside the home, leaving them without their phone for 19 hours per week — the equivalent of more than two working days. Rural Senegal (42%, 16h/week) and Urban Northern Nigeria (30%, 18h/week) are similarly affected.
- **Giving up the phone:** Coercion into surrendering the phone to others represents a third form of discontinuity, most pronounced in Nigeria. When presented with a scenario about a husband taking his wife's phone after his broke, 60% of Lagos women and 54% of Rural Northern Nigeria women described this as acceptable — often citing the husband's work needs while downplaying their own income-earning use of the device. In Kenya, views were more balanced, with 34–42% considering it unacceptable.

Barrier 2: Self-Imposed Restriction

Women also limit their own digital engagement in response to two distinct pressures:

- **Exposure to disturbing content:** Across Kenya and Nigeria, 49–70% of women say they would withdraw from social media or messaging apps after encountering violent or offensive content. The rate is highest in rural Northern Nigeria (60–70%), with over 50% in Kenya. Women's primary protective strategy is platform exit rather than content filtering.
- **Social censure and community norms:** In Senegal, community disapproval of women's phone use is a more pressing concern than online harassment. Over 90% of women expressed that they should only display moral behaviour online to avoid community backlash. Between 10–15% described leaving WhatsApp groups, deleting accounts, or stopping digital engagement entirely as their primary privacy protection strategy.

Barrier 3: Confidence and Skills

Between 70% and 80% of women in Kenya and Nigeria report feeling ashamed when asking for help with their phones. Sociolinguistic analysis reveals that the experience of shame is not uniform across regions:

- Women in Urban Kenya are most likely to withdraw from help-seeking entirely after experiencing shame — the strongest "discouraged" response pattern across all regions studied.

- Women in Lagos express comparable levels of shame but show greater resilience in pushing through it, demonstrating the most "determined" response pattern and the least likelihood of withdrawal.
- Women in Senegal show a different dynamic, tending to seek help within their households (which are considerably larger on average), resulting in a less pronounced shame response overall.

A parallel confidence gap exists around online safety. Acoustic analysis found that women in Rural Kenya and Rural Northern Nigeria who described using digital protective strategies — such as blocking contacts or using passwords — sounded statistically significantly more doubtful than confident when doing so, suggesting that stated behaviours may not reflect genuine capacity to act.

DECODIS®

Annex: Sample Distribution by Country, Region, and Segment

The study engaged a total of over 3,500 women across Kenya, Nigeria, and Senegal. Sampling was stratified by vulnerability segment using the Pathways segmentation framework, with geographic distribution emerging from where segments were naturally found rather than predetermined regional quotas. The tables below show the final distribution of respondents across segments and regions for each country.

Kenya — 1,264 respondents across Nairobi, Nakuru, and Kwale

Table A1: Kenya sample distribution by segment and region

Segment	Nairobi	Nakuru	Kwale	Total
UF2-K	✓	✓	✓	—
UF4-K	✓	✓	✓	—
UM1-K	✓	—	✓	—
UM4-K	✓	✓	✓	—
RF2-K	✓	✓	✓	—
RF4-K	—	✓	—	—
RM2-K	—	✓	✓	—
RM3-K	—	✓	✓	—
RM4-K	—	—	✓	—
Total	**169**	**568**	**527**	**1,264**

Segment codes: U = urban, R = rural; F = female-headed household, M = male-headed household; number denotes vulnerability level within the Pathways framework.

Nigeria — 1,416 respondents across Kano, Kaduna, and Lagos

Table A2: Nigeria sample distribution by segment and region

Segment	Kano	Kaduna	Lagos	Total
U1_NN	✓	—	—	—
U2_NN	✓	✓	—	—
U3_NN	✓	✓	—	—
U4_NN	✓	✓	—	—
R2_NN	✓	✓	—	—
R3.1_NN	✓	✓	—	—
R3.2_NN	✓	✓	—	—
UF1_L	—	—	✓	—
UF3.1_L	—	—	✓	—
UM1.2_L	—	—	✓	—
UM3_L	—	—	✓	—
UM4_L	—	—	✓	—
Total	**474**	**488**	**454**	**1,416**

Segment codes: U = urban, R = rural; NN = Northern Nigeria, L = Lagos; number denotes vulnerability level. Lagos segments are distinct from Northern Nigeria segments, reflecting the separate Pathways segmentation framework applied in each context. Region-level totals are shown; segment-level counts within each city were not disaggregated in the source data.

Senegal — 1,833 respondents across Dakar, Diourbel, Kolda, Saint Louis, and Thiès

Table A3: Senegal sample distribution by segment and region

Segment	Dakar	Diourbel	Kolda	Saint Louis	Thiès	Total
U1-S (least vulnerable urban)	162	—	—	—	25	187
U2.1-S (moderately vulnerable urban)	183	—	—	—	60	243
U2.2-S (moderately vulnerable urban)	257	—	—	—	15	272
U3.1-S (most vulnerable urban)	88	—	—	—	2	90
R2-S (least vulnerable rural)	—	253	—	50	222	525
R3.1-S (more vulnerable rural)	—	—	91	—	—	91
R3.2-S (more vulnerable rural)	—	268	—	19	1	288
R4-S (most vulnerable rural)	—	—	130	7	—	137
Total	**690**	**521**	**221**	**76**	**325**	**1,833**

Segment codes: U = urban, R = rural; number and letter suffix denote vulnerability level within the Pathways segmentation framework applied in Senegal. Urban segments are concentrated in Dakar and Thiès; rural segments are distributed across Diourbel, Kolda, and Saint Louis, reflecting where each segment was naturally prevalent.

Note: Across all three countries, geographic distribution was determined by where Pathways vulnerability segments were naturally found rather than by predetermined regional quotas. This approach preserved segment integrity and ensured that the characteristics of each segment were not distorted by forcing respondents into fixed geographic allocations.