

Sahel Climate Voice: An AI-Driven Multilingual Mobile System for Climate Information Translation and Speech Delivery in Bambara, Hausa, and Mooré

ELH MAMAN GARBA Ibrahim¹, Belko Abdoul Aziz DIALLO², Stephen E. Moore³, Gustavo Juan Nagy Breitenstein⁴, Adama Boua DOUMBIA¹, KAFANDO Halidou¹

¹ Graduate Study Programme on Climate Change and Education, West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL), School of Agriculture and Environmental Sciences, University of The Gambia, Banjul, The Gambia (ibrahim.e@edu.wascal.org)

² WASCAL Competence Centre (CoC), Avenue Mouammar Kadhafi, Ouagadougou, Burkina Faso (diallo.b@wascal.org)

³ University of Cape Coast, Ghana (Stephen.moore@ucc.edu.gh)

⁴ Universidad de la República (University of the Republic), Uruguay (gnagy@fcien.edu.uy)

Corresponding Author: ELH MAMAN GARBA Ibrahim

Abstract:

Background: The usefulness of climate science information is substantially diminished when it reaches end users in languages or formats outside their familiarity. This paper presents Sahel Climate Voice, an Android mobile application combining domain-specific lexical resources, neural machine translation, text-to-speech synthesis, cloud-based inference, and mobile deployment to serve speakers of Bambara, Hausa, and Mooré. **Materials and Methods:** The lexical foundation of the system is BUMANI_CIC (Climate Information Corpus), whose initial release consolidated 1,002 curated climate-change concepts with 517 weather and climate glossary terms into 1,519 aligned entries. Throughout application development, each bilingual dataset was expanded to 1,633 English–Bambara, 1,633 English–Hausa, and 1,633 English–Mooré pairs, yielding 4,899 bilingual records centered on a shared core of 1,608 case-normalized English headwords. Translation quality was ensured through functional equivalence, back-translation, and validation by expert native speakers. The AI pipeline employs NLLB-200 distilled 600M for multilingual text translation and Meta Massively Multilingual Speech (MMS)/VITS models for audio synthesis. **Results:** A FastAPI backend exposes translation and speech endpoints; the Android client caches BUMANI_CIC terminology on-device and communicates with cloud services over HTTPS. The completed application was release-signed, packaged as a 6.67 MB APK, installed on a physical Android device, and validated for translation accuracy and audio output. **Conclusion:** This work demonstrates an end-to-end workflow progressing from climate-domain corpus construction to a fully operational multilingual speech-capable mobile application. The study explicitly distinguishes technical integration validation from linguistic accuracy assessment and farmer-focused usability evaluation, both of which are designated as future research directions.

Key Word: climate information services; low-resource NLP; Bambara; Hausa; Mooré; neural machine translation; text-to-speech.

Date of Submission: 10-10-2026

Date of Acceptance: 22-10-2026

I. Introduction

Climate information services (CIS) play an important role in enabling agricultural adaptation and managing climate-related risks; however, their utility hinges on whether target populations can effectively access, comprehend, and respond to the information they convey (Vaughan et al., 2019). Systematic reviews examining African smallholder farming contexts reveal that CIS uptake is shaped by a combination of farmer-level factors, extension services, institutional frameworks, and the channels through which information is communicated (Nyoni et al., 2024). Additional reviews have consistently highlighted language barriers, the use of technical jargon, literacy limitations, and inadequate digital infrastructure as enduring obstacles to effective climate-service engagement across sub-Saharan Africa (Khatibu & Ngowi, 2025; Muita et al., 2021). Taken together, these findings indicate that linguistic accessibility must be treated as a fundamental system-design consideration rather than a secondary translation task.

African languages, in parallel, remain severely underrepresented within natural language processing research (Nekoto et al., 2020; Ranathunga et al., 2023). Research on participatory machine translation for African languages has shown that the difficulties involved go beyond simple data shortages and are deeply tied to the institutional and social contexts in which linguistic resources are developed (Nekoto et al., 2020). Recent advances in massively multilingual neural machine translation have substantially broadened language coverage: the No Language Left Behind (NLLB) initiative achieved translation across 200 languages using a single unified model architecture and specifically targeted the quality disparities that affect low-resource languages (NLLB Team et al., 2024). Progress in multilingual speech has followed a similar trajectory; the Massively Multilingual Speech (MMS) project delivered speech synthesis models covering more than 1,000 languages (Pratap et al., 2023).

In this context, the present study introduces Sahel Climate Voice, a mobile application that translates English climate information into written and spoken Bambara, Hausa, and Mooré. The system is built upon BUMANI_CIC (Climate Information Corpus), an acronym in which BU represents Burkina Faso and Mooré, MA represents Mali and Bambara, and NI represents Niger and Hausa. The scope of the contribution goes beyond the creation of a corpus or an isolated model: the paper documents the entire workflow spanning climate-knowledge acquisition, multilingual terminology development, translation modeling, speech synthesis, REST API construction, cloud deployment, Android packaging, and validation on a real device.

From a software engineering standpoint, the study examines how a specialized linguistic resource, a multilingual neural translation service, language-specific text-to-speech synthesis, a REST API, and an Android application can be combined into a deployable cloud-mobile architecture capable of supporting climate communication in low-resource language settings. The evaluation presented in this paper targets system integration, endpoint functionality, release packaging, and physical-device execution; linguistic quality and farmer-centered usability are treated as independent evaluation dimensions that cannot be inferred from technical outcomes alone.

The paper makes five distinct contributions: (i) a climate-domain multilingual terminology resource covering three West African languages; (ii) a replicable translation pipeline built on NLLB; (iii) the incorporation of multilingual speech synthesis to support oral information delivery; (iv) a hybrid cloud-mobile architecture that combines locally embedded terminology with remote AI-powered inference; and (v) comprehensive technical validation carried out on a physical Android device.

II. Related Work

2.1 Climate information accessibility

Contemporary climate-service research consistently stresses that effective information delivery must be timely, comprehensible, actionable, and appropriately matched to the resources available to users (Vaughan et al., 2019; Tarchiani & Grey, 2025). A review by Vaughan et al. (2019) encompassing 59 African studies demonstrated that the effectiveness of climate information services is jointly determined by the nature of the service, the attributes of the user, and the contextual setting in which both are situated. Khatibu and Ngowi (2025) synthesized 53 studies and found that language barriers, technical vocabulary, low levels of literacy, and inadequate digital infrastructure frequently limit service uptake, whereas participatory methods and locally adapted approaches tend to yield more favorable outcomes. Appiah et al. (2025) additionally showed that social networks can enhance CIS access and use, in part because they facilitate communication in relatable local languages and allow for direct interpersonal exchange.

2.2 Low-resource African-language machine translation

Machine translation has seen major advances through multilingual transfer learning, yet the availability of parallel corpora remains deeply unequal across languages (Ranathunga et al., 2023). Masakhane established the viability of community-driven research for African-language machine translation by generating datasets and evaluation benchmarks for more than 30 languages (Nekoto et al., 2020). NLLB subsequently extended multilingual neural translation to cover 200 languages and drew attention to the central role of data scarcity in perpetuating translation disparities (NLLB Team et al., 2024). These advances underscore the value of domain-specific resources such as BUMANI_CIC, whose specialized climate terminology can enrich and supplement the general multilingual pretraining available in large-scale models.

2.3 Multilingual speech synthesis

Translation that produces only written output does not fully resolve accessibility challenges in contexts where literacy levels are low, especially when climate services must serve users with diverse degrees of literacy and varying levels of digital access (Khatibu & Ngowi, 2025). Pratap et al. (2023) leveraged self-supervised multilingual learning to broaden speech technology coverage and reported speech synthesis models for 1,107 languages. This extensive language support makes MMS a suitable tool for pairing translated climate content with spoken output in languages that have historically lacked commercial text-to-speech solutions.

III. System design and methods

3.1 Architecture and development framework

The overall system development framework encompassed climate-source identification, terminology extraction, multilingual corpus construction and validation, neural translation, speech synthesis, API development, Android client integration, cloud deployment, and physical-device testing. Figure 1 provides an overview of the complete system architecture. The framework was conceived as a cohesive end-to-end pipeline in which linguistic resources, AI models, cloud infrastructure, and mobile technology function as tightly coupled components rather than standalone, independently operating modules.

The initial stage of development centered on establishing the climate knowledge and terminology layer. Peer-reviewed scientific literature and institutional climate resources were consulted to identify domain-relevant concepts and vocabulary, which were then screened for relevance and adapted into multilingual equivalents. These entries were consolidated into the BUMANI_CIC corpus, which serves as the specialized linguistic backbone of the system through its English-Bambara, English-Hausa, and English-Mooré paired resources. This layer ensures that the application maintains domain-specific climate vocabulary while enabling effective communication across the three target languages.

The second stage corresponds to the AI translation layer. Sahel Climate Voice integrates BUMANI_CIC terminology matching with the NLLB-200 distilled 600M model to handle sentence-level translation. This hybrid approach distinguishes between recognized climate terms present in the corpus and more general sentences that require neural machine translation. English input is processed with respect to the selected target language, and the resulting translated output is then forwarded to the speech synthesis component.

A dedicated speech-synthesis layer extends the system's capabilities beyond written-only output. Translated text is converted to speech using language-specific Meta MMS/VITS text-to-speech models corresponding to Bambara, Hausa, and Mooré; the resulting audio waveform is subsequently encoded as a WAV file for playback within the mobile application. Both the translation and speech synthesis components are made available through a REST-based cloud backend built with FastAPI, which the Android client accesses securely over HTTPS.

At the user interface level, the Android application provides language selection, English climate-information text entry, display of translated output, BUMANI_CIC terminology matching results, and synthesized audio playback. Computationally demanding AI models are hosted entirely on the cloud backend, which reduces the processing and storage footprint of the mobile client and allows model updates to be managed independently of the installed application. The development cycle concluded with release packaging and physical-device testing, thus completing the pipeline from climate knowledge acquisition and multilingual resource creation through AI inference, cloud delivery, and final user-facing access.

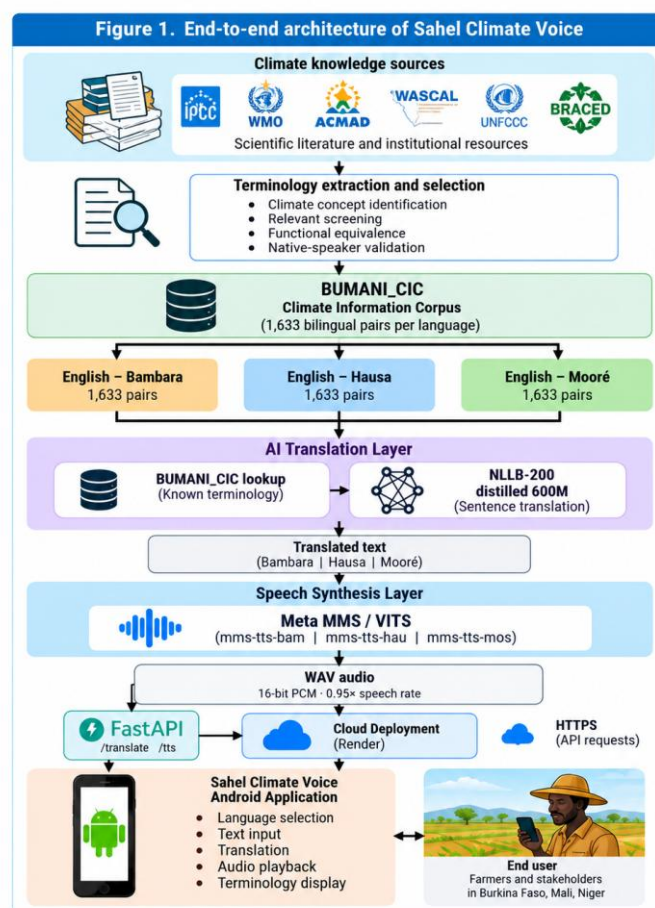


Figure 1: End-to-end architecture of the Sahel Climate Voice.

3.2 Domain resource construction: BUMANI_CIC

The foundational knowledge base was compiled through targeted desk research. It consisted of 650 scientific documents, comprising 481 peer-reviewed papers sourced from ScienceDirect and 169 papers retrieved via Google Scholar. Additional terminology and conceptual content were incorporated from climate information resources developed by WASCAL, the Intergovernmental Panel on Climate Change (IPCC), the African Centre of Meteorological Applications for Development (ACMAD), the World Meteorological Organization (WMO), the United Nations Framework Convention on Climate Change (UNFCCC), and the BRACED programme.

The initial corpus integrated an English Climate Change Concepts Corpus comprising 1,002 concepts developed specifically for this research with an additional 517 weather and climate glossary entries. The resulting multilingual working corpus contained 1,519 conceptual entries covering meteorology, adaptation, mitigation, climate policy, risk assessment, and related thematic areas.

Translation was carried out using functional equivalence rather than literal word-for-word rendering (Repeke & Dorer, 2021). In cases where direct lexical equivalents were absent in the target language, strategies such as borrowing, transliteration, compounding, explanatory paraphrase, and contextual localization were employed. Semantic integrity was verified through back-translation, while native-speaker reviewers evaluated each entry for accuracy, clarity, and cultural appropriateness. This approach was of particular importance for technical concepts that often require multi-word expressions for natural expression in the target languages.

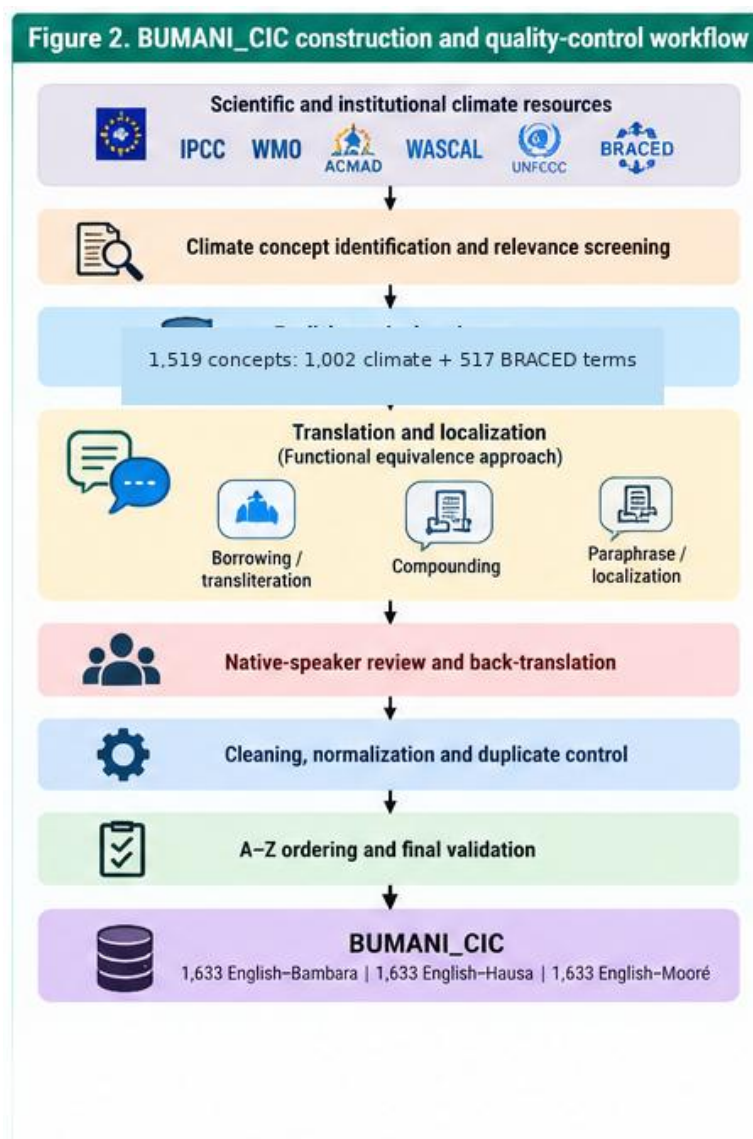


Figure 2: BUMANI_CIC corpus construction and quality-control workflow.

3.3 Application-oriented resource preparation

Further corpus preparation extended each language resource to 1,633 bilingual entry pairs. All files were sorted alphabetically by English headword, normalized for whitespace, and checked for duplicate entries. Both the Bambara and Mooré resources contain 1,633 unique case-normalized English headwords. The Hausa resource includes 1,632 unique case-normalized English headwords distributed across 1,633 rows, since the forms “Figure” and “figure” are each preserved with distinct translations. The shared cross-language headword intersection across all three language pairs consists of 1,608 entries.

Table 1: Evolution of BUMANI_CIC from the initial research resource to the application corpus.

Language pair	Initial corpus	Application corpus	Increase
English–Bambara	1,519	1,633	+114
English–Hausa	1,519	1,633	+114
English–Mooré	1,519	1,633	+114
Total bilingual records	4,557	4,899	+342

Table 2: Descriptive characteristics of the application-oriented BUMANI_CIC files.

Language pair	Rows	Unique English	Mean English tokens	Mean target tokens	Multiword target
English–Bambara	1,633	1,633	1.78	2.54	1,160
English–Hausa	1,633	1,632	1.79	2.42	1,171
English–Mooré	1,633	1,633	1.79	4.26	1,243

The relatively longer target-language expressions, most notably the mean of 4.26 tokens per entry in Mooré, validate the methodological choice to retain descriptive paraphrasing when single-word equivalents do not exist in the target language. Accordingly, BUMANI_CIC should be understood as a domain-specific terminology and concept resource rather than a sentence-level parallel translation corpus.

3.4 Translation service

The translation module is built on the NLLB-200 distilled 600M multilingual encoder-decoder architecture, which was selected because massively multilingual transfer learning has shown considerable promise for extending machine translation to low-resource languages (NLLB Team et al., 2024). English text input is identified using the eng_Latn language code; target language identifiers are bam_Latn for Bambara, hau_Latn for Hausa, and mos_Latn for Mooré. In the earlier model-development phase, the architecture was fine-tuned separately for each target language using Hugging Face Transformers and Datasets, with tokenization, padding, truncation, and Seq2SeqTrainer-based optimization. Those historical fine-tuned checkpoints have been preserved for experimental reproducibility; however, the production service detailed in Section VII currently operates with the pretrained base checkpoint rather than the fine-tuned model weights.

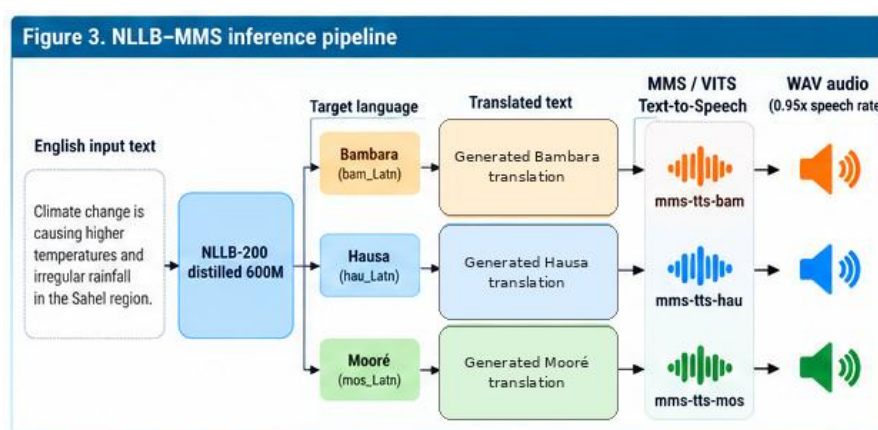


Figure 3: NLLB-MMS inference pipeline used by Sahel Climate Voice.

3.5 Speech synthesis service

Speech is synthesized using language-specific Meta MMS/VITS model checkpoints (Pratap et al., 2023), specifically facebook/mms-tts-bam, facebook/mms-tts-hau, and facebook/mms-tts-mos. The corresponding tokenizer and VITS model are used to produce an audio waveform from the translated text. This waveform is then normalized, converted to 16-bit PCM format, and returned as a WAV audio file. During iterative testing, the speech playback rate was reduced to 0.95 of the originally generated rate in order to enhance listening comfort without substantially modifying the characteristics of the synthesized voice.

3.6 REST backend and cloud architecture

The inference layer was developed as a REST API using FastAPI. The POST /translate endpoint accepts English text along with the selected target language and returns the translated output. The POST /tts endpoint accepts target-language text with the corresponding language identifier and returns WAV audio. Model objects are cached upon first load to eliminate redundant initialization overhead. The production service is hosted in the cloud and maintains secure communication with the Android application via HTTPS.

3.7 Android client implementation

Sahel Climate Voice was designed and implemented as a mobile-first Android application. While mobile and digital advisory platforms can extend the reach of climate information delivery, research evidence from African agriculture consistently shows that the practical value of such tools depends on accessibility, infrastructure quality, user literacy, local context, and user-centered design principles (Khatibu & Ngowi, 2025). The application processes climate terminology locally by matching input text against the curated BUMANI_CIC resource, while complete sentences are transmitted to the cloud-based translation service. The core user workflow is deliberately streamlined: choose Bambara, Hausa, or Mooré; type English climate information; request a translation; and listen to the synthesized local-language audio output. The interface additionally displays matched BUMANI_CIC terminology, enabling users to differentiate between curated domain-specific terms and general sentence-level translation.

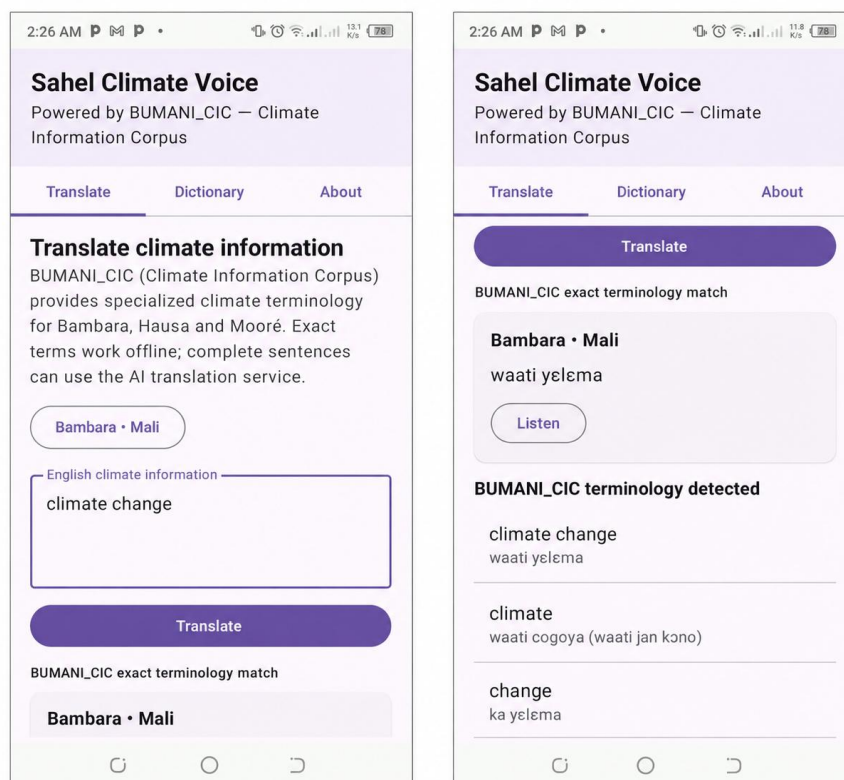


Figure 4: Real Sahel Climate Voice Android interface showing Bambara selection, English climate-information input, offline BUMANI_CIC terminology matching, translated output, and the Listen function.

IV. Technical Validation and Results

4.1 Historical model-development evidence

Table 3: Historical NLLB-200 fine-tuning dynamics.

Language	Epochs	Initial train loss	Final train loss	Initial validation loss	Final validation loss
Bambara	3	1.2133	0.0725	0.180706	0.118063
Hausa	10	0.3999	0.1271	0.302493	0.217350
Mooré	8	0.3999	0.1322	0.302493	0.217624

Bambara fine-tuning converged quickly within three epochs, whereas Hausa and Mooré exhibited extended learning trajectories. These loss figures reflect historical results from the model development phase and are not intended to represent translation accuracy on the current 1,633-entry application corpus. Training and validation losses describe the behavior of the optimization process and should not be taken as indicators of semantic quality, speech naturalness, system usability, or practical field performance.

4.2 Implemented software stack

Table 4: Sahel Climate Voice implementation stack.

Component	Technology	Role
Domain resource	BUMANI_CIC	Climate terminology
Translation	NLLB-200 distilled 600M	English to local language generation
Speech synthesis	Meta MMS/VITS	Local-language audio
Backend	Python + FastAPI	Inference API
Mobile client	Android	End-user interface
Communication	HTTPS	Mobile to cloud transport
Audio playback	Android MediaPlayer	WAV playback
Cloud deployment	Render	Production inference hosting
Version control	GitHub	Source/version management
Distribution	Signed APK	Android installation

4.3 End-to-end deployment validation

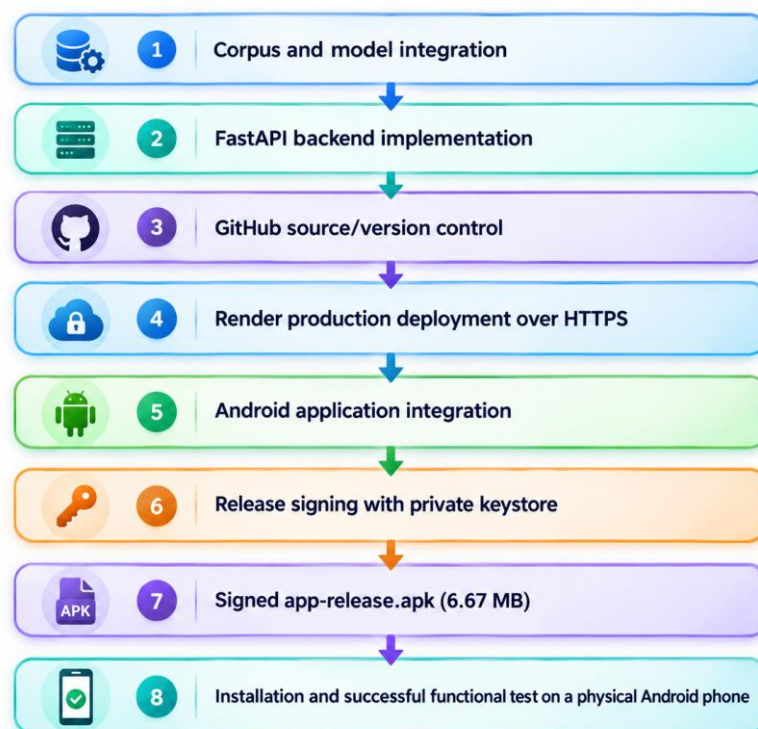


Figure 5: Deployment pathway from development to physical-device validation.

The finalized Android application was signed for release using a private keystore and compiled into app-release.apk. The signed APK weighed approximately 6.67 MB, a compact size attributable to the fact that the computationally heavy NLLB and MMS models are hosted on the cloud backend rather than bundled within the mobile package. The APK was transferred to a physical Android handset, installed without error, and tested independently of Android Studio. Both translation and audio playback functioned correctly on the physical device, confirming end-to-end technical connectivity spanning Android input, cloud-based inference, and the return of text and audio output to the device.

Table 5: Technical validation matrix.

Validation item	Environment	Outcome
Application launch	Android development environment	Passed
BUMANI CIC terminology retrieval	Android application	Passed
Translation endpoint	Cloud API	Passed
Bambara speech generation	MMS backend	Passed
Hausa speech generation	MMS backend	Passed
Mooré speech generation	MMS backend	Passed
WAV playback	Android	Passed
Signed release APK generation	Release build	Passed
APK installation	Physical Android phone	Passed
Translation and speech on a physical device	Phone–cloud–phone	Passed

V. Discussion

The primary contribution of Sahel Climate Voice lies in the integration of components that are typically studied and evaluated in isolation. This directly addresses calls for climate-service value chains that link information production, translation, communication, and practical use rather than treating content dissemination as a stand-alone technical concern (Ofogebu & New, 2022). In this system, BUMANI_CIC provides domain-specific linguistic grounding; NLLB drives multilingual generative translation; MMS converts translated text into audible speech; FastAPI makes the AI models available as network-accessible services; and Android delivers the final product to the end user. This integrated architecture transforms a research resource into an operational tool rather than leaving the corpus as an inert static dataset.

The speech synthesis layer holds particular significance for climate communication. Recent reviews of CIS continue to identify low literacy, technical language, and communication channel constraints as significant

barriers to service adoption (Ofoegbu & New, 2022). Although mobile delivery can expand the reach of climate information, prior reviews also warn that mobile-based climate services remain contingent on language suitability, literacy levels, reliability, user trust, and socioeconomic conditions (Ofoegbu & New, 2022). Sahel Climate Voice therefore targets one specific dimension of accessibility—language and delivery modality—without presupposing that technological solutions alone are sufficient to overcome the broader institutional and social factors that govern climate-service uptake.

Controlled terminology management represents another significant contribution of the work. Recent research on machine translation demonstrates that terminology-aware approaches require careful calibration between adherence to prescribed terms and overall sentence fluency, underscoring the importance of explicit domain-specific terminological resources (Abela et al., 2026). General-purpose multilingual models can generate semantically plausible output while still rendering specialized concepts inconsistently across languages (Grubišić & Korenčić, 2025). By integrating BUMANI_CIC directly into the Android application, the system maintains an explicit domain terminology layer that operates alongside neural sentence-level translation. This hybrid design lays the groundwork for future development, including terminology-constrained decoding, retrieval-augmented translation pipelines, and rigorous evaluation of climate-specific lexical fidelity.

Successful deployment and validation on a real physical device confirms the engineering feasibility of the system but must not be construed as evidence of translation accuracy, speech naturalness, usability, or behavioral impact. Addressing those dimensions requires purpose-designed linguistic and user-centered evaluation studies. This distinction is of particular importance in the climate communication domain, where a mistranslation that nonetheless sounds fluent could substantially distort the interpretation of a climate warning or advisory message.

VI. Threats To Validity, Limitations, And Future Evaluation

The evidence reported in this study is primarily technical and architectural in nature. Construct validity is restricted because successful API calls, APK installation, and audio playback provide no direct measure of semantic translation quality, speech naturalness, comprehension, or usability. External validity is similarly constrained, as the physical-device validation was conducted as an implementation test rather than a controlled field evaluation involving representative smallholder farmers in Burkina Faso, Mali, and Niger. These limitations confine the study's claims to engineering feasibility and provide the rationale for the evaluation framework outlined below.

BUMANI_CIC is primarily a terminology-oriented resource rather than a large-scale sentence-level parallel corpus. Research on low-resource machine translation has consistently highlighted the necessity of curated parallel data, community participation, domain-specific evaluation, and systematic error analysis (Ranathunga et al., 2023). Future work should therefore develop a carefully validated sentence-level climate-information benchmark and assess the deployed translation pipeline using a combination of automatic and human evaluation approaches. BLEU can serve as a standard lexical baseline (Papineni et al., 2002), chrF/chrF++ can quantify character-level overlap that is particularly informative under conditions of morphological and orthographic variability (Popović, 2015), and COMET or AfriCOMET can offer learned semantic evaluation where appropriate language support is available (Rei et al., 2020; Wang et al., 2024). Automatic metrics should be supplemented by blinded native-speaker assessments covering adequacy, terminology precision, fluency, and cultural appropriateness.

Speech quality evaluation should independently assess intelligibility, the pronunciation of climate-specific terms, overall naturalness, and speaking rate. A Mean Opinion Score protocol can be administered alongside task-based comprehension questions. In addition, field studies conducted with target users in Burkina Faso, Mali, and Niger should investigate whether participants are able to understand climate messages, navigate the application interface, correctly interpret synthesized audio, and act on the information in realistic agricultural decision-making contexts. A structured usability assessment instrument such as the System Usability Scale can provide useful complementary data to qualitative interviews, though it should not substitute for contextually grounded feedback from the farmers the system is intended to serve.

The current full-sentence translation and speech pipeline requires internet connectivity to access the cloud backend. Future engineering efforts should investigate model compression, weight quantization, edge-based inference, or partial offline functionality while maintaining acceptable levels of translation and speech quality. Wider distribution via Google Play may be pursued once the necessary developer account setup and release certification procedures have been finalized.

VII. Reproducibility and System Configuration

The production translation service operates with the pretrained base checkpoint facebook/nllb-200-distilled-600M rather than the earlier fine-tuned model checkpoints. The source language identifier is `eng_Latn`, and the target language identifiers are `bam_Latn` for Bambara, `hau_Latn` for Hausa, and `mos_Latn` for Mooré.

Speech synthesis is performed using facebook/mms-tts-bam, facebook/mms-tts-hau, and facebook/mms-tts-mos. The backend technology stack consists of Python, FastAPI, Transformers, PyTorch, SentencePiece, SciPy, and Librosa. Translation and text-to-speech functionalities are published as separate REST endpoints. The Android client communicates with the backend service over HTTPS and plays the returned WAV audio on-device. Security credentials, signing passwords, and private keystores have been excluded from the reproducibility package.

VIII. Conclusion

Sahel Climate Voice illustrates a complete development pathway from domain-specific climate-language resource construction to a fully operational multilingual mobile application. BUMANI_CIC grew from its original 1,519-concept form into three application-ready bilingual datasets of 1,633 pairs each. This resource was combined with NLLB-200 multilingual translation, Meta MMS/VITS speech synthesis, a FastAPI-powered cloud backend, and an Android user interface. The release-signed application was successfully installed and functionally validated on a physical Android device.

The contribution thus goes beyond corpus development or isolated model experimentation. It shows how climate-domain terminology resources, multilingual AI capabilities, and speech synthesis technology can be assembled into a fully functional information delivery pipeline serving Bambara, Hausa, and Mooré speakers. The next critical research step is a rigorous evaluation with native speakers and intended end users, so that the demonstrated technical feasibility can be corroborated by evidence of linguistic accuracy, system usability, and effective climate information comprehension.

Data and Code Availability

BUMANI_CIC consists of English-Bambara, English-Hausa, and English-Mooré climate terminology resources used in the application. Source code for the Android application and backend is maintained in the public repository: <https://github.com/Elhadjzz/BUMANI-Climate-Voice>. A stable archival release with explicit corpus provenance and licensing metadata is recommended for long-term reproducibility.

Ethics and Responsible Deployment

The corpus-development work uses linguistic and documentary resources and does not, by itself, establish safety-critical decision authority. Climate translations and generated speech are intended as communication support and should not replace official meteorological warnings or advisories. Because translation errors can alter the interpretation of climate information, future deployment should include clear user guidance, error-reporting procedures, and traceability to authoritative information sources.

Future farmer-centered studies should use informed consent, minimize the collection of personal data, document procedures for handling translation or speech errors, and assess whether intended users correctly understand the delivered information. Evaluation protocols should also account for literacy, language preference, device access, connectivity, and local communication practices.

References

- [1]. Abela, K., Tanti, M., & Borg, C. (2026). Balancing Fluency and Adherence: Hybrid Fallback Term Injection in Low-Resource Terminology Translation. In Proceedings for the Ninth Workshop on Technologies for Machine Translation of Low Resource Languages (LoResMT 2026), 78-86. <https://doi.org/10.18653/v1/2026.loresmt-1.6>
- [2]. Al Farouq, M. H., Wassie, A. K., & Moslem, Y. (2025). Bemba speech translation: Exploring a low-resource African language. In Proceedings of the 22nd International Conference on Spoken Language Translation (IWSLT 2025). <https://doi.org/10.18653/v1/2025.iwslt-1.37>
- [3]. Alabi, J. O., Hedderich, M. A., Adelani, D. I., & Klakow, D. (2025). Charting the landscape of African NLP: Mapping progress and shaping the road ahead. arXiv. <https://doi.org/10.48550/arxiv.2505.21315>
- [4]. Alfonsi, R. M., Naidoo, M., & Gasparatos, A. (2024). Adoption and desirable characteristics of information and communication technologies for urban small-scale food producers in South Africa. *Frontiers in Sustainable Food Systems*, 8, 1332978. <https://doi.org/10.3389/fsufs.2024.1332978>
- [5]. Appiah, C. E., Quarmin, W., Osei-Amponsah, C., Okem, A. E., & Sarpong, D. B. (2025). Improving smallholder farmers' access to and utilization of climate information services in sub-Saharan Africa through social networks: A systematic review. *Climate Services*, 37, 100528. <https://doi.org/10.1016/j.cliser.2024.100528>
- [6]. Bah, A. S., Wang, Y., Abdullahi, N. M., Maleko, S. A., & Nkhoma, N. (2025). Does the use of information and communication technologies improve cereal production in Sub-Saharan Africa? A method of moments quantile regression approach. *Canadian Journal of Agricultural Economics*. <https://doi.org/10.1111/cjag.70003>
- [7]. Belay, T. D., Ahmad, I. S., Abdulmumin, I., Nakatumba-Nabende, J., Yimam, S. M., & Muhammad, S. H. (2025). Automatic speech recognition (ASR) for African low-resource languages: A systematic literature review. arXiv. <https://doi.org/10.48550/arxiv.2510.01145>
- [8]. Bhushan, S., Mishra, V. P., Rishiwal, V., Sharmila, A., & Agarwal, U. (2025). Advancing text-to-speech systems for low-resource languages: Challenges, innovations, and future directions. *IEEE Access*, 13. <https://doi.org/10.1109/access.2025.3605236>
- [9]. Buzaaba, H., Wettig, A., Adelani, D. I., & Fellbaum, C. (2025). Lughu-Llama: Adapting large language models for African languages. arXiv. <https://doi.org/10.48550/arxiv.2504.06536>

- [10]. Gouroubera, M. W., Segnon, A. C., Tonle, F. B. N., & Zougmore, R. B. (2025). Farmers' willingness to pay for weather and climate information services in sub-Saharan Africa: A systematic review with meta-analysis. *Frontiers in Climate*, 7, 1692594. <https://doi.org/10.3389/fclim.2025.1692594>
- [11]. Grubišić, I., & Korenčić, D. (2025). IRB-MT at WMT25 Terminology Translation Task: Metric-guided Multi-agent Approach. In *Proceedings of the Tenth Conference on Machine Translation*, 1302-1334. <https://doi.org/10.18653/v1/2025.wmt-1.110>
- [12]. Imam, S. H., Sani, B., Gete, D. K., Ahmed, B., Ahmad, I. S., Abdulmumin, I., Yimam, S. M., Bello, M. Y., & Muhammad, S. H. (2025). Automatic speech recognition for African low-resource languages: Challenges and future directions. In *Proceedings of the 5th Workshop on African Natural Language Processing (AfricaNLP 2025)*. <https://doi.org/10.18653/v1/2025.africanlp-1.13>
- [13]. Inuwa-Dutse, I. (2025). NaijaNLP: A survey of Nigerian low-resource languages. *arXiv*. <https://doi.org/10.48550/arxiv.2502.19784>
- [14]. Jimoh, T. A., De Wille, T., & Nikolov, N. S. (2025). Bridging gaps in natural language processing for Yoruba: A systematic review of a decade of progress and prospects. *arXiv*. <https://doi.org/10.48550/arxiv.2502.17364>
- [15]. Katumba, A., Kagumire, S., Nakatumba-Nabende, J., Quinn, J. M., & Murindanyi, S. (2025a). A curated crowdsourced dataset of Luganda and Swahili speech for text-to-speech synthesis. *Data in Brief*, 59, 111915. <https://doi.org/10.1016/j.dib.2025.111915>
- [16]. Katumba, A., Kagumire, S., Nakatumba-Nabende, J., Quinn, J. M., & Murindanyi, S. (2025b). Building text-to-speech models for low-resourced languages from crowdsourced data. *Applied AI Letters*, 6, e117. <https://doi.org/10.1002/ail.2.117>
- [17]. Khatibu, S., & Kissoka, G. (2025). Gender dynamics in climate information services: A systematic review of intersectional influences and strategies for smallholder farmers in Sub-Saharan Africa. *Frontiers in Climate*, 7, 1681671. <https://doi.org/10.3389/fclim.2025.1681671>
- [18]. Khatibu, S., & Ngowi, E. E. (2025). Effectiveness of climate information services in Sub-Saharan Africa's agricultural sector: A systematic review of what works, what doesn't work, and why. *Frontiers in Climate*, 7, <https://doi.org/10.3389/fclim.2025.1616691>
- [19]. Muita, R., Dougill, A. J., Mutemi, J., Aura, S., Graham, R., Awolala, D. O., Nkiaka, E., Hiron, L., & Opijah, F. J. (2021). Understanding the Role of User Needs and Perceptions Related to Sub-Seasonal and Seasonal Forecasts on Farmers' Decisions in Kenya: A Systematic Review. *Frontiers in Climate*, 3, <https://doi.org/10.3389/fclim.2021.580556>
- [20]. Nekoto, W., Marivate, V. N., Matsila, T., Fasubaa, T., Fagbohunge, T., Akinola, S. O., Muhammad, S. H., Kabenamualu, S. K., Osei, S., Sackey, F., Niyongabo, R. A., Macharm, R., Ogayo, P., Ahia, O., Berhe, M. M., Adeyemi, M., Mokgesi-Seling, M., Okegbemi, L., Martinus, L., et al. (2020). Participatory Research for Low-resource Machine Translation: A Case Study in African Languages. *Findings of EMNLP 2020*, 2144-2160. <https://doi.org/10.18653/v1/2020.findings-emnlp.195>
- [21]. NLLB Team, Costa-jussà, M. R., Cross, J., Çelebi, O., Elbayad, M., Heffernan, K. S., Kalbassi, E., Lam, J., Licht, D. J., Maillard, J., Wang, A., Wang, S., Wenzek, G., Youngblood, A., et al. (2024). Scaling neural machine translation to 200 languages. *Nature*, 630(8018), 841–846. <https://doi.org/10.1038/s41586-024-07335-x>
- [22]. Nyoni, R. S., Bruelle, G., Chikowo, R., & Andrieu, N. (2024). Targeting smallholder farmers for climate information services adoption in Africa: A systematic literature review. *Climate Services*, 34, 100450. <https://doi.org/10.1016/j.cliser.2024.100450>
- [23]. Ofoegbu, C., & New, M. G. (2022). Evaluating the effectiveness and efficiency of climate information communication in the African agricultural sector: A systematic analysis of climate services. *Agriculture*, 12(2), 160. <https://doi.org/10.3390/agriculture12020160>
- [24]. Ogun, S., Owodunni, A., Olatunji, T., Alese, E., Oladimeji, B., Afonja, T., Olaleye, K., Etori, N. A., & Adewumi, T. (2024). 1000 African voices: Advancing inclusive multi-speaker multi-accent speech synthesis. *Proceedings of Interspeech 2024*. <https://doi.org/10.21437/interspeech.2024-2281>
- [25]. Olatunbosun, A., Ismaila, E. O., Adeoba, M., Oluwadare, O., Arayombo, T. J., Oladunni, A., Ajiferuke, O. E., & Igwenagu, M. O. (2026). AI-enabled climate-smart agriculture for smallholder resilience in Sub-Saharan Africa: A systematic review. *Journal of Global Agriculture and Environment*, 18(1). <https://doi.org/10.56557/jogae/2026/v18i110191>
- [26]. Ouedraogo, R. S., Timilsena, B. P., Tiendrebogo, N. R., Morr, D., & Hughes, D. (2026). Scaling climate-smart agriculture through multi-channel advisory services: Economic evaluation of weather information systems in sub-Saharan Africa. *SSRN*. <https://doi.org/10.2139/ssrn.5858183>
- [27]. Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311-318. <https://doi.org/10.3115/1073083.1073135>
- [28]. Popović, M. (2015). chrF: Character n-gram F-score for automatic MT evaluation. *Proceedings of the Tenth Workshop on Statistical Machine Translation*, 392-395. <https://doi.org/10.18653/v1/W15-3049>
- [29]. Pratap, V., Tjandra, A., Bowen, S., Tomasello, P., Arun, B., Kundu, S., Elkahky, A. M., Ni, Z., Vyas, A., Fazel-Zarandi, M., Baevski, A., Adi, Y., Xiaohui, Z., Hsu, W.-N., Conneau, A., & Auli, M. (2023). Scaling Speech Technology to 1,000+ Languages. *arXiv*. <https://doi.org/10.48550/arXiv.2305.13516>
- [30]. Qorbani, A., Ramezani, R., Baraani, A., & Kazemi, A. (2025). Multilingual neural machine translation for low-resource languages by twinning important nodes. *Neurocomputing*, 623, 129890. <https://doi.org/10.1016/j.neucom.2025.129890>
- [31]. Ranathunga, S., Lee, E.-S. A., Prifti Skenduli, M., Shekhar, R., Alam, M., & Kaur, R. (2023). Neural Machine Translation for Low-resource Languages: A Survey. *ACM Computing Surveys*, 55(11), 1-37. <https://doi.org/10.1145/3567592>
- [32]. Rei, R., Stewart, C., Farinha, A. C., & Lavie, A. (2020). COMET: A Neural Framework for MT Evaluation. *Proceedings of EMNLP 2020*, 2685-2702. <https://doi.org/10.18653/v1/2020.emnlp-main.213>
- [33]. Repke, L., & Dorer, B. (2021). Translate Wisely! An Evaluation of Close and Adaptive Translation Procedures in an Experiment Involving Questionnaire Translation. *International Journal of Sociology*, 51(2), 135-162. <https://doi.org/10.1080/00207659.2020.1856541>
- [34]. Singh, M., Ouedraogo, M., Jimenez, D., Talsma, T., Ouedraogo, A., Kagabo, D. M., Ramirez, J., & Laderach, P. (2025). Mapping the climate and agronomic digital advisory services landscape in West and Central Africa. *PLOS ONE*, 20(1), e0338010. <https://doi.org/10.1371/journal.pone.0338010>
- [35]. Solaja, S. O., Bamiro, O. M., Ayeni, M. D., Awe, T. E., Faronbi, O. A., & Okonta, O. W. (2025). Mobile agricultural information and climate change adaptation strategies: Insight from smallholder farmers in Kwara, Nigeria. *Journal of the Nigerian Institution of Production Engineers*, 7(4). <https://doi.org/10.37933/nipes/7.4.2025.si133>
- [36]. Tambol, T., Derbile, E. K., & Soule, M. (2025). Use of climate smart agriculture technologies in West Africa peri-urban Sahel in Niger. *Scientific Reports*, 15. <https://doi.org/10.1038/s41598-024-82813-w>
- [37]. Tarchiani, V., & Grey, S. (2025). Bridging the communication gap in agrometeorological services: Enhancing the uptake and effectiveness for users in developing countries. *Climate Services*, 37, 100531. <https://doi.org/10.1016/j.cliser.2024.100531>
- [38]. Tuyizere, D., & Tonguz, O. K. (2025). Edge-based speech transcription and synthesis for Kinyarwanda and Swahili languages. *arXiv*. <https://doi.org/10.48550/arxiv.2510.16497>
- [39]. Vaughan, C., Hansen, J., Roudier, P., Watkiss, P., & Carr, E. R. (2019). Evaluating agricultural weather and climate services in Africa: Evidence, methods, and a learning agenda. *WIREs Climate Change*, 10(4), e586. <https://doi.org/10.1002/wcc.586>

- [40]. Wang, J., Adelani, D. I., Agrawal, S., Masiak, M., Rei, R., Briakou, E., Carpuat, M. J., He, X., Bourhim, S., Bukula, A., Mohamed, M., Olatoye, T., Adewumi, T., Mokayed, H., Mwase, C., Kimotho, W., Yuehgoh, F., Aremu, A., Ojo, J., et al. (2024). AfriMTE and AfriCOMET: Enhancing COMET to Embrace Under-resourced African Languages. Proceedings of NAACL-HLT 2024, 5997-6023. <https://doi.org/10.18653/v1/2024.naacl-long.334>
- [41]. Wang, K., Zhang, Z., Lin, A., & Gabriel, S. (2025). The African languages lab: A collaborative approach to advancing low-resource African NLP. arXiv. <https://doi.org/10.48550/arxiv.2510.05644>
- [42]. Yu, H., Xu, T., Hedderich, M. A., Hamidouche, W., Zamir, S. W., & Adelani, D. I. (2026). AfriqueLLM: How data mixing and model architecture impact continued pre-training for African languages. arXiv. <https://doi.org/10.48550/arxiv.2601.06395>
- [43]. Zhang, M., & Adelani, D. I. (2025). AfriMTEB and AfriE5: Benchmarking and adapting text embedding models for African languages. arXiv. <https://doi.org/10.48550/arxiv.2510.23896>