

# NekoSpeak



## Engineering Intelligence at the Edge

**100% Offline • Ultra Low Latency • Privacy First**

**Sivasubramanian Ramanathan**

*Product Owner | Fintech & Digital Innovation  
(Ex-BIS Innovation Hub Singapore)*

# The Problem: AI has a "Last Mile" Issue

In my work exploring **Fintech & RegTech**, I've seen how reliance on cloud APIs creates bottlenecks. For Voice AI on Android, this manifests as:

1. ● **Latency**: Waiting for server responses breaks natural conversation flow.
2. ● **Privacy Risks**: Sending sensitive audio data to the cloud is unacceptable for many use cases.
3. ● **Robotic Fallback**: Traditional offline engines (`espeak`) sound unnatural.

“**Goal**: Build a "Zero-Compromise" engine that runs mostly on-device.”

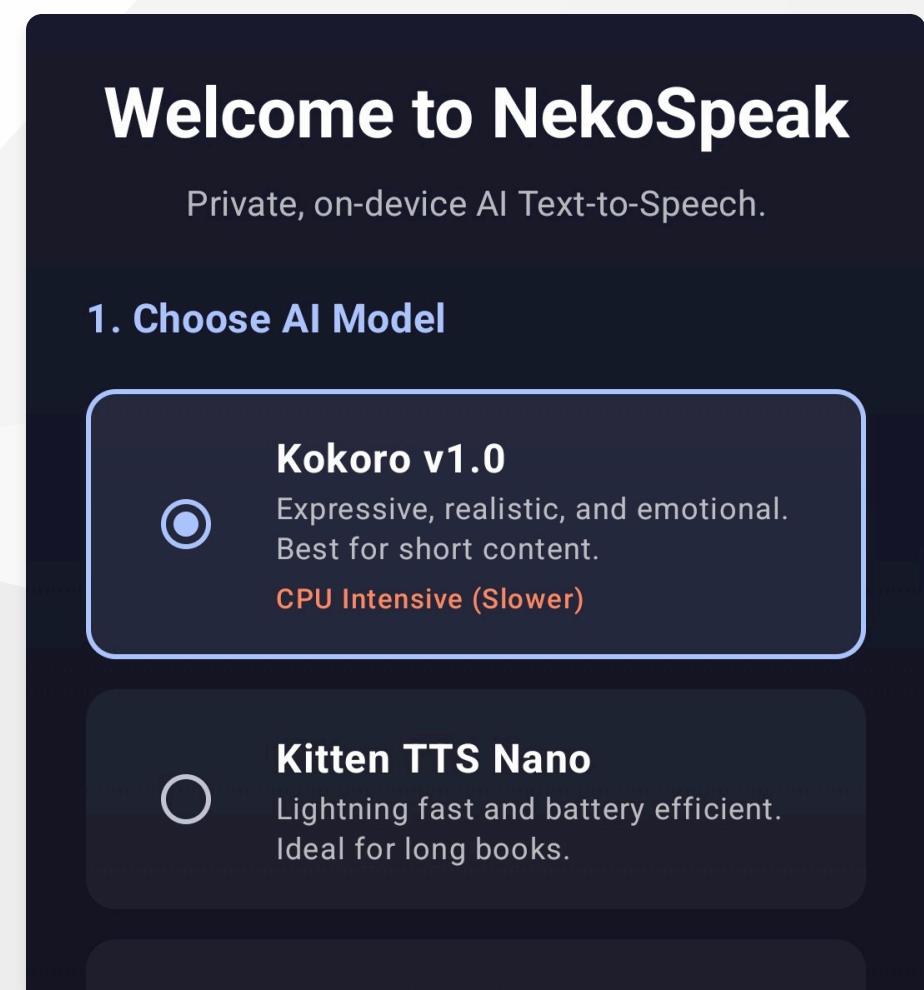
# The Solution: NekoSpeak

A drop-in replacement for the Android TTS ecosystem, bringing heavily quantized Large Audio Models (LAMs) to the mobile edge.

- **Triple Engine Architecture:**

-  **Kokoro (82M)**: Human-level expressiveness.
-  **Piper**: High-speed multilingual inference.
-  **Kitten (Nano)**: Ultra-lightweight fallback.

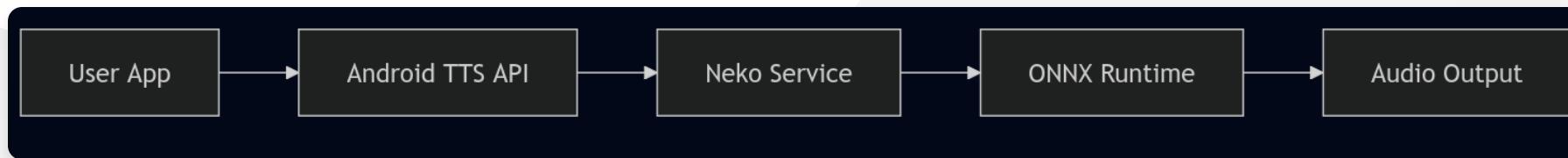
- **100% Offline**: No internet required after download.



# Technical Architecture



I architected a custom pipeline using **ONNX Runtime** and **C++ JNI Bridges** to optimize performance on mobile CPUs.

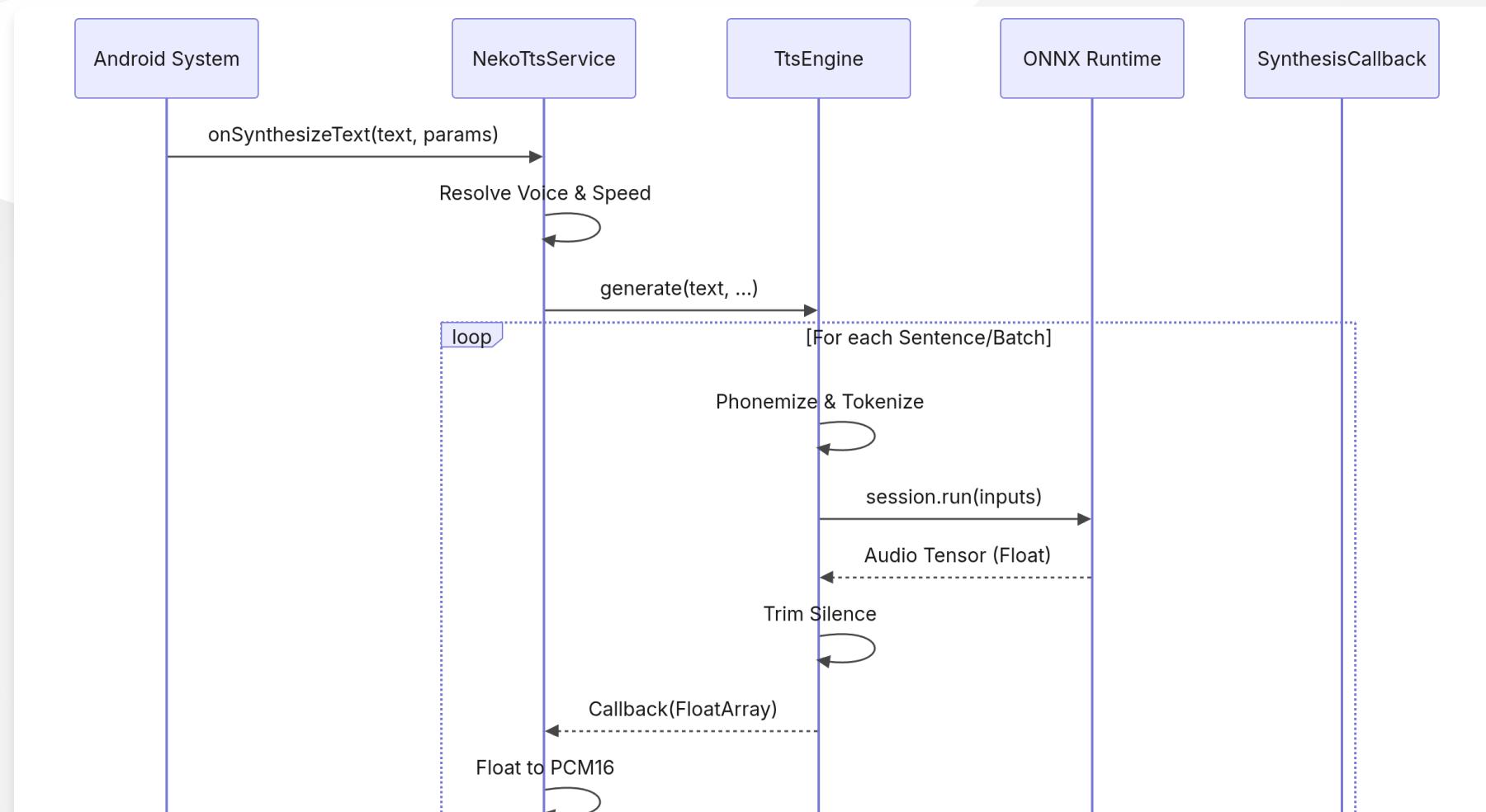


- **Smart Batching:** Dynamic buffering balances latency vs. context window.
- **Native Bridge:** Custom C++ wrapper for `libespeak-ng` phonemization.

# Deep System Integration



Unlike simple "wrapper" apps, NekoSpeak integrates deep into the **Android Framework**.



# Product Showcase



Polished UX focusing on accessibility and ease of use.

The image displays two screenshots of the NekoSpeak mobile application interface. The left screenshot shows the "Voice Selection" screen with the title "Welcome to NekoSpeak" and the subtitle "Private, on-device AI Text-to-Speech.". It features a section titled "2. Choose Starter Voice" with three buttons: "Heart" (highlighted in blue), "Adam", and "Bella". At the bottom are "Back" and "Next" buttons. The right screenshot shows the "Settings & Config" screen with the title "Voice Selection" and "Settings & Config". It includes sections for "AI Model" (with options for Kokoro v1.0 (Standard), Kitten TTS Nano, and Piper (ONNX) - the active voice), "Performance" (CPU Threads: 6 slider), "Stream Buffer: Auto tokens" (reset button), "Speech Speed: 1.0x" (reset button), "Battery" (warning about battery optimization), and "General" (System TTS Settings). Navigation icons for "Voices" and "Settings" are at the bottom.

# Engineering Philosophy & Impact



This project reflects my approach to Product Engineering:

1. **Solve Real Problems**: Bridges the gap between "Cool AI Demo" and "Daily Driver Utility".
2. **Robust Engineering**: "Zero-Crash" architecture with graceful degradation (Cloud -> Local -> Nano).
3. **User-Centric**: Privacy by default, with no hidden analytics.

*Similar to my work on the **Singapore Location Intelligence MCP** and **Client-Side OCR**.*

# About the Builder



**Sivasubramanian Ramanathan**

*Product Owner | Fintech, RegTech & Digital Innovation  
PMP | PSM II | PSPO II*

I specialize in taking messy, real-world complexity and structuring it into reliable products.

**I am looking for my next role in Singapore:** 

- **Focus:** Product Management, Payment Infrastructure, Digital Assets.
- **Value:** Bridging the gap between Policy, Tech, and Business.

# Lets Connect



I am ready to bring this level of engineering rigor and product thinking to your team.

- **Portfolio:** [sivasub.com](http://sivasub.com)
- **LinkedIn:** [linkedin.com/in/sivasub987](https://linkedin.com/in/sivasub987)
- **Code:** [github.com/siva-sub/NekoSpeak](https://github.com/siva-sub/NekoSpeak)

**Download NekoSpeak v1.0.10:**  
[github.com/siva-sub/NekoSpeak/releases](https://github.com/siva-sub/NekoSpeak/releases)