

МОДЕЛЬ СИНТЕЗУ МОВИ NEUTTS AIR

Старкова А.В.

anastasiia.starkova@nure.ua

Науковий керівник – к.т.н., старший викладач Желанов О. О.

Харківський національний університет радіоелектроніки, каф. МІРЕС
м. Харків, Україна

Speech synthesis is a technology that converts written text into audio. Various applications expand the capabilities of this technology, including voice cloning, emotional synthesis, multilingual synthesis, dialect synthesis, and whisper or shout synthesis. Currently, these tasks are accomplished using artificial intelligence models that require external resources for data storage and processing. However, models now exist that can run autonomously on the user's device.

Синтез мови (Text-to-Speech, TTS) – це технологія перетворення текстової інформації на звукову. Сучасні методи TTS забезпечують природне звучання голосу, інтонації та правильні наголоси. Активно застосовуються в голосових помічниках, навігаторах, аудіокнигах, та для людей з обмеженими можливостями.

Більшість сучасних голосових рішень залежать від наявності з'єднання з Інтернет, оскільки спираються на велику кількість даних. Це приводить до обмежень, пов'язаних з апаратною частиною користувача (великий об'єм даних потребує значні ресурси пам'яті). Тому, на сьогодні, завдання створення моделі TTS для локальної роботи є актуальним.

Восени 2025 року компанія Neuphonic представила розробникам і користувачам модель NeuTTS Air – відкриту модель перетворення тексту на мову. Головною особливістю даної моделі є те, що вона працює повністю локально без необхідності підключення до мережі Інтернету або хмарних сервісів. Це однієї особливістю є те, що ця модель дозволяє створювати реалістичну мову прямо на смартфоні, ноутбучі чи іншій апаратній платформі (наприклад, Raspberry Pi, Orange Pi), не надсилаючи дані на сторонні сервери. Модель не просто синтезує голос, вона вміє миттєво клонувати його, використовуючи лише 3 секундний зразок аудіо. Все це відкриває нові можливості для розробників:

- створення голосових помічників, які працюють офлайн (освіта, туризм, побутові рішення);
- програми для людей з обмеженими можливостями;
- інструменти для контент-індустрії (озвучування відео, книг тощо);
- вбудовані системи з голосовим інтерфейсом (розумний дім, робототехніка);
- рішення для сфер із підвищеними вимогами до конфіденційності (медицина, фінанси, юриспруденція).

Моделі NeuTTS Air і NeuTTS-Nano побудовані на основі невеликих LLM-магістралей – легких мовних моделей, оптимізованих для розуміння та генерації тексту. В основі лежить оптимізований трансформер Qwen 0.5 B, який відповідає за розуміння та генерацію тексту.

Переваги даної моделі полягають у наступному.

1. Низька затримка. Відсутність очікування відповіді на мережеві запити, що робить взаємодію між користувачем і пристроєм.

2. Конфіденційність та безпека. Відсутнє завантаження конфіденційної інформації в хмару, що є критично важливим для програм, які потребують захисту конфіденційності користувачів.

3. Офлайн-доступність. Головна функціональність моделі не залежить від підключення до Інтернету.

4. Зменшення витрат. Локальна робота не потребує платних запитів для обробки даних.

5. Оптимізація для локального розгортання. Проект можна легко розгорнути на мобільних платформах.

Однак дана модель має свої обмеження та особливості.

1. Якість вхідних даних: Модель чутлива до якості аудіо. Зашумовані записи, сильні акценти або нестандартна вимова можуть знизити точність клонування.

2. Мовна підтримка: наразі тільки англійська, іспанська, німецька, французька. Архітектура потенційно дозволяє збільшити кількість мов.

3. Продуктивність. В залежності від апаратних можливостей пристрою робота моделі зайняти деякий час.

4. Розмір контексту. Модель оптимізована під короткі фрази та речення. Для синтезу довгих текстів може знадобитися розбиття на частини.

Слід зазначити, що модель NeuTTS включає вбудований водяний знак Perth у кожен згенерований аудіофайл. Цей перцептивний пороговий водяний знак дозволяє ідентифікувати та відстежувати згенерований контент, залишаючись непомітним для слухачів. Це сприяє етичному використанню технології синтезу голосу та допомагає запобігти зловживанню.

Звичайно, «хмарні» рішення пропонують деякі переваги - більші бібліотеки готових голосів, продвинуте управління емоціями, підтримку безлічі мов. Але NeuTTS Air виграє там, де важлива автономність, продуктивність, конфіденційність, а також, відсутність регулярних витрат.

Список використаних джерел: 1. NeuTTS [Електронний ресурс] URL: <https://github.com/neuphonic/neutts?tab=readme-ov-file>, (дата звернення 09.03.2026 р.) 2. About NeuTTS Air [Електронний ресурс] URL: <https://neutts.org/about> , (дата звернення 09.03.2026 р.) 3. Deploy NeuTTS Air: Ultra-Realistic On-Device Voice AI with Instant Voice Cloning Air [Електронний ресурс] URL: <https://www.spheron.network/blog/neutts-air-spheron-voice-ai/> , (дата звернення 09.03.2026 р.)