

ВИКОРИСТАННЯ «TEXT-TO-SPEECH» МОДЕЛІ AUDIOAPI ВІД OPENAI

Зубко М.В., Лебединський А.В.

Харківський національний автомобільно-дорожній університет, Харків

У цій роботі ми розглянемо значення та можливості використання моделі Text-to-Speech (TTS) від OpenAI, відомої як AudioAPI, у сучасному інформаційному середовищі. «Text-to-Speech» є ключовою технологією, яка дозволяє автоматично перетворювати текст на мовлення людського стилю, що має великий потенціал у різних сферах життя.

Модель AudioAPI від OpenAI відрізняється високою якістю мовлення та інтонації, завдяки застосуванню глибоких нейронних мереж. Її застосування розглядається як в сфері технологій, так і в розважальній індустрії, що відкриває безліч можливостей для інновацій та покращення користувацького досвіду.

AudioAPI від OpenAI – це продукт, який дозволяє створювати реалістичне мовлення з використанням штучного інтелекту. Огляд моделі AudioAPI від OpenAI включає в себе детальне розглядання технологічних особливостей та потенціалу цієї системи.

Ця модель базується на передових методах глибокого навчання та нейронних мережах, що забезпечує високу якість мовлення. Вона натренована на великому обсязі аудіоданих, що дозволяє їй відтворювати мовлення з природним звучанням та інтонацією. Однією з ключових переваг моделі AudioAPI є її здатність генерувати реалістичне мовлення, яке легко розрізнити від синтезованого. Ця модель володіє широким набором голосових характеристик, що дозволяє користувачам вибирати підходящий для них стиль мовлення.

OpenAI's Text-To-Speech API автоматично розпізнає мову введення та генерує мову мовою введення тексту. API підтримує різні мови, зокрема англійську, китайську, гінді, іспанську, французьку, арабську, бенгальську, російську, португальську та індонезійську [2]. У той час як голоси за замовчуванням оптимізовано для англійської мови, API добре працює на інших мовах, адаптуючись до мови введення, щоб створювати розмовний звук потрібною мовою [3].

Користувачі повідомляють, що такі мови, як німецька, французька, іспанська та англійська, звучать майже ідеально без акцентів, демонструючи здатність API ефективно обробляти різні мови [2, 3]. Крім того, API дозволяє здійснювати потокове аудіо в реальному часі за допомогою кодування передачі фрагментів, що дозволяє відтворювати аудіо до того, як буде створено повний файл і стане доступним.

Функціональні можливості:

- гнучкість вибору голосів: AudioAPI надає можливість вибору різних голосів для генерації мовлення відповідно до ваших потреб;
- реалістичне мовлення: модель здатна створювати мовлення з природним інтонацією, ритмом та акцентом, надаючи вражаючу автентичність;
- підтримка різних мов: AudioAPI може працювати з різними мовами, що розширює її застосування на міжнародному рівні.

Застосування в технологіях:

- синтезатори мовлення для додатків: Модель AudioAPI може бути використана для створення синтезаторів мовлення у додатках [4], що дозволяє користувачам отримувати аудіо відтворення тексту, наприклад, для читання книг або повідомлень;
- асистенти для людей з обмеженими можливостями: Технологія TTS може бути корисною для людей з вадами зору або іншими обмеженнями, які затрудняють читання тексту. Модель AudioAPI може бути використана для створення асистентів, що допомагають їм отримувати аудіо відтворення текстової інформації.

Застосування в розвагах:

- аудіокниги та подкасти: використання моделі AudioAPI дозволяє створювати високоякісні аудіокниги та подкасти, що забезпечують природне мовлення та підвищують враження від аудіального контенту;
- відеоігри: технологія TTS може бути використана для надання голосових характеристик персонажам у відеоіграх, що покращує іммерсію гравців та забезпечує додатковий рівень взаємодії.

У висновку можна сказати, що використання моделі AudioAPI від OpenAI відкриває безліч можливостей у багатьох сферах, включаючи розваги, технології доступності та електронну комерцію. AudioAPI забезпечує можливість контролювати інтонацію мовлення, що робить її більш гнучкою для різних застосувань. Завдяки постійному вдосконаленню технологій штучного інтелекту, модель AudioAPI продовжує покращуватися в якості та швидкості генерації мовлення. Вона є потужним інструментом для розробників та дослідників, які працюють у сферах, пов'язаних з обробкою природної мови та аудіовиведенням. Її потенціал вже застосовується у різних галузях, від веб-розробки до поліпшення доступності контенту.

Список використаних джерел

1. Учасники проєктів Вікімедіа. Синтез мовлення – Вікіпедія. *Вікіпедія*. URL: https://uk.wikipedia.org/wiki/Синтез_мовлення (дата звернення: 11.04.2024).
2. Can I choose the TTS language?. *OpenAI Developer Forum*. URL: <https://community.openai.com/t/can-i-choose-the-tts-language/477491> (date of access: 11.04.2024).
3. Text to Speech Guide. *OpenAI*. URL: <https://platform.openai.com/docs/guides/text-to-speech> (date of access: 11.04.2024).
4. Building With Bubble. How To Add OpenAI Text-To-Speech To Your Bubble App (Complete Guide), 2024. *YouTube*. URL: <https://www.youtube.com/watch?v=6dicvYNHyDw> (date of access: 11.04.2024).