

УДК 004.4

П'ятерня П.О.¹, Гладкова О.М.²

¹ студ. гр. КНТ-222м НУ «Запорізька політехніка»

² канд. техн. наук, доц. НУ «Запорізька політехніка»

ДОСЛІДЖЕННЯ ТА ПРОГРАМНА РЕАЛІЗАЦІЯ НАВЧАННЯ ГОЛОСОВОЇ МОДЕЛІ ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ

Об'єкт дослідження – сучасні підходи та засоби для навчання голосової моделі.

Предмет дослідження – дослідження та навчання голосової моделі з можливістю перетворення вхідного голосу на голос моделі.

Мета роботи – розробка програмного забезпечення, яка навчає голосову модель за допомогою штучного інтелекту для можливості перетворення голосу.

Матеріали, методи та технічні засоби: штучний інтелект, глибоке навчання, датасет, віддалена лабораторія Google Colab.

Розроблена комп'ютерна система, яка перетворює голос людини в голос навченої моделі.

Дана система є оптимальною з точки зору апаратних витрат, оскільки вона обробляє аудіо з датасету на платформі Colab у хмарному середовищі та не потребує додаткової інфраструктури для своєї роботи.

У роботі досліджено актуальну проблему клонування голосу людини. Розглянуто переваги та недоліки існуючих технологій та систем для її вирішення, сфери застосування таких систем та сучасні підходи до їх розв'язання.

Досліджено засоби для вирішення поставленої проблеми. Розглянуті моделі нейронних мереж придатні для навчання голосової моделі.

Було проведено порівняння сучасних технологій для навчання голосових моделей. У результаті порівняння між SVC-SO-VITS, Whisper і Vox2Vox було обрано SVC-SO-VITS, як архітектура, що найліпше підходить до вирішення поставленого завдання.

Проведено навчання моделі нейронної мережі для перетворення голосу людини на голос моделі: налаштоване середовище для навчання, підготовлені записи голосу людини з відповідного датасету, створені необхідні для тренування файли та успішно виконано навчання моделі.

Розроблена готова до впровадження комп'ютерна система, яка перетворює вхідний голос людини на голос навченої моделі Петра Порошенка.

Навчена нейронна мережа дозволяє клонувати голос людини. Абсолютна відсоткова похибка становить не більше 20%. При цьому виявлено, що точність розробленої системи залежить від якості навчання моделі нейронної мережі, яка у свою чергу залежить від якості датасету.

Наукова новизна роботи полягає у застосуванні та навчанні голосової моделі SVC-SO-VITS. А також у використанні інноваційного підходу до навчання, який дозволяє досягти значного успіху в області клонування голосу людини.

Поставлена проблема навчання голосової моделі за допомогою штучного інтелекту повністю вирішена з допустимою точністю.

Галузь використання – індустрія розваг, аудіо-контент, системи автоматизованого виклику, синтез голосу для віртуальних асистентів та в інших областях, де необхідно відтворення голосового контенту з високою якістю та без значних витрат.