

УДК 004.934

**І. Гарбар, студент 2-го курсу магістратури кафедри програмних систем і технологій;
К. Духновська, к. т. н., доцент кафедри програмних систем і технологій
(Київський національний університет імені Тараса Шевченка, Україна)**

АДАПТИВНА СИСТЕМА РОЗПІЗНАВАННЯ МОВЛЕННЯ ТА ПЕРЕКЛАДУ

UDC 004.934

**I. Harbar, 2nd year Master's student, Department of Software Systems and Technologies;
K. Dukhnovska, PhD., Associate Professor, Department of Software Systems and Technologies**

ADAPTIVE SPEECH RECOGNITION AND TRANSLATION SYSTEM

Об'єктом дослідження обрано корейську мову, автоматична обробка якої ускладнена контекстуальною фонетикою та розгалуженою ієрархією ввічливості. На сучасному етапі користувачі змушені шукати компроміс між високоточними хмарними сервісами, використання яких пов'язане з ризиками для приватності, та захищеними, проте менш функціональними автономними інструментами [1].

У роботі обґрунтовано програмна архітектура, спрямована на вирішення оптимізаційної задачі співвідношення точності обробки мовлення та конфіденційності даних. Наукова новизна дослідження полягає у комплексній інтеграції офлайн-моделей та хмарних сервісів у межах єдиної системи автоматичного розпізнавання мовлення (ASR) [2] та машинного перекладу. Реалізовано модульний підхід із використанням стандартизованого конвеєра попередньої обробки аудіосигналів. З метою нівелювання впливу акустичних завад застосовано методи спектрального віднімання, нормалізації амплітуди та сегментації сигналу, що гарантує адаптивність системи до гетерогенних умов експлуатації. Для забезпечення гнучкості підсистеми ASR розроблено гібридну архітектуру, яка поєднує дві моделі з комплементарними характеристиками. Зокрема, модель Vosk забезпечує ефективну конфіденційну обробку на пристроях з обмеженими обчислювальними ресурсами, тоді як модель Whisper гарантує максимальну точність транскрипції навіть за наявності шумів та діалектних варіацій. Такий підхід уможливорює контекстно-залежний вибір між оперативністю автономного режиму та прецизійністю ресурсоємних обчислень. Для підвищення лінгвістичної точності підсистеми машинного перекладу, застосовано метод зваженого ансамблю. Базуючись на синергії «універсальної» моделі та «спеціалізованої» моделі, даний метод демонструє значну перевагу над мономоделними рішеннями.

Результати експериментальних досліджень підтверджують ефективність запропонованої гібридної системи: досягнуто показників WER на рівні 5% та BLEU – 0.85. Модульна архітектура успішно забезпечує баланс між точністю та приватністю, що дозволяє рекомендувати розробку для впровадження у прикладні задачі різного профілю.

Література

1. Radford, A., Kim, J. W., Xu, T., et al. (2023). Robust Speech Recognition via Large-Scale Weak Supervision. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*.
2. Povey, D., Ghoshal, A., Boulianne, G., et al. (2011). The Kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*.