

Лабораторна №1

Загальні правила

- Здача роботи зараховується через заповнення та надсилання Google Form:
<https://forms.gle/Fm1FskfWbr5cSXaH9>
- **Дедлайн здачі всіх матеріалів роботи - 16.03.2026, дедлайн захисту - 20.03.2026. ВАЖЛИВО для студентів з КПІ: Якщо робота не буде захищена до 20.03.2026 - студент буде НЕ атестований**
- Проект можна виконувати групами до Зъох людей. **Бажано об'єднатися в групи 3 людини**
- Треба виконати **ОБИДВА завдання**
- Відповіді на теоретичні питання в Google Form будуть оцінені під час захисту й будуть впливати на фінальну оцінку, проте лектор може додатково поставити теоретичні питання

Завдання № 1

Варіант 1

Побудувати модель розпізнавання фонем, використовуючи ознаки з моделей навчання без вчителя (self-supervised learning (SSL) models)

1. Використайте [TIMIT Dataset](#)
2. Побудуйте спеціалізовану модель поверх SSL ознак
 - a. Спробуйте кілька SSL-моделей — можна використати моделі з лекції та/або дослідити [HF Model Hub](#) і/або переглянути [SUPERB Leaderboard](#).
 - b. У лекції ми використовували last_hidden_state. Поекспериментуйте з іншими шарами; наприклад, перегляньте [HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units](#).
3. Для оцінювання використовуйте PER (Phoneme Error Rate). Використовуйте попередньо визначений розподіл на train/test.

ВАЖЛИВО: СТС ф-ція втрат не є єдиним варіантом вирішення цієї задачі, цю задачу можливо вирішити використовуючи “обережне” розбиття вікнами й СЕ ф-ція втрат. В разі використання СТС ф-ції втрат студент має знати математичний і програмний принцип її роботи.

Варіант 2

Побудувати модель розпізнавання мовлення, використовуючи ознаки з моделей навчання без вчителя (self-supervised learning (SSL) models)

1. Використайте [OpenTTS UK dataset](#)
2. Те саме, що і пункт 2 з Варіанта 1
3. Оцініть модель на тестовому наборі [Common Voice 10](#). Для оцінювання використовуйте WER (Word Error Rate) та CER (Character Error Rate).

ВАЖЛИВО: СТС ф-ція втрат не є єдиним варіантом вирішення цієї задачі, цю задачу можливо вирішити використовуючи “обережне” розбиття вікнами й СЕ ф-ція втрат. В разі використання СТС ф-ції втрат студент має знати математичний і програмний принцип її роботи.

В Звіт необхідно додати

- Огляд даних, основні особливості й проблеми
- Опис використаних підходів
- Таблиці метрик
- Гіпотези й думки, щодо отриманих результатів

Завдання № 2

Застосувати аудіо фільтри та інші методи обробки аудіо (без використання методів, які базуються на глибоких нейронних мережах) для покращення якості даних

Варіант 1

Покращити якість TIMIT датасету

1. Розробити й протестувати аудіо фільтри для покращення якості TIMIT датасету
2. Протестувати чи було отримане покращення по PER метриці
3. Запропонувати інші об'єктивні метрики для оцінки якості даних

Варіант 2

Покращити якість OpenTTS UK dataset та Common Voice 10 та опціонально [Toronto dataset](#)

4. Розробити й протестувати аудіо фільтри для покращення якості моделей розпізнавання мовлення
5. Протестувати чи було отримане покращення по метриках WER та CER
6. Запропонувати інші об'єктивні метрики для оцінки якості даних
7. [Опціонально] Перевірити попередні 2 пункти на [Toronto dataset](#)

В Звіт необхідно додати

- Опис запропонованих алгоритмів фільтрування
- Абляційне дослідження по впливу алгоритмів фільтрування на фінальну якість моделей

Що необхідно підготувати

- Звіт у форматі PDF (LaTex) або Markdown
- Програмний код на GitHub
- Заповнити Google Form

Політика щодо використання ШІ

Студент може використовувати інструменти ШІ, але має валідувати, перевіряти й розуміти отримані результати

Позитивний сценарій використання ШІ: “Написати ф-цю фільтрування за допомогою Chat GPT, після цього перевірити її коректність й зрозуміти логіку роботи”

Негативний сценарій використання ШІ: “Завантажити файл завдання Лабораторної, щоб ШІ зробив її усю”

Суворо забороняється використання ШІ під час захисту робіт!!!

Студент має розуміти кожен рядок свого коду та звіту й мати змогу пояснити його й відповісти на питання щодо нього !!!

Система оцінювання (приблизна)

Елемент	Максимальний бал
Завдання 1 (код)	3
Завдання 1 (звіт)	2
Завдання 1 (захист + питання)	4
Завдання 2 (код)	3
Завдання 2 (звіт)	2
Завдання 2 (захист + питання)	4