

УДК 004.8:004.934:316.77

Руденський Роман Анатолійович*доктор економічних наук, професор, професор кафедри комп'ютерних наук,
Національний університет біоресурсів і природокористування України*ORCID: <https://orcid.org/0009-0002-3682-9702>E-mail: roman.rudensky@nubip.edu.ua**Кравченко Володимир Миколайович***доктор економічних наук, доцент, професор кафедри економічної кібернетики,
Національний університет біоресурсів і природокористування України*ORCID: <https://orcid.org/0000-0002-8033-3985>E-mail: v.kravchenko@nubip.edu.ua**СИСТЕМА КОМПЛЕКСНОГО АНАЛІЗУ КОМУНІКАТИВНОЇ ПОВЕДІНКИ В ПУБЛІЧНИХ ДИСКУСІЯХ НА ОСНОВІ ДІАРИЗАЦІЇ МОВЦІВ**

Анотація. Актуальність дослідження зумовлена зростанням кількості онлайн-зустрічей та публічних дискусій у цифровому форматі, що створює потребу в автоматизованих інструментах аналізу групової комунікації. Традиційні методи ручного кодування та транскрипції є надзвичайно трудомісткими та суб'єктивними, що обмежує можливості масштабного дослідження комунікативних патернів. Мета дослідження полягає в розробці та апробації комплексної системи автоматизованого аналізу комунікативної поведінки, яка поєднує сучасні технології діаризації мовців, автоматичного розпізнавання мови та статистичного аналізу для надання детальної картини групової динаміки в публічних дискусіях. Методи. Система реалізована на основі мікросервісної архітектури з використанням Python 3.10+, FastAPI та React. Для діаризації мовців застосовано алгоритм ruannote.audio, що поєднує конволюційні енкодеру з попередньо навченими моделями WavLM. Автоматичне розпізнавання мови здійснюється через трансформерні архітектури (Whisper, AssemblyAI, Conformer). Аналіз комунікативної поведінки включає обчислення статистичних метрик активності, мережевий аналіз взаємодій та оцінку стилю комунікації. Результати. Розроблена система успішно інтегрує діаризацію мовців з точністю до 0.5 секунди, автоматичну транскрипцію та багатомірний аналіз комунікативних патернів. Модульна архітектура забезпечує гнучкість адаптації під різні домени застосування. Система генерує детальні часові мітки активності учасників, візуалізує розподіл часу мовлення та надає комплексну аналітику для покращення процесів прийняття рішень. Перспективи. Подальший розвиток системи включає інтеграцію мультимодального аналізу з урахуванням невербальної комунікації, покращення стабільності роботи в умовах фонового шуму, доменну адаптацію для специфічних галузей та впровадження real-time аналізу живих дискусій. Система відкриває нові можливості для дослідження групової динаміки в корпоративному, освітньому та державному секторах.

Ключові слова: діаризація мовців, автоматичне розпізнавання мови, аналіз комунікативної поведінки, штучний інтелект, публічні дискусії.

Актуальність. Особливо актуальними є дослідження комунікативних патернів у контексті цифрової трансформації суспільства та збільшення кількості віртуальних взаємодій. Пандемія COVID-19 значно прискорила перехід до онлайн-форматів зустрічей, що створило нові виклики для аналізу групової динаміки в цифровому середовищі.

В свою чергу, аналіз групової комунікації відіграє ключову роль у розумінні динаміки колективного сприйняття ситуації та прийняття рішення, налагодженні якісного публічного дискурсу й ефективної командної роботи. З розвитком цифрових технологій кількість аудіо та відео записів зустрічей, дискусій та публічних обговорень зростає експоненційно, що створює потребу в автоматизованих інструментах їх аналізу.

В умовах, коли наявний запис дискусії не має розведення учасників по різних аудіо каналах, або відсутні метадані, які прямо вказують який фрагмент аудіо пов'язаний із кожним учасником, перед проведенням аналізу необхідно забезпечити розпізнавання учасників на запису. Такий процес автоматичного визначення "хто і коли говорить" в аудіозаписі називається діаризацією мовців. Саме це являє собою одну з ключових технологій розуміння динаміки групової взаємодії, оскільки традиційні методи аналізу групової комунікації базуються на ручному кодуванні та транскрипції, що є надзвичайно трудомістким процесом.

Експерти витрачають години на аналіз навіть коротких записів, що робить масштабне дослідження комунікативних патернів практично неможливим. Більше того, людський фактор вносить суб'єктивність у процес аналізу, що може впливати на достовірність результатів.

Досягнення в області автоматичного розпізнавання мови (ASR) та діаризації мовців пройшли значну еволюцію від статистичних методів і прихованих Марківських моделей до глибокого навчання і трансформерних архітектур. Наразі вони відкривають нові можливості для автоматизації аналізу групових дискусій.

Аналіз останніх досліджень та публікацій. Дослідження в галузі автоматичного аналізу групових дискусій активно розвиваються протягом останнього десятиліття, зокрема завдяки прогресу в технологіях глибокого навчання. Бредін Х. разом зі співавторами [1] представили ruannotate.audio – нейронну систему для діаризації мовців, що стала де-факто стандартом у галузі завдяки високій точності та модульній архітектурі [2]. Радфорт А. і колеги [3] розробили Whisper – трансформерну модель для розпізнавання мови, навчену на 680 000 годинах багатомовних даних, що демонструє значну стійкість до фонового шуму та акцентів. Колектив науковців на чолі з Гулаті А. [4] запропонували архітектуру Conformer, яка поєднує конволюційні та трансформерні шари для ефективного моделювання як локальних, так і глобальних залежностей в аудіосигналі, досягаючи state-of-the-art результатів на benchmark датасетах LibriSpeech.

У контексті аналізу комунікативної поведінки, Ао Дж. та співавтори [5] представили SpeechT5 – уніфіковану енкодер-декодерну архітектуру для обробки усного мовлення, що дозволяє виконувати множину задач, включаючи розпізнавання, синтез та аналіз тональності. Дослідження показують, що інтеграція діаризації мовців з автоматичним розпізнаванням мови значно покращує точність транскрипції в умовах групових дискусій порівняно з традиційними підходами. Більш того, Парк Т. Дж. та співавтори [6] запропонували AutoVC для zero-shot стилістичного переносу голосу, що розширює можливості аналізу комунікативних стилів у мультимодальних дискусіях без додаткового тренування.

Однак більшість існуючих рішень зосереджені на окремих аспектах аналізу – або на діаризації, або на транскрипції, або на аналізі тональності. Комплексні системи, що поєднують усі ці компоненти з можливістю детального аналізу комунікативних ролей та патернів взаємодії, залишаються недостатньо представленими в науковій літературі. Крім того, існуючі рішення не забезпечують цільового аналізу дискусій відповідно до заданих тем чи ключових слів, що обмежує можливості оцінки глибини розкриття окремих тематичних блоків та їх представленості в комунікативному процесі.

Мета дослідження полягає в розробці та апробації комплексної системи автоматизованого аналізу комунікативної поведінки, яка поєднує сучасні технології діаризації мовців, автоматичного розпізнавання мови та статистичного аналізу для надання детальної картини групової динаміки в публічних дискусіях.

Результати дослідження та їх обговорення. Трансформерні архітектури, такі як Whisper, революціонізували галузь завдяки механізмам уваги (attention mechanisms), які дозволяють моделі фокусуватися на релевантних частинах аудіосигналу. Це особливо важливо під час обробки довгих аудіозаписів та роботи в умовах значного фонового шуму. У табл. 1 наведено порівняльний аналіз провідних ASR систем, що використовуються в дослідженні. Кожна система має унікальні переваги та оптимальні сценарії використання.

Для здійснення діаризації мовців, що є критично важливим в аналізі групових дискусій, використано алгоритм навчання без вчителя, реалізований в системі ruannotate.audio. Ця система являє собою «state-of-the-art» рішення, що поєднує конволюційні енкодери з попередньо навченими моделями типу WavLM для досягнення високої точності розділення мовців.

Алгоритм розділення учасників дискусії (рис. 1) функціонує за наступними етапами:

1. *Паралельна обробка вхідного сигналу:* аудіосигнал одночасно обробляється через 1D конволюційний енкодер (для вилучення акустичних ознак) та WavLM модель (для отримання семантичних представлень).

Таблиця 1 – Порівняльна характеристика ASR систем

Модель	Характеристики	Рекомендації для використання
Whisper	Трансформерна архітектура з механізмом уваги для багатомовного ASR. Підтримує 99 мов та демонструє високу стійкість до шумів	<ul style="list-style-type: none"> • Багатомовні додатки • Транскрипція інтерв'ю та лекцій • Субтитри для відео • Робота з шумним аудіо
AssemblyAI	Хмарна платформа для ASR з підтримкою реального часу, діаризації мовців та аналізу настроїв	<ul style="list-style-type: none"> • Бізнес-застосування (call centers) • Аналіз розмов та зустрічей • API-інтеграції
Conformer	Гібридна архітектура, що поєднує конволюційні та трансформерні шари для ефективного моделювання локальних та глобальних залежностей	<ul style="list-style-type: none"> • Високоточна транскрипція • Критичні застосування (медицина, право) • Production системи з великими ресурсами

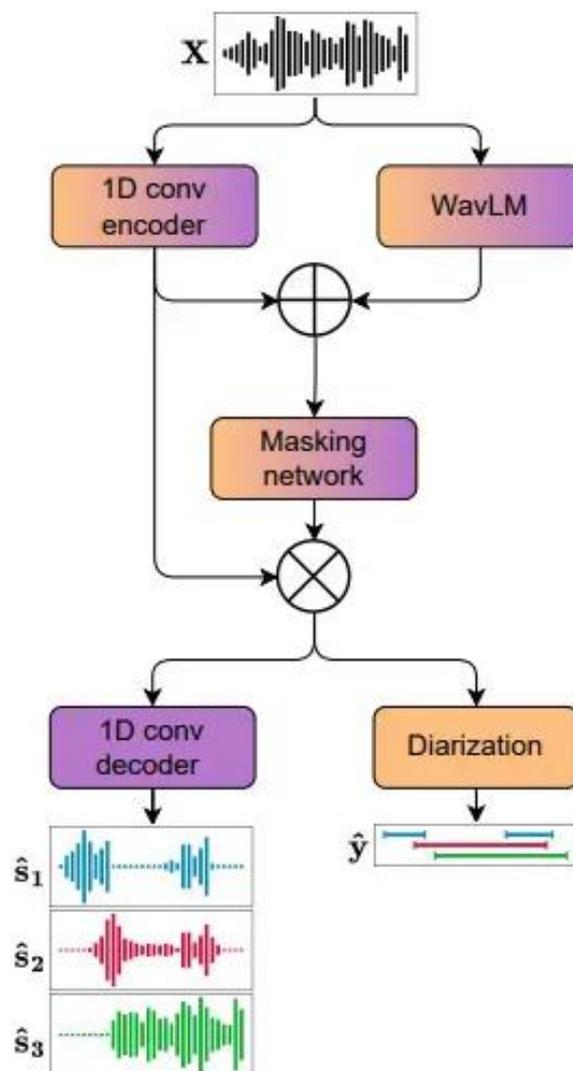


Рисунок 1 – Схема архітектури системи ruannote для діаризації мовців

2. *Об'єднання ознак*: результати з обох гілок комбінуються для створення багатовимірного представлення аудіосигналу із подальшою кластеризацією.

3. *Застосування *masking network**: нейронна мережа створює маски для розділення мовців на основі об'єднаних ознак та кластерів.

4. *Подвійний вихід*: система генерує як розділені аудіосигнали для кожного мовця, так і часові мітки активності.

Ця комплексна архітектура дозволяє ефективно розділяти накладені голоси та точно визначати часові інтервали активності кожного мовця, що є критично важливим для подальшого аналізу комунікативної поведінки.

Система аналізу комунікативної поведінки.

Комунікативна поведінка в групових дискусіях може бути охарактеризована через кілька ключових аспектів, які піддаються кількісному аналізу. Розроблена методологія базується на інтеграції статистичних методів, мережевого аналізу та машинного навчання для всебічного розуміння групової динаміки. В табл. 2 окреслено основні поведінкові аспекти та відповідні методи аналізу, які були адаптовані для роботи з результатами автоматичної діаризації мовців.

Таблиця 2 – Система комплексного аналізу комунікативної поведінки

Поведінковий аспект	Суть аналізу	Методи та підходи
Час і активність у розмові	Визначення рівня залученості кожного учасника, виявлення домінування або пасивності в обговореннях	Підрахунок часу мовлення, аналіз пауз, статистика активності
Ролі та лідерство	Виявлення неформальних лідерів, розуміння того, хто ініціює ідеї та веде обговорення	Аналіз взаємодій, виявлення впливових учасників, мережевий аналіз
Стиль комунікації	Оцінка складності подачі інформації, професійності мови, адаптації під аудиторію	Аналіз словника, оцінка складності мови, частотний аналіз

Система спроектована за принципами мікросервісної архітектури, що забезпечує масштабованість, надійність та легкість підтримки. Архітектура включає три основні рівні: рівень представлення (frontend), рівень бізнес-логіки (backend) та рівень даних (рис 2).

Система розроблена з використанням сучасного технологічного стеку, що забезпечує високу продуктивність та масштабованість. Backend реалізовано на Python 3.10+ з фреймворком FastAPI для створення RESTful API, який було обрано через високу продуктивність, автоматичну генерацію документації та нативну підтримку асинхронного програмування. Frontend побудовано на React для створення інтерактивного веб-інтерфейсу з підтримкою real-time оновлень через WebSocket з'єднання. Для забезпечення портабельності та консистентності розгортання у різних середовищах застосовано контейнеризацію на основі Docker, а Nginx використовується як reverse проху для балансування навантаження та обслуговування статичних файлів. Для реалізації функціоналу машинного навчання інтегровано спеціалізовані AI/ML бібліотеки: ruannotate.audio для діаризації, AssemblyAI для хмарного ASR, Rydub та MoviePy для обробки медіафайлів, NLTK для аналізу тексту.

Запропонована система реалізує багатоетапну процедуру підготовки та обробки даних, побудовану за принципом послідовного перетворення вхідної інформації від необробленого медіаконтенту до структурованої аналітики комунікативної поведінки. Цей процес можна розділити на п'ять основних фаз, кожна з яких вирішує специфічні завдання обробки та збагачення даних.

Фаза 1: Завантаження та попередня обробка. Система приймає файли у широкому спектрі форматів, включаючи аудіоформати MP3 та WAV, а також відеоформати MP4, AVI та

інші популярні формати. На цьому етапі виконується валідація завантажених файлів для перевірки їх цілісності та коректності структури. Ключовим кроком є нормалізація аудіосигналу, що включає приведення всіх вхідних файлів до єдиного sample rate 16kHz, який є оптимальним для більшості моделей автоматичного розпізнавання мови. Для відеофайлів виконується автоматична екстракція звукової доріжки з подальшим її збереженням у проміжному форматі для обробки. Додатково застосовуються алгоритми фільтрації фонового шуму для підвищення якості вхідного сигналу, що критично важливо для точності подальших етапів діаризації та транскрипції. Система також виконує сегментацію довгих аудіозаписів на керовані фрагменти для оптимізації використання пам'яті та прискорення обробки.

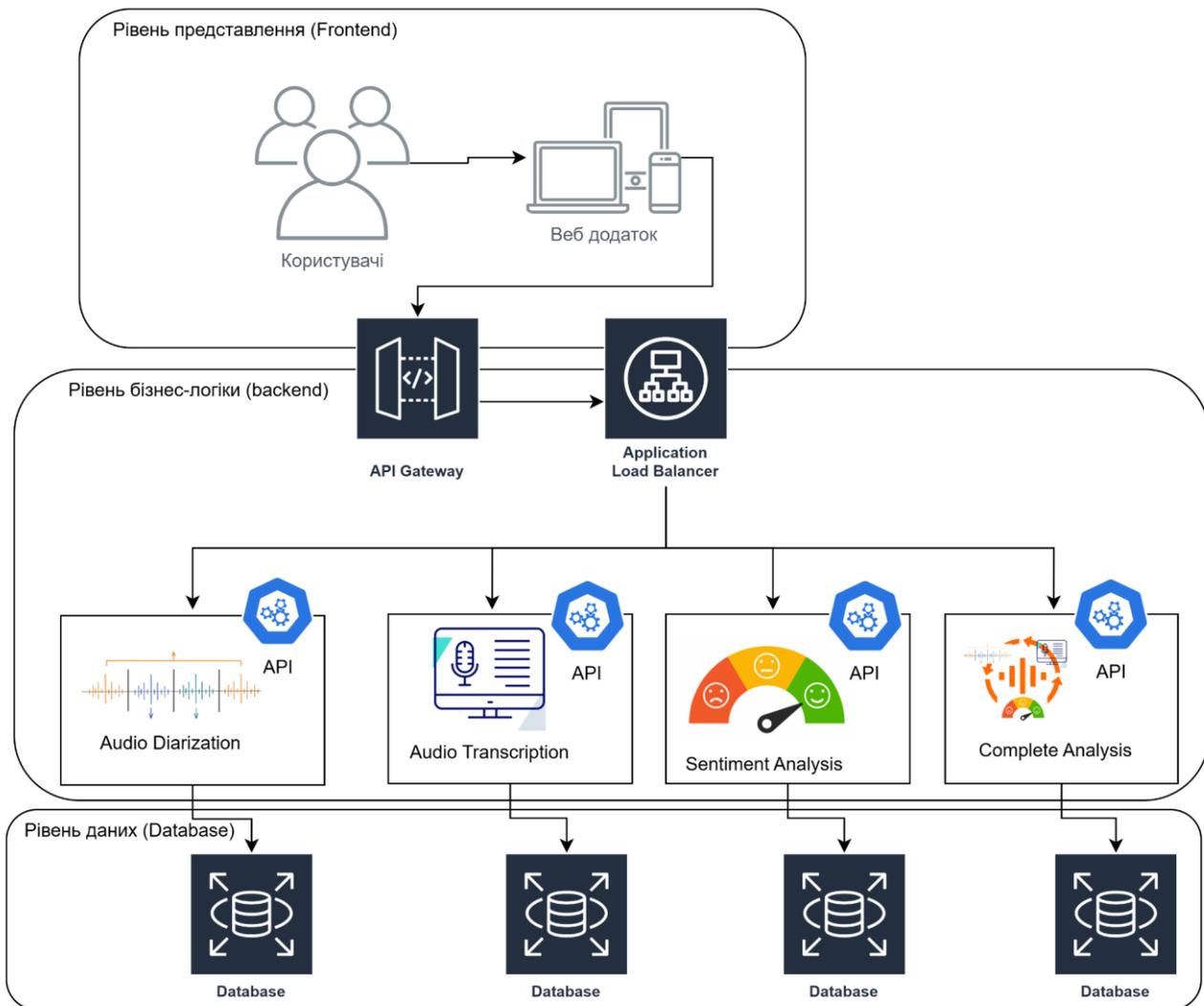


Рисунок 2 – Мікросервісна архітектура системи аналізу комунікативної поведінки

Фаза 2: Діаризація мовців. На цьому етапі застосовується алгоритм `ruannotate.audio` для ідентифікації та розділення мовців у попередньо обробленому аудіосигналі. Система автоматично визначає кількість унікальних голосів в аудіозаписі та присвоює кожному мовцю унікальний ідентифікатор. Алгоритм аналізує акустичні характеристики голосу, включаючи тембр, тональність та частотні особливості, для побудови голосових профілів учасників. Генеруються детальні часові мітки активності кожного учасника з точністю до 0.5 секунди, що включає початок та кінець кожного сегменту мовлення. Система також виявляє та обробляє випадки накладення голосів, коли декілька учасників говорять одночасно, що є типовою ситуацією в живих дискусіях. Результати діаризації зберігаються у структурованому

форматі з метаданими про кожного мовця, тривалість їхньої активності та розподіл часу протягом всієї дискусії.

Фаза 3: Автоматичне розпізнавання мови. Кожен виділений на попередньому етапі сегмент мовлення обробляється через вибрану систему автоматичного розпізнавання мови (ASR) для перетворення аудіо в текстовий формат. Система підтримує інтеграцію з декількома ASR-движками, включаючи Whisper, AssemblyAI та Conformer, що дозволяє обирати оптимальну модель залежно від мови контенту, акустичних умов та вимог до точності. Транскрибований текст автоматично прив'язується до конкретного мовця та часових міток, забезпечуючи повну відстежуваність кожного фрагменту розмови. Система зберігає інформацію про рівень впевненості розпізнавання для кожного слова, що дозволяє ідентифікувати потенційно некоректно розпізнані фрагменти. Додатково виконується постобробка тексту, включаючи нормалізацію пунктуації, виправлення типових помилок розпізнавання та сегментацію на речення для покращення читабельності результатів.

Фаза 4: Аналіз комунікативної поведінки. На основі отриманих на попередніх етапах даних система виконує комплексний аналіз комунікативних патернів учасників дискусії. Обчислюються статистичні метрики активності, включаючи загальний час мовлення кожного учасника, середню тривалість реплік, частоту перебувань та розподіл пауз між висловлюваннями. Система виявляє патерни взаємодії між учасниками через аналіз послідовностей реплік, що дозволяє ідентифікувати домінуючих мовців, пасивних учасників та динаміку зміни ролей протягом дискусії. Виконується аналіз емоційної динаміки на основі лексичного аналізу транскрибованого тексту з використанням словників тональності та класифікаторів настроїв. Додатково аналізується складність мови кожного учасника через обчислення метрик лексичного різноманіття, довжини речень та використання професійної термінології. За наявності вказаних ключових слів або тем система проводить цільовий аналіз їх згадування різними учасниками, оцінюючи глибину розкриття тематичних блоків та емоційне забарвлення відповідних фрагментів дискусії.

Фаза 5: Візуалізація та звітність. Завершальна фаза присвячена представленню результатів аналізу в зручному для інтерпретації форматі. Система генерує інтерактивні звіти, що включають часові діаграми активності учасників з можливістю перегляду деталей конкретних сегментів дискусії. Візуалізується розподіл часу мовлення у вигляді кругових діаграм та гістограм, що дозволяє швидко оцінити баланс участі в обговоренні. Створюються графіки емоційної динаміки, що показують зміну тональності висловлювань протягом часу для кожного учасника та дискусії в цілому. Статистика участі включає кількісні показники активності, середні тривалості реплік, частоту ініціювання нових тем та реакції на висловлювання інших учасників. Всі результати супроводжуються аналітичними висновками щодо комунікативних ролей учасників, ефективності взаємодії та ключових моментів дискусії. Система надає можливість експорту звітів у різних форматах, включаючи PDF для презентацій, CSV для подальшого статистичного аналізу та JSON для інтеграції з іншими аналітичними платформами.

Компоненти веб-додатку та функціональність.

Компоненти веб-додатку та функціональність. Розроблений веб-додаток структуровано у вигляді модульної системи, що дозволяє користувачам обирати відповідний рівень деталізації аналізу залежно від конкретних потреб дослідження. Архітектура включає три основні компоненти, кожен з яких реалізує специфічну функціональність аналізу та може використовуватися як окремо, так і в комбінації з іншими модулями. Така модульна побудова забезпечує гнучкість використання системи для різних сценаріїв — від швидкого розділення мовців до комплексного дослідження групової динаміки з детальною аналітикою комунікативних ролей.

Audio Diarization. Цей компонент є спеціалізованим інструментом для виявлення та розділення мовців в аудіофайлах, призначеним для випадків, коли основною метою є ідентифікація учасників та визначення структури їхньої активності без необхідності повної транскрипції. Модуль автоматично визначає кількість унікальних голосів у записі та присвоює

кожному мовцю унікальний числовий ідентифікатор (Speaker 0, Speaker 1 тощо). Система генерує детальні часові мітки активності кожного учасника із зазначенням точного початку та завершення кожного сегменту мовлення, що дозволяє відстежувати динаміку участі протягом всієї дискусії. Компонент візуалізує розподіл часу мовлення у вигляді інтерактивних часових діаграм, кругових діаграм та гістограм, що надає наочне уявлення про баланс участі та домінування окремих учасників. Користувачі можуть переглядати статистику для кожного мовця окремо, включаючи загальний час мовлення, кількість реплік, середню тривалість висловлювань та частоту активності в різні періоди дискусії. Компонент також надає можливість ручного коригування результатів діаризації, що особливо корисно в складних акустичних умовах або при необхідності об'єднання сегментів одного мовця, які система могла помилково розділити.

Audio Transcription & Sentiment. Цей модуль поєднує функціональність автоматичної транскрипції аудіофайлів з можливістю цільового аналізу тональності за ключовими словами, що робить його оптимальним інструментом для дослідження специфічних тем у дискусіях. Система перетворює аудіо в текстовий формат із прив'язкою до часових міток, що дозволяє швидко навігувати до конкретних фрагментів обговорення. Ключовою особливістю модуля є підтримка пошуку специфічних термінів або фраз у транскрибованому тексті з автоматичним виділенням всіх згадувань та їх контексту. Для кожного знайденого ключового слова система проводить емоційну оцінку навколишнього текстового контексту, використовуючи лексичні словники тональності та алгоритми класифікації настроїв для визначення позитивного, негативного або нейтрального забарвлення висловлювань. Модуль генерує детальну статистику згадувань по учасниках, що включає кількість використань кожного терміна різними мовцями, розподіл згадувань у часі та переважну тональність обговорення теми кожним учасником. Користувачі можуть створювати списки ключових слів для моніторингу, що дозволяє автоматично відстежувати обговорення важливих тем, таких як назви проєктів, імена клієнтів, технічні терміни або проблемні питання. Система візуалізує динаміку згадувань у вигляді часових графіків, що показують інтенсивність обговорення певної теми протягом дискусії, а також надає можливість порівняння частоти використання різних термінів різними учасниками для виявлення розбіжностей у фокусі уваги або пріоритетах.

Full Audio/Video Analysis. Найбільш комплексний аналітичний модуль системи, що інтегрує всі попередні компоненти в єдиний workflow і додає розширені можливості глибинного аналізу комунікативної поведінки. Цей компонент поєднує діаризацію мовців, повну транскрипцію контенту та багатовимірний сентимент-аналіз для створення цілісної картини групової динаміки. Модуль автоматично розпізнає та розділяє мовців, транскрибує їхні висловлювання з прив'язкою до конкретних особистостей та часових міток, після чого виконує комплексний аналіз комунікативних патернів. Система ідентифікує комунікативні ролі учасників, визначаючи лідерів дискусії (які ініціюють нові теми та домінують за часом мовлення), активних учасників (які регулярно долучаються до обговорення), модераторів (які підтримують структуру розмови та надають слово іншим) та пасивних спостерігачів (з мінімальною участю). Аналізується стиль комунікації кожного учасника через оцінку складності їхньої мови, використання професійної термінології, лексичного різноманіття та структури аргументації. Модуль виявляє патерни взаємодії між учасниками, включаючи частоту звертань один до одного, випадки перебивань, підтримуючі або суперечливі реакції на висловлювання колег. Система відстежує емоційну динаміку дискусії в цілому та кожного учасника окремо, виявляючи моменти підвищення напруги, консенсусу або конфлікту. Результати представляються у вигляді багаторівневих інтерактивних звітів, що включають executive summary з ключовими висновками, детальну статистику участі, візуалізацію комунікативних мереж, часові діаграми емоційних станів та повну транскрипцію з можливістю фільтрації за мовцем, часом або темою.

Висновки та перспективи. Розроблена система комплексного аналізу комунікативної поведінки демонструє ефективність інтеграції сучасних технологій штучного інтелекту для автоматизації аналізу групових дискусій. Інтеграція різних аспектів комунікативної поведінки

в єдину аналітичну платформу забезпечує комплексний підхід та надає цілісне уявлення про групову динаміку та індивідуальну поведінку учасників. Система надає детальну аналітику комунікативних патернів, яка може використовуватися для покращення процесів прийняття рішень та підвищення ефективності зустрічей. Модульна архітектура дозволяє легко інтегрувати додаткові компоненти аналізу та адаптувати систему під специфічні потреби різних доменів застосування.

Незважаючи на позитивні результати, дослідження виявило кілька обмежень та викликів. Точність системи знижується в умовах значного фонового шуму або при накладенні голосів учасників, що впливає на якість діаризації та транскрипції. Оптимальна продуктивність досягається при кількості учасників від двох до п'яти; при більшій кількості учасників точність діаризації суттєво знижується через складність розділення голосів. Хоча система підтримує багато мов, найкращі результати досягаються для одномовного контенту, що пов'язано з особливостями навчання базових моделей розпізнавання мови.

Перспективи подальшого розвитку системи охоплюють декілька ключових напрямків. Інтеграція відео-аналізу дозволить враховувати невербальну комунікацію, жести та міміку учасників, що суттєво розширить можливості аналізу комунікативної поведінки. Розробка спеціалізованих алгоритмів для підвищення точності діаризації в умовах значного фонового шуму та при накладенні голосів покращить стабільність роботи системи в реальних умовах застосування. Створення спеціалізованих моделей для аналізу специфічних доменів, таких як медичні консультації, юридичні засідання та освітні семінари, забезпечить більш точний та релевантний аналіз у кожній галузі. Впровадження можливості аналізу живих дискусій з генерацією рекомендацій у реальному часі відкриє нові сценарії використання системи для підтримки модераторів та учасників під час проведення зустрічей.

Розроблена система відкриває нові можливості для дослідження групової динаміки та може знайти застосування в широкому спектрі галузей. У корпоративному секторі система може використовуватися для аналізу ефективності зустрічей, оптимізації командної роботи та підготовки менеджерів. В освітній сфері інструменти аналізу дозволять оцінювати участь студентів у семінарах та удосконалювати методики групового навчання. У сфері державного управління система надасть можливість здійснювати моніторинг громадських слухань, аналізувати публічні дебати та покращувати процеси консультацій з громадськістю. Впровадження таких систем сприятиме підвищенню ефективності колективної роботи, покращенню якості прийняття рішень та розвитку цифрових навичок у сфері аналізу комунікацій.

Список використаних джерел

1. Bredin, H., Laurent, A., Rouvier, M., Meignier, S., & Duponchel, L. (2020). Pyannote.audio: Neural building blocks for speaker diarization. In 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 7124–7128). IEEE. <https://doi.org/10.48550/arXiv.1911.01255>.
2. Bredin, H. (2023). pyannote.audio 2.1 speaker diarization pipeline: Principle, benchmark, and recipe. In Proc. Interspeech 2023. International Speech Communication Association. <https://doi.org/10.21437/Interspeech.2023-1294>.
3. Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In Proceedings of the 40th International Conference on Machine Learning (Vol. 202, pp. 28492–28518). PMLR. <https://proceedings.mlr.press/v202/radford23a.html>. <https://doi.org/10.48550/arXiv.2212.04356>.
4. Gulati, A., Qin, J., Chiu, C.-C., Parmar, N., Zhang, Y., Yu, J., Han, W., Wang, S., Zhang, Z., Wu, Y., & Pang, R. (2020). Conformer: Convolution-augmented Transformer for speech recognition. In Proceedings of Interspeech 2020 (pp. 5036–5040). ISCA. <https://doi.org/10.21437/Interspeech.2020-3015>.
5. Ao, J., Wang, R., Yang, L., Zhou, J., Liu, S., Wei, L., Qian, C., & Li, X. (2022). SpeechT5: Unified-modal encoder-decoder pre-training for spoken language processing. In Proceedings of

- the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 5723–5738). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.acl-long.393>.
6. Park, T. J., Zhang, N., Lu, X., Wu, Y., & Glass, J. (2021). AutoVC: Zero-shot voice style transfer with only autoencoder loss. In International Conference on Machine Learning (pp. 8291–8300). PMLR. <https://proceedings.mlr.press/v139/park21b.html>. <https://doi.org/10.48550/arXiv.1905.05879>.

Rudensky Roman

Doctor of Economic Sciences, Professor, Professor of the Department of Computer Science, National University of Life and Environmental Sciences of Ukraine

ORCID: <https://orcid.org/0009-0002-3682-9702>

E-mail: roman.rudensky@nubip.edu.ua

Kravchenko Volodymyr

Doctor of Economic Sciences, Associate Professor, Professor of the Department of Economic Cybernetics,

National University of Life and Environmental Sciences of Ukraine

ORCID: <https://orcid.org/0000-0002-8033-3985>

E-mail: v.kravchenko@nubip.edu.ua

A SYSTEM FOR COMPREHENSIVE ANALYSIS OF COMMUNICATIVE BEHAVIOR IN PUBLIC DISCUSSIONS BASED ON SPEAKER DIARIZATION

Abstract. *The relevance of this study is driven by the increasing volume of online meetings and public discussions in digital formats, creating a demand for automated tools to analyze group communication. Traditional manual coding and transcription methods are highly labor-intensive and subjective, limiting large-scale research on communication patterns. The aim of this research is to develop and validate a comprehensive system for automated analysis of communicative behavior that integrates modern speaker diarization technologies, automatic speech recognition, and statistical analysis to provide a detailed picture of group dynamics in public discussions. Methods. The system is implemented using a microservice architecture with Python 3.10+, FastAPI, and React. Speaker diarization is performed using the pyannote.audio algorithm, which combines convolutional encoders with pre-trained WavLM models. Automatic speech recognition is carried out using transformer architectures (Whisper, AssemblyAI, Conformer). Communicative behavior analysis includes calculation of activity statistics, network analysis of interactions, and assessment of communication style. Results. The developed system successfully integrates speaker diarization with 0.5-second precision, automatic transcription, and multidimensional analysis of communication patterns. The modular architecture ensures flexibility for adaptation to various application domains. The system generates detailed timestamps of participant activity, visualizes speaking time distribution, and provides comprehensive analytics to improve decision-making processes. Prospects. Further development of the system includes integration of multimodal analysis considering non-verbal communication, improvement of stability in noisy conditions, domain adaptation for specific sectors, and implementation of real-time analysis of live discussions. The system opens new opportunities for studying group dynamics in corporate, educational, and governmental sectors.*

Keywords: *Speaker Diarization, Automatic Speech Recognition, Analysis of Communicative Behavior, Artificial Intelligence, Public Discussions.*