Хмельницький національний університет
Факультет інформаційних технологій
Кафедра комп'ютерних наук

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

на тему *Метод автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами*

Галузь знань *12 – Інформаційні технології*
Шифр і назва галузі знань

Спеціальність *122 – Комп'ютерні науки*
Шифр і назва спеціальності

Освітня програма *Комп'ютерні науки*
Назва освітньої програми

Виконав: *студент групи КН-21-2* (Група виконавця) _____ (Підпис) *Юрій АНТОНЮК* (Ім'я, ПРІЗВИЩЕ)

Керівник: *к.т.н., доц. каф. КН* (Науковий ступінь, посада) _____ (Підпис) *Руслан БАГРІЙ* (Ім'я, ПРІЗВИЩЕ)

Нормоконтроль: *к.т.н., доц. каф. КН* (Науковий ступінь, посада) _____ (Підпис) *Руслан БАГРІЙ* (Ім'я, ПРІЗВИЩЕ)

До захисту допускаю:
Зав. кафедри КН, д.т.н., професор _____ (Підпис) *Олександр БАРМАК* (Ім'я, ПРІЗВИЩЕ)
19 *06* 2025 р.

Хмельницький 2025

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет *інформаційних технологій*
Кафедра *комп'ютерних наук*
Освітній ступінь *бакалавр*
Галузь знань *12 – Інформаційні технології*
Спеціальність *122 – Комп'ютерні науки*

ЗАТВЕРДЖУЮ
Завідувач кафедри *комп'ютерних наук*

_____
(підпис)
д.т.н., професор Олександр БАРМАК
«*10*» *02* *2025* року

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА

1. Тема кваліфікаційної роботи бакалавра: *«Метод автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами.»*

2. Завдання видано студенту *Юрію АНТОНЮКУ*
(Ім'я, прізвище)

3. Керівник роботи *доцент кафедри КН Руслан БАГРІЙ*
(посада, ім'я, прізвище)

4. Затверджено наказом університету від «*07*» *02* *2025* р. № *23*

5. Дата видачі завдання студенту: «*10*» *02* *2025* р.

6. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Мета кваліфікаційної роботи бакалавра – підвищення релевантності та точності автоматичного резюмування навчальних відеоматеріалів засобами глибокого навчання. Для досягнення мети становлено наступний перелік завдань: дослідити сучасні методи та технології обробки мовлення та автоматичного резюмування тексту; розробити метод автоматичного резюмування навчальних відеоматеріалів з використанням нейромережевих засобів; створити програмну реалізацію методу автоматичного резюмування навчальних відеоматеріалів для обробки аудіо- та текстових даних; оцінити ефективність методу автоматичного резюмування.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

№	Назва етапів (розділів) кваліфікаційної роботи бакалавра	Термін виконання	Примітка
1	*Вибір напряму дослідження та узгодження тематики кваліфікаційної роботи бакалавра з керівником, складання календарного графіка виконання*	*січень 2025*	*Виконано*
2	*Ознайомлення з предметною областю, формулювання мети і задач дослідження, визначення об'єкта та предмета дослідження*	*лютий 2025*	*Виконано*
3	*Проєктування та розроблення методу вирішення завдання, загальної архітектури програмного забезпечення, інтерфейсу користувача, вибір засобів реалізації програмного забезпечення*	*березень 2025*	*Виконано*
4	*Створення та тестування програмного забезпечення, дослідження ефективності, висновки з виконаної роботи*	*квітень 2025*	*Виконано*
5	*Написання пояснювальної записки, урахування зауважень керівника, оформлення згідно з вимогами*	*травень 2025*	*Виконано*
6	*Розробка презентаційних матеріалів та попередній захист кваліфікаційної роботи*	*травень 2025*	*Виконано*
7	*Отримання відгуку керівника, рецензії, перевірка на плагіат, нормоконтроль*	*червень 2025*	*Виконано*
8	*Підготовка до захисту та захист кваліфікаційної роботи*	*червень 2025*	*Виконано*

Виконавець: *студент групи КН-21-2* (Група виконавця) _____ (Підпис) *Юрій АНТОНЮК* (Ім'я, ПРІЗВИЩЕ)

Керівник: *к.т.н., доц. каф. КН* (Науковий ступінь, посада) _____ (Підпис) *Руслан БАГРІЙ* (Ім'я, ПРІЗВИЩЕ)

Анотація

Тема кваліфікаційної роботи бакалавра: *«Метод автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами»*

Виконавець кваліфікаційної роботи бакалавра: *студент групи КН-21-2 Юрій Антонюк*

Керівник кваліфікаційної роботи бакалавра: *к.т.н., доцент кафедри КН Руслан БАГРІЙ*

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
51	23	2	55	2

Метою кваліфікаційної роботи бакалавра є підвищення релевантності та точності автоматичного резюмування навчальних відеоматеріалів за допомогою методів глибокого навчання. Розроблений підхід ґрунтується на використанні ASR моделі Whisper для розпізнавання мовлення та LLM моделі BART для створення стислих резюме.

Головним результатом є реалізований метод, що забезпечує ефективне отримання коротких резюме з освітніх відеоматеріалів різних форматів та змісту. Він може бути використаний для швидкого ознайомлення студентів з ключовими аспектами лекцій, спрощення пошуку інформації у великих обсягах відеоконтенту або для автоматичного створення конспектів, значно підвищуючи ефективність освітнього процесу.

Ключові слова: ASR, LLM, Whisper, BART, модель, навчальні відеоматеріали, метод, штучний інтелект.

Виконавець: *студент групи КН-21-2* _____ *Юрій АНТОНЮК*

Група виконавця | Підпис | Ім'я, ПРІЗВИЩЕ

Зміст

Перелік скорочень

Скорочення, термін, позначення	Пояснення
ASR	Автоматичне розпізнавання мовлення
HMM	Прихована марковська модель
DNN	Глибокі нейронні мережі
ШІ	Штучний інтелект
LLM	Велика мовна модель
RNN	Рекурентна нейронна мережа
NLP	Нейролінгвістичне програмування
PCM	Формат цифрового кодування аналогових сигналів
WER	Коефіцієнт помилок слів

Вступ

Метою кваліфікаційної роботи бакалавра є створення методу автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами.

В епоху стрімкого розвитку цифрової освіти та масового поширення онлайн-курсів, вебінарів і відеолекцій, обсяги навчальних відеоматеріалів сильно зростають. Це створює значне навантаження на здобувачів освіти та викладачів, яким доводиться витрачати велику кількість часу на перегляд повного контенту для виявлення ключових концепцій та інформації. Використання методу автоматичного резюмування навчальних відеоматеріалів дозволяє суттєво оптимізувати процес засвоєння знань, забезпечуючи швидкий доступ до суті лекцій та тренінгів. Розробка такого методу сприятиме підвищенню доступності освітнього контенту та загальній оптимізації освітнього процесу, що є невід'ємною частиною модернізації сучасної системи освіти.

Об’єкт дослідження – процес автоматичного резюмування навчальних відеоматеріалів.

Предмет дослідження – методи глибокого навчання та обробки природної мови для автоматичного резюмування навчальних відеоматеріалів.

Мета кваліфікаційної роботи бакалавра – підвищення релевантності та точності автоматичного резюмування навчальних відеоматеріалів засобами глибокого навчання.

Завдання кваліфікаційної роботи бакалавра: дослідити сучасні методи та технології обробки мовлення та автоматичного резюмування тексту; розробити метод автоматичного резюмування навчальних відеоматеріалів з використанням нейромережевих засобів; створити програмну реалізацію методу автоматичного резюмування навчальних відеоматеріалів для обробки аудіо- та текстових даних; оцінити ефективність методу автоматичного резюмування.

Розділ 1 Огляд технологій штучного інтелекту для резюмування навчальних відеоматеріалів

1.1 Автоматизоване резюмування відеоматеріалів

Щороку генерується безліч цінної інформації різного формату що потребує обробки. Однак узагальнювати, класифікувати і форматувати дані вручну є дуже клопіткою і енергозатратною роботою. Тому автоматичне резюмування тексту може стати чудовим рішенням. З використанням такої технології людина делегуватиме тяжку роботу машині. Таким чином можна зменшити витрати коштів та скоротити строки резюмування, без ризиків упустити якусь важливу інформацію.

Однак серед усього різноманіття форматів даних, однозначно домінує відео. З кожним роком відеоконтент захоплює все більшу частку світового інтернет трафіку. Відеоконтент більше не є просто можливістю; це необхідність. Згідно з нещодавнім дослідженням Cisco, станом на 2024 рік відео становить 82% усього інтернет-трафіку споживачів [1].

Відео можна класифікувати за призначенням: освітні, розважальні, рекламні, інформаційні та користувацькі [2].

Навчальні відео стали важливою частиною вищої освіти, забезпечуючи важливий інструмент доставки контенту в багатьох перевернутих, змішаних і онлайн-класах. Ефективне використання відео як навчального інструменту підвищується, коли викладачі враховують три елементи: як керувати когнітивним навантаженням відео; як максимізувати взаємодію студентів з відео; і як сприяти активному навчанню за допомогою відео. У цьому есе розглядається література, що стосується кожного з цих принципів, і пропонуються практичні способи, як викладачі можуть використовувати ці принципи під час використання відео як навчального інструменту.

Але яким би корисним відео не було для навчання, дуже часто буває що для всього перегляду відео часу може не вистачати. З усього відео, потрібна

частина займає всього десяту частину від всієї наданої в ньому інформації. Тоді в нагоді може стати використання засобів резюмування тексту.

Резюмування тексту – це метод стиснення оригінального тексту для формування резюме, яке надасть той самий зміст та інформацію, що й оригінальна версія. Це допомагає заощадити час і підвищити ефективність роботи.

Узагальнення тексту допомагає створювати зведення, заголовки, резюме, короткий опис, знаходити ключові слова, тощо. Інші приклади – заголовки новин в газеті, назви книг і багато іншого. Узагальнення тексту необхідне, тому що величезне зростання обсягу інформації вимагає високого рівня обслуговування [3].

В резюмуванні можна виділити два відмінні підходи для вирішення відповідних завдань.

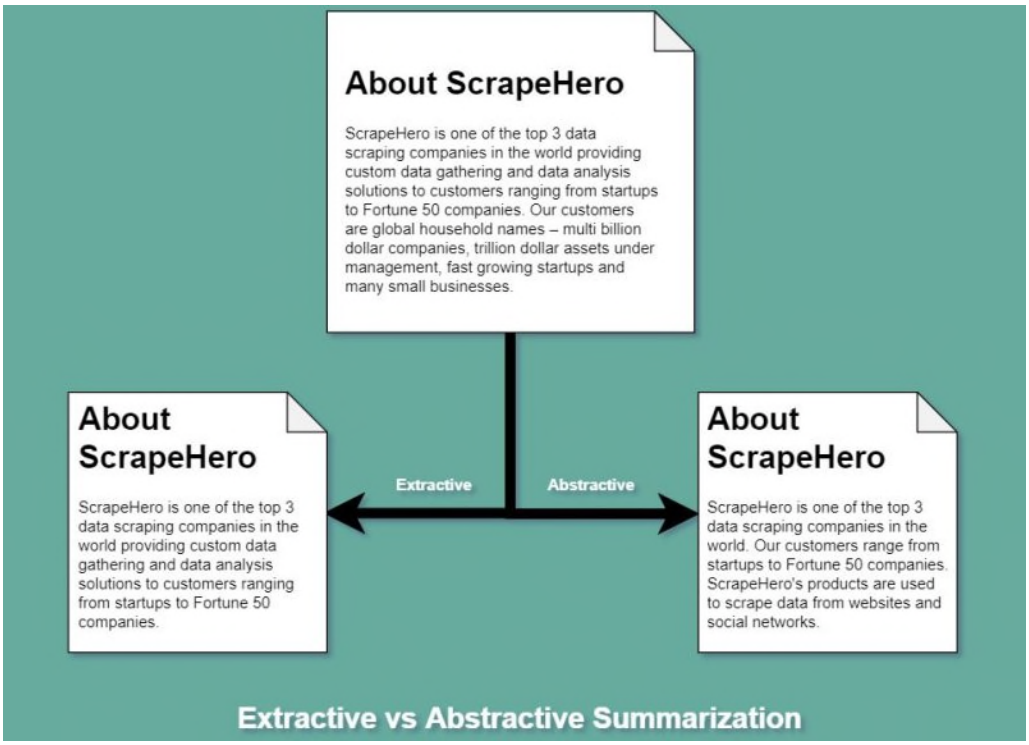


Рисунок 1.1 – Відмінності екстрактивного та абстрактного резюмування [4]

Екстрактивне реферування – це створення реферату, що містить підмножину речень оригінального тексту, після визначення важливих речень [5].

Абстрактне резюмування тексту – це завдання створення короткого і стислого резюме, яке відображає основні ідеї вихідного тексту. Згенеровані

резюме можуть містити нові фрази та речення, які можуть не з'являтися у вихідному тексті [6].

Далеко не для кожного відео можна застосувати резюмування, бо воно вимагає текстового файлу. Тому для резюмування відео без задокументованих даних потрібно комбінувати резюмування тексту з автоматичним розпізнаванням мовлення.

Розпізнавання мовлення – це технологія, що дозволяє машинам обробляти розмовну мову як послідовність акустичних сигналів, щоб вони могли інтерпретувати значення, контекст і намір у вихідний текст. Простіше кажучи, це технологія, яка перекладає або перетворює мову на текст [7].

Технологія розпізнавання мовлення існує з 1952 року, коли сумнозвісна компанія Bell Labs створила «Audrey», розпізнавач цифр. Спочатку Audrey можна було використовувати лише для транскрипції вимовлених чисел, але через десять років дослідники змогли змусити його транскрибувати рудиментарні розмовні слова, як-от «привіт».

Пізніше дослідники використовували класичні технології машинного навчання, такі як приховані марковські моделі, для посилення моделей розпізнавання мовлення, хоча згодом точність таких класичних моделей знизилася на фоні стрімкого розвитку інших сучасних підходів до виконання такого роду завдань [8].

Підбиваючи підсумки, можна сказати, що зважаючи на зростаючу в мережі кількість відеоматеріалів потреба в якісному ресурсі з обробки буде лише зростати.

1.2 Огляд технологій штучного інтелекту для автоматичного розпізнавання мовлення та резюмування.

Розпізнавання мовлення, змінило взаємодію людей з комп’ютерними пристроями, адже – це технологія, яка розуміє і може виконувати усні команди. Ця чудова інновація полегшила застосування, сприяючи підвищенню

продуктивності в різних галузях, таких як охорона здоров'я, обслуговування клієнтів і телекомунікації.

Розпізнавання мовлення не є універсальним рішенням, бо має безліч нюансів, і його типи різняться залежно від багатьох функціональних можливостей. Ці функції включають ідентифікацію мовлення та системи розпізнавання дикторів [9].

Програмне забезпечення для розпізнавання мовлення має адаптуватися до дуже мінливої та контекстно-залежної природи людського мовлення. Алгоритми програмного забезпечення, які обробляють і організовують аудіо в текст, тренуються на різних мовних моделях, стилях мовлення, мовах, діалектах, акцентах і фразах. Програмне забезпечення також відокремлює розмовне аудіо від фонового шуму, який часто супроводжує сигнал.

Щоб задовольнити ці вимоги, системи розпізнавання мови ASR використовують два типи моделей: акустичні моделі, що відображають зв'язок між лінгвістичними одиницями мови і звуковими сигналами, а також мовні які зіставляють звуки з послідовностями слів, щоб розрізняти слова, які звучать схоже [10].

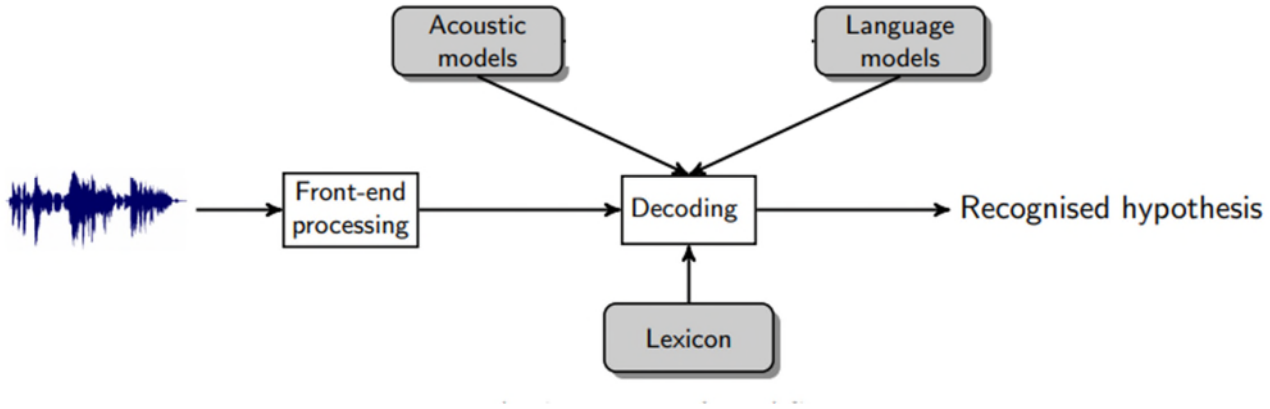


Рисунок 1.2 – Архітектура ASR системи [11]

Акустичне моделювання зазвичай відноситься до процесу встановлення статистичних представлень для послідовностей векторів ознак, обчислених з форми мовного сигналу, а також включає в себе «моделювання вимови», яке описує, як послідовність або кілька послідовностей основних мовних одиниць

(наприклад, фони або фонетичних ознак) використовуються для представлення більших мовних одиниць, таких як слова або фрази, які є об'єктом розпізнавання мовлення. Акустичне моделювання може також включати використання інформації зворотного зв'язку від розпізнавача для зміни форми векторів ознак мови для досягнення завадостійкості при розпізнаванні мови. [12].

Найпоширеніші типи акустичних моделей включають приховані марковські моделі (HMM) і глибокі нейронні мережі (DNN). HMM були основою традиційних систем розпізнавання мовлення, тоді як DNN представляють собою авангард сучасних досягнень, пропонуючи неперевершену точність і можливості навчання. Обидві моделі мають свої сильні сторони, але перехід до глибокого навчання відображає постійний розвиток галузі [13].

Приховані марковські моделі, скорочено відомі як HMM, – це статистичні моделі, що описують еволюцію спостережуваних подій, які самі по собі залежать від внутрішніх факторів, які не можна безпосередньо спостерігати. Прихована модель Маркова складається з двох різних стохастичних процесів, які можна визначити як послідовності випадкових величин (змінних), що залежать від випадкових подій [14].

Глибока нейронна мережа (DNN) – це тип моделі машинного навчання, який імітує те, як людський мозок. На відміну від традиційних алгоритмів, які дотримуються заздалегідь визначених правил, DNN можуть слідувати заданим шаблонам і робити прогнози на основі попереднього досвіду. DNN є основою глибокого навчання, що забезпечує такі програми, як агенти ШІ, розпізнавання зображень, голосові помічники, чат-боти ШІ. Вона складається з кількох шарів вузлів, які отримують вхідні дані від інших рівнів і створюють вихідні дані до досягнення кінцевого результату [15].

Мовна модель – використовує машинне навчання для обчислення розподілу ймовірностей за словами. Мовні моделі вивчають текст і можуть використовуватися для створення оригінального тексту, передбачення наступного слова в тексті, розпізнавання мовлення, оптичного розпізнавання символів і розпізнавання рукописного тексту [16]. Серед мовних моделей можна

виділити два основних типи: статистичні (імовірнісні) моделі, та моделі на основі нейронних мереж.

Статистичні моделі – це моделі, які можуть передбачити наступне слово в послідовності, враховуючи слова, які йому передують. Статистична модель мови вивчає ймовірність появи слова на прикладах тексту. Простіші моделі можуть розглядати контекст короткої послідовності слів, тоді як більші моделі можуть працювати на рівні речень або абзаців. Найчастіше діють на рівні слів. Ці моделі зазвичай використовуються на в середині більш складної моделі для завдання, яке вимагає розуміння мови. Статистичні моделі можна поділити на типи:

– N-грами: модель суть якої полягає в тому, що замість того, щоб обчислювати ймовірність слова, враховуючи всю його історію, ми можемо приблизно оцінити історію лише за кількома останніми словами. [17].

– Експоненціальні (Моделі максимальної ентропії): клас статистичних моделей, які забезпечують підхід до вивчення даних шляхом максимізації ентропії основного розподілу ймовірностей [18].

Резюмування тексту з використанням штучного інтелекту передбачає використання технологій для стиснення великих обсягів тексту, аудіо чи відеоданих у більш керовану та узгоджену форму. Цей процес зберігає основну інформацію або теми, забезпечуючи легше розуміння та швидке засвоєння суттєвих матеріалів. Технологія покладається на алгоритми машинного навчання для визначення ключових елементів і шаблонів у даних [19].

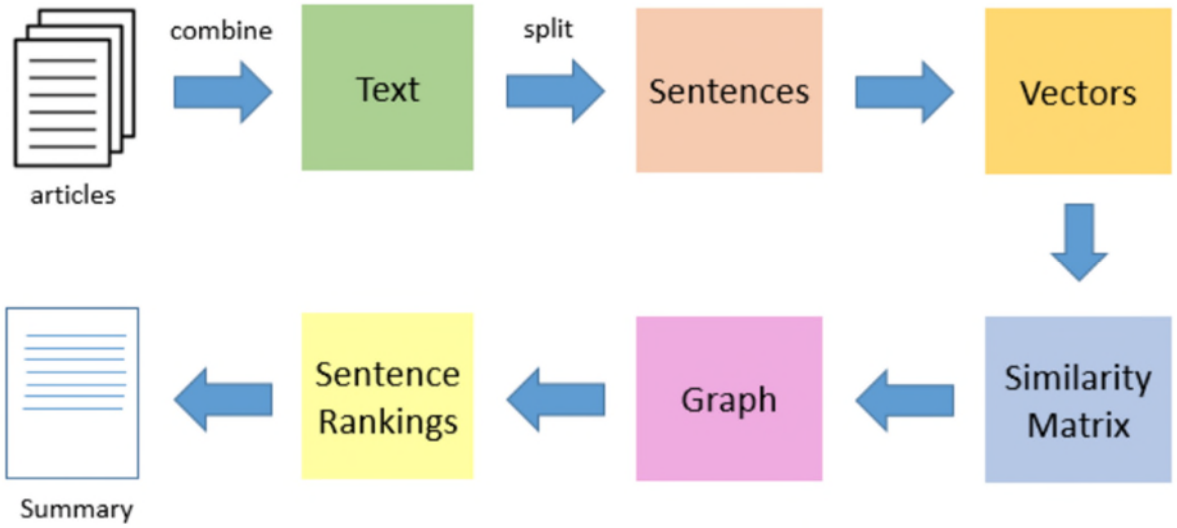


Рисунок 1.3 – Базовий алгоритм резюмування тексту [20]

В резюмування тексту явно домінують такі архітектурні моделі як: Sequence-to-Sequence (Seq2Seq), трансформери та LLM.

Моделі послідовності (Seq2Seq) – це тип архітектури нейронної мережі, який використовується в обробці природної мови (NLP) і завданнях машинного перекладу. Вони складаються з мережі кодера та декодера. Кодер приймає вхідну послідовність і перетворює її на контекстний вектор фіксованого розміру, фіксуючи інформацію вхідної послідовності. Потім декодер генерує вихідну послідовність на основі цього вектора контексту [21].

Модель трансформера – це нейронна мережа, яка вивчає контекст відстежуючи зв’язки в послідовних даних [22]. Головною особливістю моделей трансформерів є їхній механізм самоуваги, завдяки якому моделі отримують здатність виявляти зв’язки (або залежності) між кожною частиною вхідної послідовності [23].

Велика мовна модель (LLM) – це передова система штучного інтелекту, яка розуміє, обробляє та створює вміст, подібний до людини. LLM навчаються на величезний кількості великі дані, що дозволяє їм розпізнавати складні шаблони та структури для створення узгоджених, контекстуально релевантних відповідей на широкий спектр завдань [24].

Наведені вище технології є провідними для вирішення завдань що стосуються обробки мовлення чи роботи з текстом. Використовуючи такий набір технологій можна бути певним про якість та надійність архітектури, оскільки перераховані технологій пройшли перевірку часом, зайняли ключові позиції та мають велику базу інформації для коректної роботи з ними.

1.3 Огляд існуючих засобів для трансформації аудіо в текст та резюмування тексту

Аналіз конкуренції на ринку також грає важливу роль в досягненні успіху. Лише проаналізувавши продукти конкурентів, можна отримати повну

картину затребуваності в тому чи іншому продукті чи його функціях. І на основі отриманих даних можна скласти план розробки майбутнього продукту.

Otter.ai – дозволяє користувачам записувати та транскрибувати розмови та зустрічі на мобільному телефоні (iOS та Android) або у веб-браузері (через розширення Chrome) [25].

На наступному скріншоті відображено приклад запису розмови з використанням комп'ютера (Рисунок 1.4).

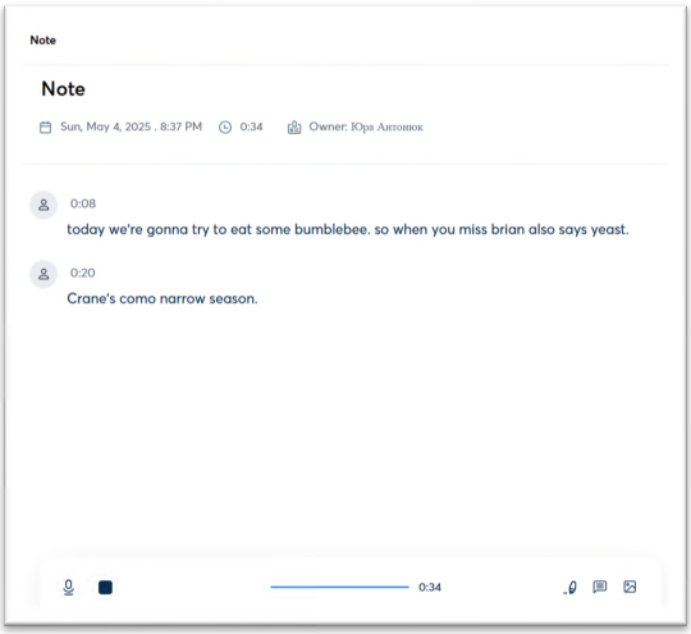


Рисунок 1.4 – Приклад запису розмови в програмі Otter.ai

Speechnotes – інструмент що дозволяє транскрибувати відео та аудіо в текст. Має доволі широкий спектр використання [26].

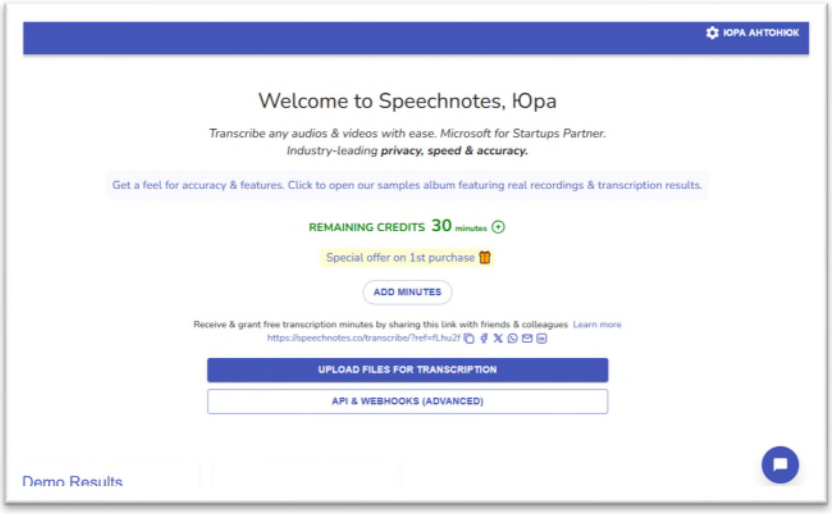


Рисунок 1.5 – Скріншот інтерфейсу speechnotes

Типовою програмою для резюмування тексту є SSMRY – веб-ресурс що резюмує текст. Також може працювати з завантаженими файлами чи посиланнями на ресурси [27].

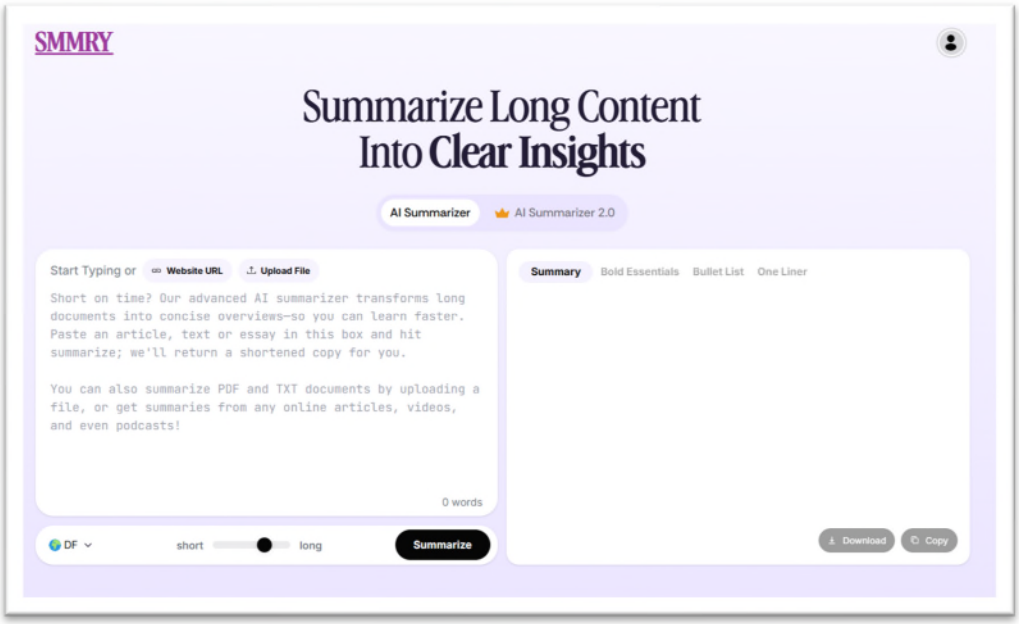


Рисунок 1.6 – Скріншот інтерфейсу SSMRY

Також розглянемо продукт Quillbot AI – продукт, що має великий спектр функцій, таких як перевірка граматики, гуманізація тексту, детектор штучного інтелекту, переклад, генератор цитат та інші [28].

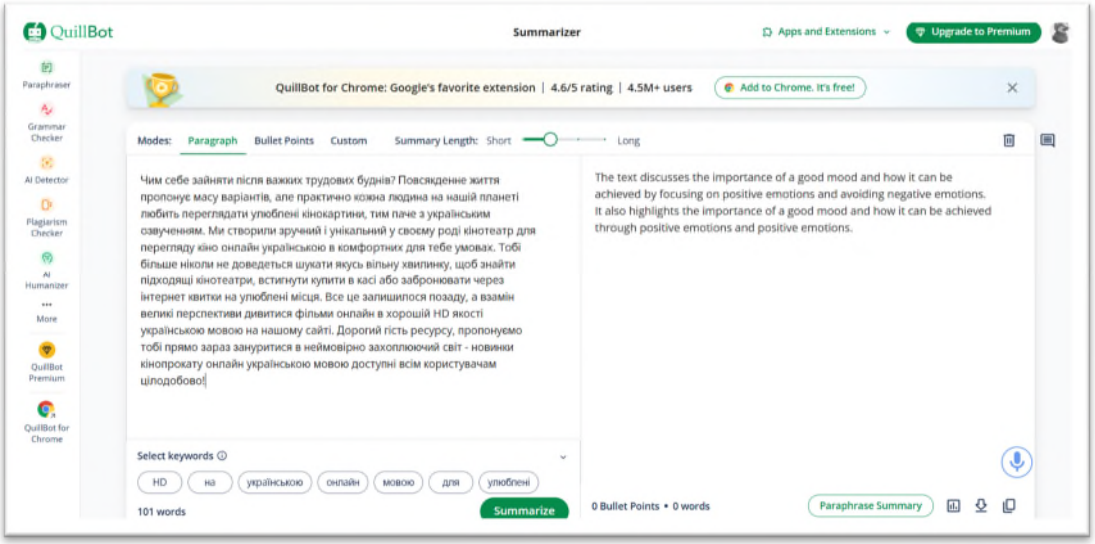


Рисунок 1.8 – Приклад резюмування тексту

Zoom надає можливості використання помічника на базі штучного інтелекту Zoom AI Companion. Він використовується для транскрипції в режимі реального часу та для створення коротких резюме по онлайн зустрічах [29].

Проаналізувавши доступні в інтернеті ресурси було виявлено, що усі ці ресурси є популярними та активно підтримуються, що вказує на те, що такого роду програмні продукти є популярними та затребуваними на ринку. І на основі користувацького досвіду було занотовано усі успішні аспекти реалізації вищезгаданих продуктів, для того, щоб реалізований продукт відповідав усім сучасним вимогам на ринку.

1.4 Мета та завдання кваліфікаційної роботи

Мета кваліфікаційної роботи бакалавра – підвищення релевантності та точності автоматичного резюмування навчальних відеоматеріалів засобами глибокого навчання.

Для досягнення поставленої мети було визначено наступні цілі:

– дослідити сучасні методи та технології обробки мовлення та автоматичного резюмування тексту;

– розробити метод автоматичного резюмування навчальних відеоматеріалів з використанням нейромережевих засобів;

– створити програмну реалізацію методу автоматичного резюмування навчальних відеоматеріалів для обробки аудіо- та текстових даних;

– оцінити ефективність методу автоматичного резюмування.

Розділ 2 Проєктування методу автоматичного резюмування відеоматеріалів

2.1 Загальна ідея методу автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами.

Автоматичне резюмування навчальних відеоматеріалів є складним завданням, що поєднує обробку мультимодальних даних, аналіз мовлення та текстову компресію з використанням нейромережевих технологій. Метою цього процесу є створення стислих і змістовно релевантних узагальнень відеоконтенту, які зберігають ключові ідеї та полегшують засвоєння інформації. Загальна ідея методу автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами зображена на рисунку 2.1.

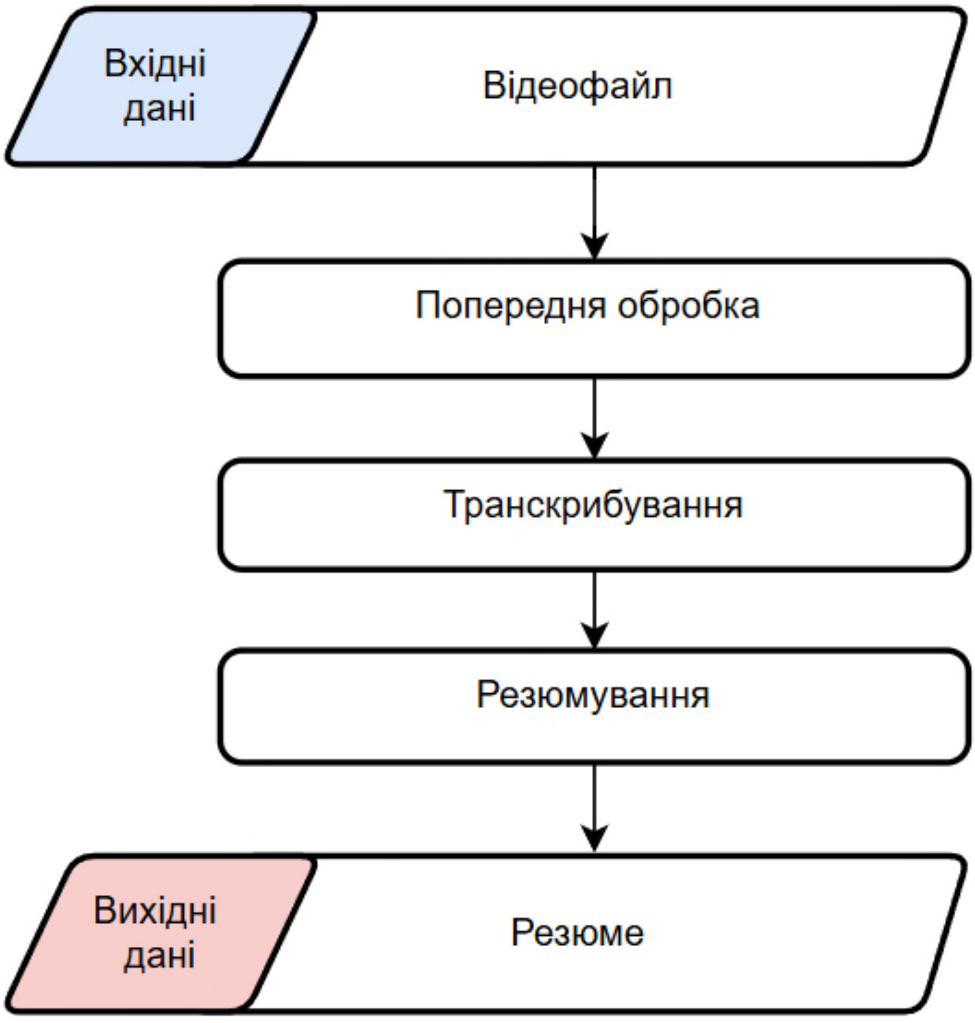


Рисунок 2.1 – Загальна ідея автоматичного резюмування навчальних відеоматеріалі

Як початкові дані для роботи методу подається відеофайл різних форматів.

Етап попередньої обробки готує аудіодані до наступного етапу (транскрибування), оптимізуючи їхній формат і структуру для ефективної роботи з нейромережевими засобами. Результатом є аудіофайл, збережений за шляхом, який слугує основою для подальшого аналізу.

Коли аудіофайл отримано, запускається другий етап. На цьому етапі відбувається транскрибування вхідних даних і результатом виконання етапу є створений текстовий файл, що містить розпізнані слова з переданих вхідних даних.

Процес резюмування тексту базується на систематичному підході, який включає кілька етапів для забезпечення ефективного стиснення змісту. Спочатку вхідний текст розбивається на менші сегменти, щоб полегшити обробку та уникнути перевантаження системи, при цьому враховується оптимальна довжина кожного фрагмента для збереження смислової цілісності. Далі кожен сегмент аналізується з використанням спеціалізованих алгоритмів, які виділяють ключові ідеї та основні положення, формуючи стислі узагальнення. Якщо текст є надто об’ємним, процес обробки відбувається послідовно для всіх частин, після чого отримані результати об’єднуються в єдине резюме. Результатом виконання етапу є сформоване резюме.

По завершенні всіх попередніх етапів користувачу стає доступний текстовий файл-резюме, що в повній мірі передає зменшений в обсязі контент оригінального файлу, та числова оцінка якості розпізнавання мовлення та резюмування тексту.

2.2 Архітектура ASR моделей

Сучасні моделі ASR використовують революційний підхід до обробки мовлення з наскрізним глибоким навчанням. Один із перших підходів до

побудови наскрізної системи був представлений у 2014 році дослідниками Алексом Грейвсом з Google DeepMind і Навдіпом Джайтлі з Університету Торонто.

Складна нейронна мережа в сучасній моделі ASR замінює багатоетапні моделі в застарілих системах, що мінімізує затримку та суттєво покращує продуктивність і точність. Ця архітектура також позбавляється незалежних мовних, акустичних та лексиконних моделей, а отримана сучасна система ASR функціонує як єдина нейронна мережа [30].

Базова архітектура яка використовується в сучасних ASR системах зображена на рисунку 2.2.

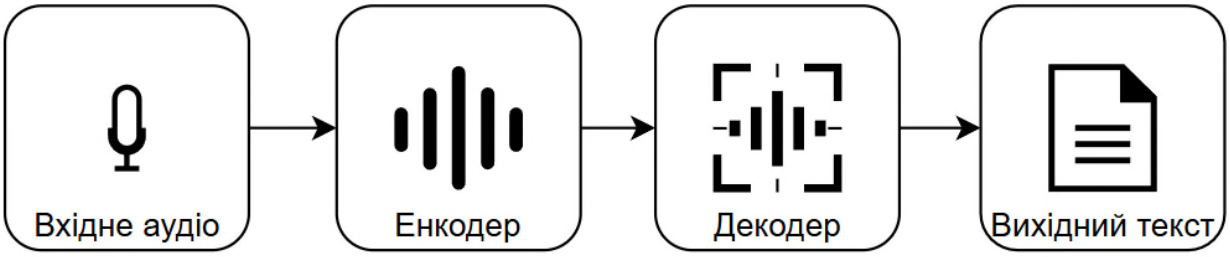


Рисунок 2.2 – Схема роботи сучасної ASR архітектури.

Яскравим прикладом сучасної ASR системи є модель Whisper від OpenAI, що складається з двох основних компонентів (енкодера та декодера), які приймають послідовність на вхід та генерують відповідну послідовність на виході.

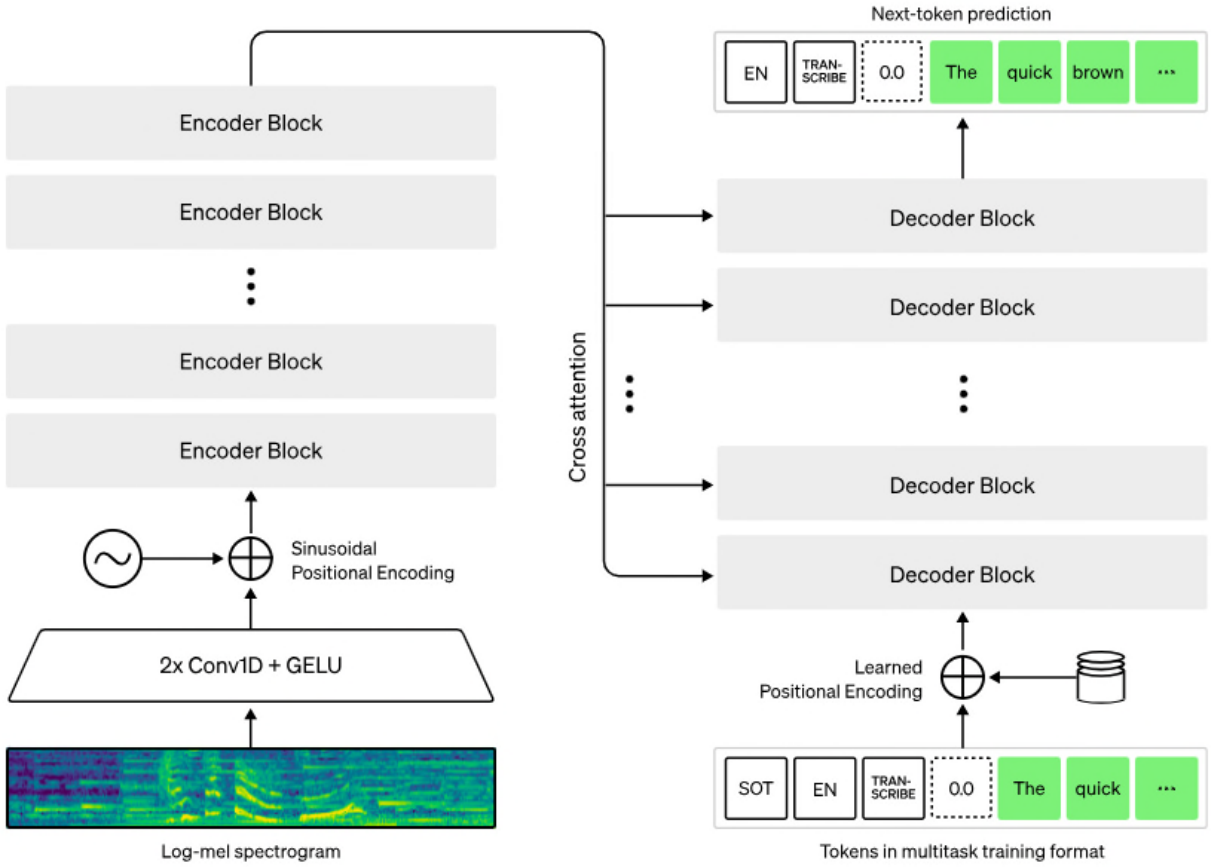


Рисунок 2.3 – Архітектура моделі Whisper [31]

Енкодер: Енкодер використовує нейронні мережі для обробки вхідної послідовності та створення векторного представлення фіксованого розміру. Енкодер також захоплює контекст із вхідної послідовності та передає його декодеру.

Декодер: Модуль «декодера» приймає вихідний вектор кодера та створює вихідну послідовність. Декодер передбачає наступні токени вихідної послідовності на основі отриманого контексту та власних попередніх прогнозів.

Багато моделей ASR використовують наскрізну архітектуру трансформерів для розуміння контексту та значення вхідного аудіо. Модель спочатку розділяє вхідне аудіо на невеликі фрагменти, перш ніж передати їх кодеру. Декодер передбачає текстові підписи.

Також для енкодера та декодера можливий варіант використання рекурентних нейронних мереж (RNN), специфічний тип нейронної мережі, що

добре підходить для прогнозування послідовностей. Комірки RNN можуть запам'ятовувати інформацію про раніше побачені елементи послідовності через свою внутрішню пам'ять і використовувати її для визначення вихідного результату. На відміну від трансформерів, RNN є послідовними моделями.

Традиційні застарілі системи перетворення мовлення в текст використовують комбінацію акустичних, лексиконових і мовних моделей для прогнозування тексту. Але вони мають суттєві обмеження в точності та потребують спеціальних знань для фонетичного навчання.

Сучасні моделі ASR стають дедалі складнішими та можуть обробляти різні типи введення, включно з кількома мовами. Спрощена наскрізна модель глибокого навчання забезпечує мінімальну затримку, виняткову точність і високу масштабованість.

2.3 Архітектура LLM моделей

LLM – це спеціально розроблена підмножина машинного навчання, відомого як глибоке навчання, яке використовує алгоритми, навчені на великих наборах даних, щоб розпізнавати складні шаблони. LLM навчаються на величезній кількості тексту. На базовому рівні вони вчаться відповідати на запити користувачів релевантним, контекстним вмістом, написаним людською мовою – тим типом слів і синтаксису, які люди використовують під час звичайної розмови [32].

Для глибшого розуміння внутрішньої роботи LLM моделей, зокрема їхньої архітектури, розглянуто одну з таких моделей – BART (Bidirectional Encoder and Autoregressive Transformer). Спрощена архітектура цієї моделі зображена на рисунку 2.4.

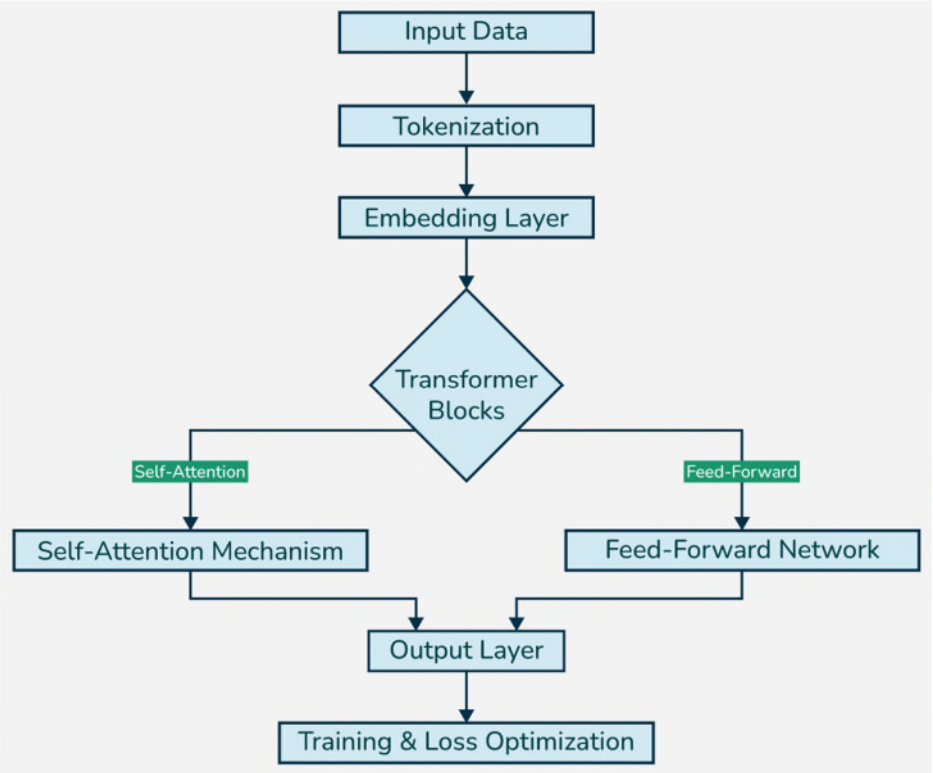


Рисунок 2.4 – Архітектура великої мовної моделі BART [33]

Токенізація є першим і останнім кроком обробки та моделювання тексту. Токенізація розбиває текст на лексеми, і кожному токену присвоюється числове представлення або індекс, який можна використовувати для введення в модель. У типовому робочому процесі LLM спочатку кодується введений текст у токени за допомогою токенізатора. Кожному унікальному токену присвоюється певний індексний номер у словнику токенізатора. Після того, як текст токенізовано, ці маркери пропускаються через модель, яка зазвичай включає шар вбудовування та блоки трансформера. Рівень вбудовування перетворює токени в щільні вектори, які фіксують семантичні значення. Потім блоки трансформатора обробляють ці вектори вбудовування, щоб зрозуміти контекст і створити результати. Останнім кроком є декодування, яке детокенізує вихідні маркери назад у текст, який читається людиною. Це робиться шляхом зіставлення токенів із відповідними словами за допомогою словника токенізатора [34].

Рівень вбудовування є першим шаром у цих моделях, відповідальним за перетворення необроблених текстових даних у щільні векторні представлення, які можуть бути оброблені наступними шарами. Рівень вбудовування в LLM

зазвичай реалізується як таблиця пошуку, де кожне слово або символ у словнику пов’язується з унікальним вектором. Вектори ініціалізуються випадковим чином на початку навчання, а потім ітеративно оновлюються, коли модель вивчає дані [35].

Моделі LLM зазвичай базуються на архітектурі трансформера та складаються з кількох рівнів нейронних мереж, кожна з яких має параметри, які можна точно налаштувати під час навчання, які додатково посилюються численним рівнем, відомим як механізм уваги, який звертається до певних частин наборів даних [36].

Механізм уваги – це техніка, яка використовується в штучному інтелекті (ШІ) та машинному навчанні, яка імітує людську когнітивну увагу. Це дозволяє моделі вибірково концентруватися на найбільш значущих частинах вхідних даних – наприклад, на конкретних словах у реченні або області на зображенні, при складанні прогнозів чи генерації вихідних даних. Замість того, щоб однаково ставитись до всіх вхідних даних, така виборча концентрація покращує продуктивність, особливо при роботі з великими обсягами інформації, такими як довгі текстові послідовності або зображення високої роздільної здатності. Це дозволяє моделям ефективніше справлятися зі складними завданнями і стало ключовим нововведенням, популяризованим в основній статті «Attention Is All You Need» [37], в якій була представлена архітектура Transformer [38].

Цей механізм є ключовим компонентом моделі трансформера та являє собою специфічний тип механізму уваги.

Самоувага – це механізм, що використовується в машинному навчанні, зокрема в обробці природної мови (NLP) та завданнях комп'ютерного зору, для фіксації залежностей та зв'язків у вхідних послідовностях. Це дозволяє моделі ідентифікувати та зважувати важливість різних частин вхідної послідовності, звертаючи на себе увагу. Вона працює шляхом перетворення вхідної послідовності на три вектори: запит, ключ і значення. Ці вектори отримуються за допомогою лінійних перетворень вхідних даних. Механізм уваги обчислює зважену суму значень на основі подібності між векторами запиту та ключа.

Отримана зважена сума разом з вихідними вхідними даними потім пропускається через нейронну мережу прямого зв'язку для отримання кінцевого результату. Цей процес дозволяє моделі зосередитися на релевантній інформації та враховувати довгострокові залежності [39].

Хоча самоувага є потужним механізмом, але покладаючись на один набір показників уваги, можна обмежити здатність моделі зосереджуватися на різних аспектах вхідних даних. Тут на допомогу приходить багатоголова увага (Multi-Head Attention). Замість обчислення однієї функції уваги, увага кількох голів запускає декілька операцій уваги паралельно, дозволяючи моделі звертати увагу на різні частини вхідної послідовності одночасно. Багатоголова увага розділяє вектори запитів, ключів та значень на кілька менших векторів та виконує самоаналіз незалежно від кожного з них. Результати з цих головок потім об'єднуються та лінійно перетворюються для отримання кінцевого результату. Ключова ідея полягає в тому, що кожен фокус уваги може зосереджуватися на різних частинах послідовності, що дозволяє моделі фіксувати різні зв'язки між словами [40].

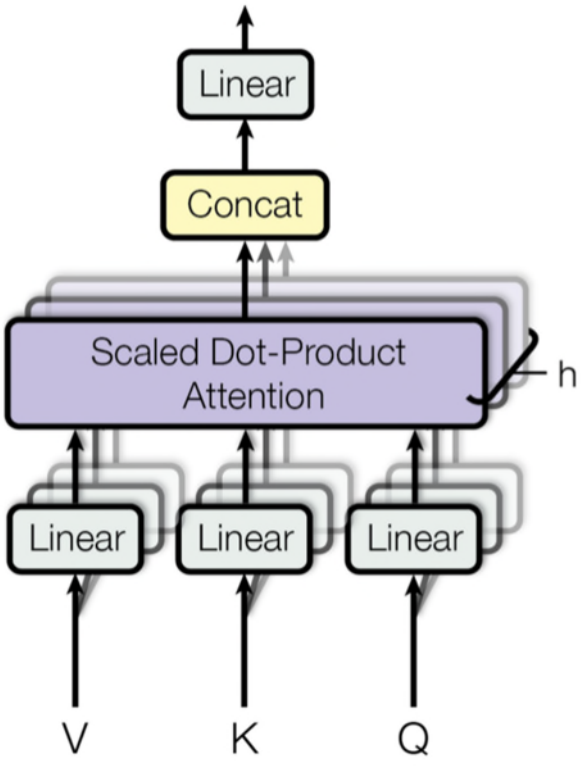


Рисунок 2.5 – Механізм багатоголової уваги [37]

Таким чином LLM на основі трансформерів (Transformer) – це потужні інструменти, які зробили революцію в області обробки природної мови, дозволивши моделям розуміти та генерувати текст, схожий на людину, з більшою точністю та ефективністю.

2.4 Метод автоматичного резюмування відеоматеріалів

Метод автоматичного резюмування, базується на інтеграції технологій автоматичного розпізнавання мовлення (ASR) та великих мовних моделей (LLM). Детальна схема методу, автоматичного резюмування відеоматеріалів зображена на рисунку 2.6.

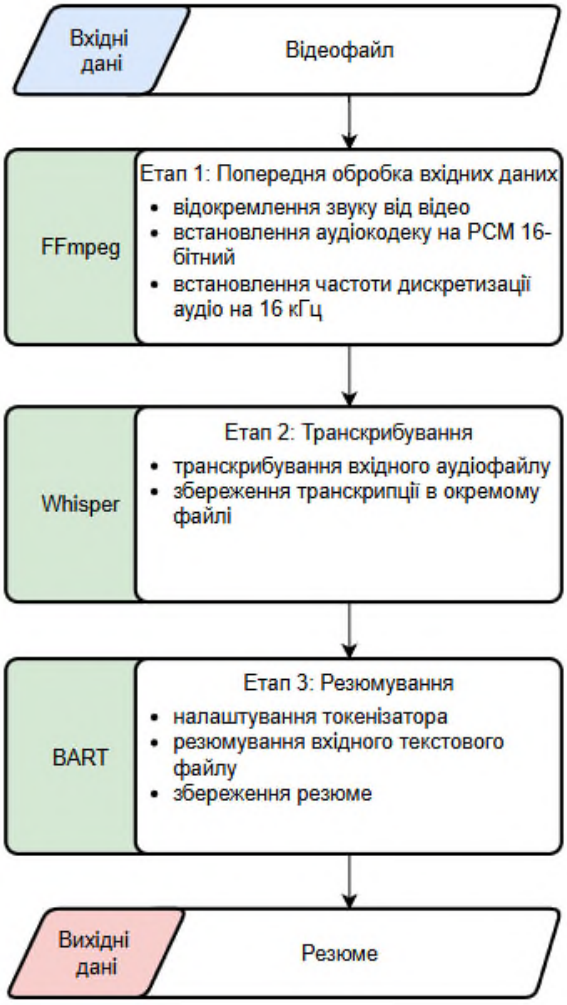


Рисунок 2.6 – Схема методу автоматичного резюмування відеоматеріалів

Метод як вхідні дані приймає відеофайли. Коли вхідний відеофайл отримано запускається етап 1: обробка вхідних даних. Використовуючи інструмент командного рядка FFmpeg [41], аудіоряд відокремлюється від відео, конвертується у формат PCM, що є необхідним кроком, оскільки PCM це стандартний формат вхідних даних для моделей ASR. Після зміни формату перевизначається частота до 16кГц, бо моделі краще розпізнають мовлення саме в такому форматі.

Коли вхідний файл пройшов етап обробки, починається процес транскрибування (етап 2). На цьому етапі відформатований аудіофайл використовується моделлю ASR Whisper і на його основі створює текстовий файл, що містить транскрипцію вхідних даних.

Автоматичне розпізнавання мовлення (ASR) у моделі Whisper реалізується за допомогою архітектури «кодер-декодер» на основі трансформерів. Вхідний аудіосигнал спочатку перетворюється у послідовність двійкових спектрограм, які потім подаються на вхід кодера. Кодер, що складається з багатошарових трансформерних блоків, обробляє ці спектрограми і на виході формує контекстне представлення вхідного аудіо. Декодер, генерує послідовність текстових токенів на основі виходу кодера та попередньо згенерованих токенів. Для підвищення точності застосовуються механізми уваги, що дозволяють моделі фокусуватися на частинах вхідного аудіо під час генерації кожного токена.

Після отримання транскрипції запускається етап 3, суть якого полягає в створенні короткого резюме шляхом використання LLM моделі BART. Першим кроком на цьому етапі є ініціалізація токенізатора, чиє завдання полягає в правильному підрахунку токенів та фрагментації вхідних даних. Якщо вхідні дані є завеликими, тоді виконується поділ даних на частини, які окремо проходять резюмування, а потім об’єднуються в один вихідний файл.

Модель BART від Facebook AI є потужною архітектурою для генерації резюме, що базується на трансформерах. Процес резюмування в BART поділяється на кілька етапів: спочатку вхідний текст кодується за допомогою

енкодера BART, який обробляє послідовність токенів, отримуючи представлення кожного слова. Ці представлення, що містять інформацію про взаємозв'язки між словами у вхідному тексті, потім передаються декодеру. Декодер BART послідовно генерує токени вихідного резюме, використовуючи інформацію з енкодера та раніше згенеровані токени. На кожному кроці генерації токена модель за допомогою механізму уваги обирає найрелевантніші частини вхідного тексту, що допомагає їй зберігати ключову інформацію та забезпечувати когерентність і зв'язність резюме.

По завершенні виконання всіх етапів користувачу стає доступним резюме, яке в повній мірі передає контекст оригіналу.

Таким чином, передбачивши всі необхідні кроки виконання ще на етапі проєктування, можна бути певним, що під час розробки методу не виникатиме неочікуваних критичних сценаріїв, що в свою чергу підвищить швидкість розробки.

2.5 Критерії та метрики оцінювання ефективності методу автоматичного резюмування відеоматеріалів

Ефективність методу автоматичного резюмування відеоматеріалів оцінюється через показники релевантності та точності. Це є комплексним завданням, що складається з кількох етапів, оцінювання роботи яких суттєво відрізняється.

На початковому етапі застосовується модель автоматичного розпізнавання мовлення (ASR), яка тренується на великих наборах даних і часто потребує глибокого навчання, оскільки ручна обробка значних обсягів даних є надзвичайно складною для людини. Для створення базової моделі інженери використовують великі масиви немаркованих даних, після чого модель точно налаштовується для досягнення оптимального коефіцієнта помилок слів (WER). Оцінка точності перетворення мовлення в текст є ключовим аспектом аналізу роботи систем ASR. Коефіцієнт помилок у словах (WER), як одна з

найпоширеніших і стандартизованих метрик, забезпечує об’єктивне порівняння вихідних даних системи ASR з еталонними транскрипціями, що дозволяє чітко визначити ефективність розпізнавання.

Коефіцієнт помилок слів (WER) – це показник, який оцінює точність систем ASR шляхом аналізу результатів перетворення мовлення на текст. Нижчий показник WER вказує на кращу продуктивність ASR, і навпаки [42]. Щоб отримати WER, треба додати кількість замін, вставок та видалень, які відбуваються в послідовності розпізнаних слів. А потім поділити отримане число на загальну кількість слів в оригінальному тексті. Результатом є WER [43].

Математична формула обислення WER виглядає наступним чином (формула 2.1):

$$WER = \frac{(S+D+I)}{N} \tag{2.1}$$

де: S (помилка підстановки) – заміна слова з оригінального тексту на неправильне, D (помилка видалення) – видалення слова з оригінального тексту та I (помилка вставки) – Заміна слова з оригінального тексту на неправильне [44].

Таким чином обчисливши значення WER на основі тестового значення, можна визначити якість виконання етапу розпізнавання мовлення.

Етап резюмування слідує відразу за етапом розпізнавання і його оцінювання якості суттєво відрізняється від ASR. Оцінка LLM є складним процесом, оскільки, на відміну від традиційної розробки програмного забезпечення, де результати передбачувані, а помилки можна налагодити, LLM є чорною скринькою з нескінченною кількістю можливих вхідних даних і відповідних вихідних даних.

Наприклад, для модульного тесту, що оцінює якість резюме, створеного LLM, критеріями можуть бути те, чи містить резюме достатньо інформації та чи містить воно будь-які галюцинації з оригінального тексту. Оцінка критеріїв здійснюється за допомогою так званої метрики оцінювання LLM (Рисунок 2.7)

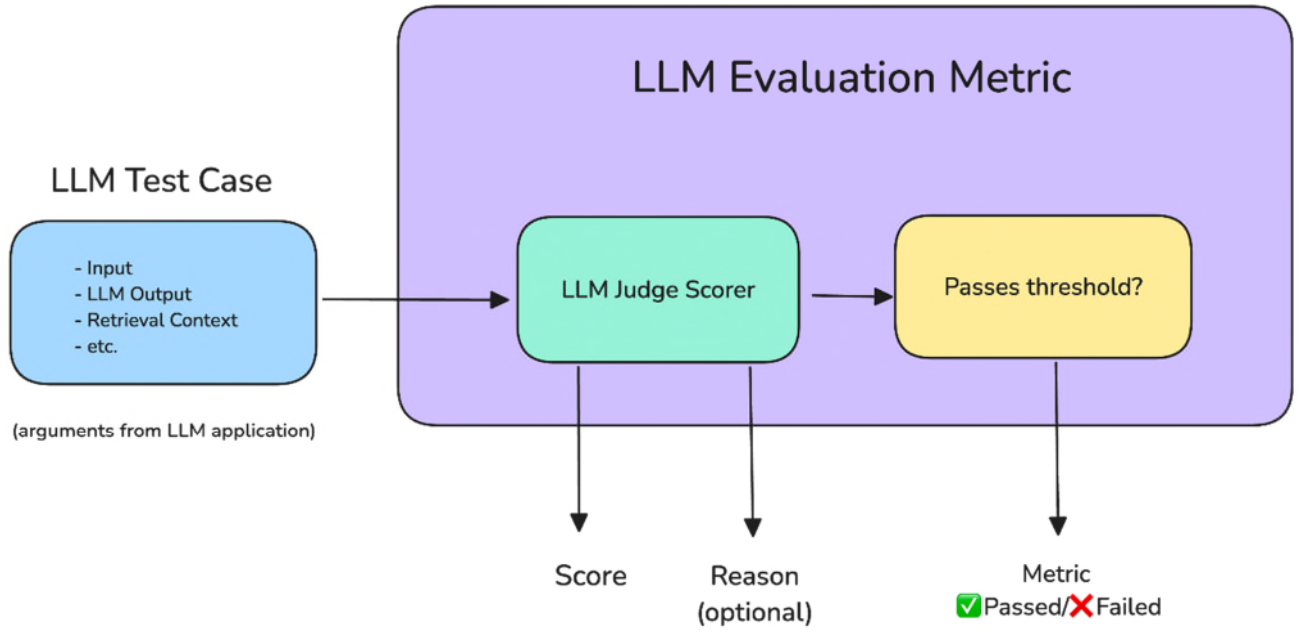


Рисунок 2.7 – Архітектура метрик оцінювання LLM [45]

Процес автоматичного резюмування тексту передбачає дотримання ключових критеріїв, що визначають якість роботи великих мовних моделей. Релевантність забезпечує збереження основних ідей і суттєвих деталей вихідного тексту в резюме. Лаконічність характеризується високою інформаційною насиченістю, уникненням повторів і надмірної багатослівності. Послідовність викладу сприяє логічній структурованості резюме, що полегшує його сприйняття. Достовірність змісту виключає наявність інформації, не підтвердженої вихідним матеріалом.

Для оцінки якості автоматично створених резюме застосовується підхід, заснований на аналізі великою мовною моделлю (LLM). Цей метод дозволяє оцінювати не лише лексичну відповідність між еталонним і згенерованим резюме, але й глибші аспекти, такі як змістова точність, релевантність і стисліть.

Оцінка ефективності великих мовних моделей здійснюється за допомогою фреймворку DeepEval, центральним елементом якого є G-Eval. DeepEval, що базується на великій мовній моделі, забезпечує семантичний аналіз згенерованого резюме порівняно з вихідним текстом, оцінюючи його за

заздалегідь визначеними критеріями. G-Eval аналізує якість резюме на основі порівняння з еталонними даними або оцінки відповідності заданим параметрам, таким як точність, повнота охоплення інформації, лаконічність, зв’язність і відсутність недостовірних даних [46].

Алгоритм оцінки включає формування набору даних, що складається з оригінального тексту, згенерованого резюме та запиту з чітко визначеними критеріями якості. LLM-оцінювач аналізує резюме в контексті вихідного документа, генеруючи оцінку, яка відображає ступінь відповідності критеріям у діапазоні від 0.0 (незадовільна якість) до 1.0 (висока якість). Оцінка в діапазоні від 0.0 – 0.5 вважається поганою або вказує на значні проблеми. Значення 0.5 – 0.7 вважається прийнятним, тоді як значення 0.7 – 0.9 вважається хорошим. Ітеративне застосування цього алгоритму на різних етапах розробки моделей резюмування сприяє кількісному відстеженню прогресу та виявленню аспектів, що потребують удосконалення.

Оцінювання передбачає подання великої мовної моделі спеціально сформульованого запиту, що включає вихідний текст, еталонне резюме та згенерований результат. На основі контекстуального аналізу модель повертає кількісну оцінку, яка залежить від точності передачі інформації, відповідності ключовим аспектам тексту та стислості викладу. На відміну від традиційних метрик, таких як ROUGE чи BLEU, LLM-орієнтоване оцінювання здатне виявляти семантичні невідповідності та недостовірні факти, що наближає результати до експертного аналізу, особливо в задачах, де важлива не лише лексична, а й змістова точність.

Таким чином для того, щоб оцінити ефективність роботи методу потрібно окремо протестувати основні модулі цього методу. Впевнитися що ASR модель коректно розпізнає мовлення, та впевнитися що LLM модель створює якісні резюме на основі вхідних даних.

2.6 Висновки до розділу 2

Для методу автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами запропоновано використання моделі ASR для автоматичного розпізнавання мовлення та LLM моделі для створення лаконічних резюме.

Детально описано структуру сучасних ASR моделей на прикладі Whisper. Вони забезпечують мінімальну затримку, виняткову точність і високу масштабованість та можуть обробляти різні типи введення, включно з кількома мовами.

Опираючись на модель BART розглянуто структуру LLM моделей, що реалізовані на базі транфсформерів, які в свою чергу використовують механізм уваги – техніку, що дозволяє моделі вибірково концентруватися на найбільш значущих частинах вхідних даних.

Розроблено метод автоматичного резюмування відеофайлів, реалізація якого передбачає використання FFmpeg (інструмент командного рядка), що використовується для попередньої обробки відео. Модель ASR Whisper, яка розпізнає мовлення та генерує текстовий файл. І модель LLM BART, що створює короткі резюме на основі вхідних відеофайлів.

Описано критерії оцінювання що обраховуватимуться шляхом обчислення значень WER для ASR моделей, та показника G-Eval, що використовує для оцінювання сторонню LLM модель.

Розділ 3 Особливості реалізації та результати тестування зпроєктованого методу.

3.1 Особливості розробки методу автоматичного резюмування відеоматеріалів.

3.1.1 Засоби розробки методу

Для реалізації методу автоматичного резюмування навчальних відеоматеріалів засобами штучного інтелекту обрано наступний стек технологій: мова програмування Python та інтегроване середовище розробки PyCharm. Також для повної реалізації методу було використано модель ASR Whisper, модель LLM для резюмування BART, модель gemini 2.0 flash для оцінювання якості резюме, фреймворк FFmpeg для роботи з медіафайлами, фреймворк DeepEval який спеціалізується на модульному тестуванні великих мовних моделей (LLM) та пакет JiWER для обчислення якості роботи ASR системи.

Whisper – це передовий інструмент ASR, який використовує методи глибокого навчання для розпізнавання мовлення з аудіофайлів. Це модель з відкритим вихідним кодом, який доступний будь кому. Whisper побудований на архітектурі Transformer, тієї ж архітектури, яка використовується в мовній моделі GPT-3 від OpenAI та DALL-E, іншої революційної моделі ШІ.

Однією з унікальних особливостей Whisper є його підтримка багатомовності. Він може розпізнавати мовлення на різних мовах, що робить його універсальним інструментом для дослідників та розробників, які працюють з багатомовними наборами даних. Він також включає функцію ідентифікації мови, яка може автоматично визначати вимовлене слово. Ця функція корисна при роботі з багатомовними наборами даних або при створенні чат-ботів, які повинні розпізнавати та відповідати кількома мовами, як ChatGPT [47].

BART – це модель кодера-декодера (seq2seq) на базі трансформера із двонаправленим (BERT-подібним) кодувальником і авторегресійним (GPT-подібним) декодером. BART попередньо навчається шляхом спотворення тексту

за допомогою довільної функції шуму та навчання моделі для реконструкції вихідного тексту [48].

BART особливо ефективний, для створення тексту (наприклад, резюме, переклад), але також добре працює для завдань розуміння (наприклад, класифікація тексту, відповідь на запитання) [49].

Gemini 2.0 Flash – це найновіша високошвидкісна та економічно ефективна модель штучного інтелекту від Google, представлена на початку 2025 року. Вона базується на успіху Gemini 1.5 Flash, пропонуючи покращену продуктивність, мультимодальні можливості та агентні функції, що робить її придатною для широкого кола застосувань [50].

FFmpeg – це провідна мультимедійна платформа, здатна декодувати, кодувати, перекодовувати, мультиплексувати, демультиплексувати, потоково передавати, фільтрувати та відтворювати майже все, що створили люди та машини. Він підтримує від найдавніших форматів аж до найсучасніших. Неважливо, чи були вони розроблені високопоставленим комітетом, спільнотою чи корпорацією.

FFmpeg намагається забезпечити найкраще технічно можливе рішення як для розробників програм, так і для кінцевих користувачів. Щоб досягти цього, це рішення поєднує в собі найкращі доступні варіанти безкоштовного програмного забезпечення [51]. Використання FFmpeg в своєю методі як крок обробки вхідних відеофайлів є чудовим рішенням, бо цей фреймворк пропонує усі необхідні рішення, щоб перетворювати файли в коректний формат.

DeepEval – це простий у використанні фреймворк з відкритим вихідним кодом для оцінювання та тестування систем на базі великих мовних моделей LLM. Він схожий на Pytest, але спеціалізований на модульному тестуванні результатів LLM. DeepEval включає останні дослідження для оцінки результатів LLM на основі таких показників, як G-Eval, галюцинація, релевантність відповіді, RAGAS тощо, які використовують LLM та різні інші моделі NLP, що запускаються локально на вашому комп’ютері для оцінки.

JiWER – це простий та швидкий пакет на Python для оцінки системи автоматичного розпізнавання мовлення (ASR). Він підтримує обчислення WER шляхом обчислення з використанням мінімальної відстані для редагування між одним або кількома реченнями, посилання та гіпотези [52].

У підсумку слід зазначити, що застосування зазначених технологій є необхідною умовою коректного функціонування методу, оскільки кожне з обраних рішень оптимально відповідає специфіці реалізації окремих його складових.

3.1.2 Структура та особливості реалізації методу автоматизованого резюмування навчальних відеоматеріалів

Для того, щоб візуально відобразити структуру роботи методу було реалізовано діаграму класів, що показує які функції і змінні використовуються в кожному класі (Рисунок 3.1).

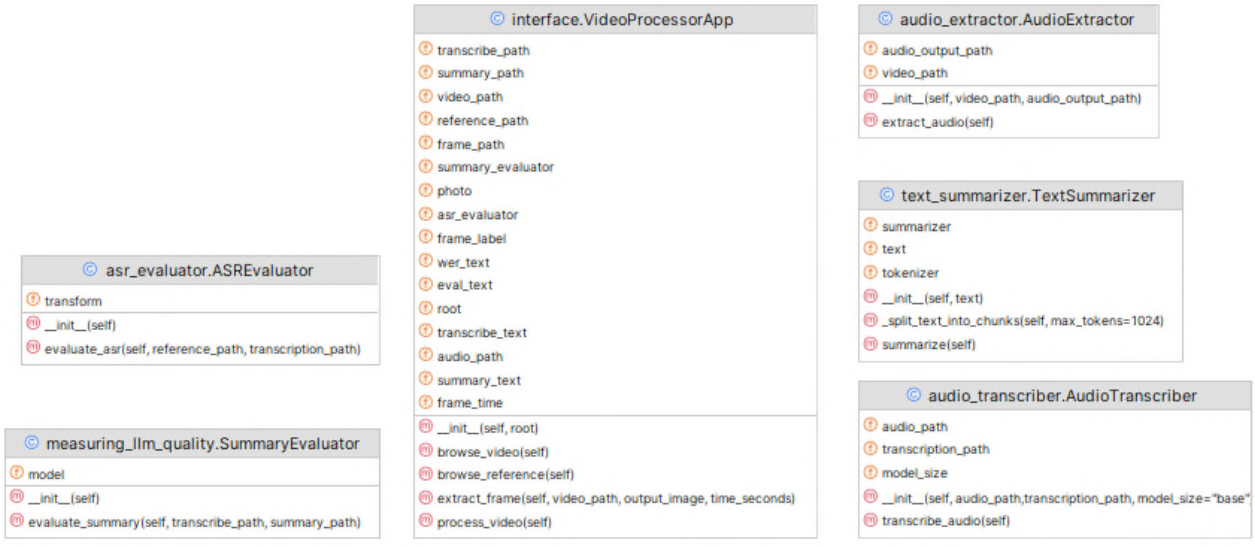


Рисунок 3.1 – Діаграма класів

На цій діаграмі зображено три ключові класи: AudioExtractor, AudioTranscriber і TextSummarizer. Вони забезпечують коректну виконання

кожного етапу, починаючи з обробки вхідних даних та закінчуючи формуванням відповіді.

Клас AudioExtractor реалізовує метод extract_audio(), в роботі якого використовується фреймворк FFmpeg. Цей метод відокремлює аудіо від відео та конвертує його в формат, який найбільше підходить для подальшого використання.

Клас AudioTranscriber для реалізації методу transcribe_audio() використовує сучасну модель Whisper, що спеціалізується на автоматичному розпізнаванні мовлення. Метод виконує функцію транскрипції: розпізнає мовлення з вхідного аудіофайлу та формує текстовий документ з розпізнаним текстом.

Клас TextSummarizer відповідає за резюмування вхідних текстових файлів. В ньому реалізовано два ключових методи без якого функціональність програми не є можливою. У випадках, коли вхідний текст є надто великим, модель LLM не може обробити дані одночасні, тому метод _split_text_into_chunks() розділяє вхідні дані на частини. Після чого метод summarize() проводить резюмування окремих частин і об’єднує результат в фінальне резюме. Візуалізацію процесу резюмування великих файлів зображено на рисунку 3.2.

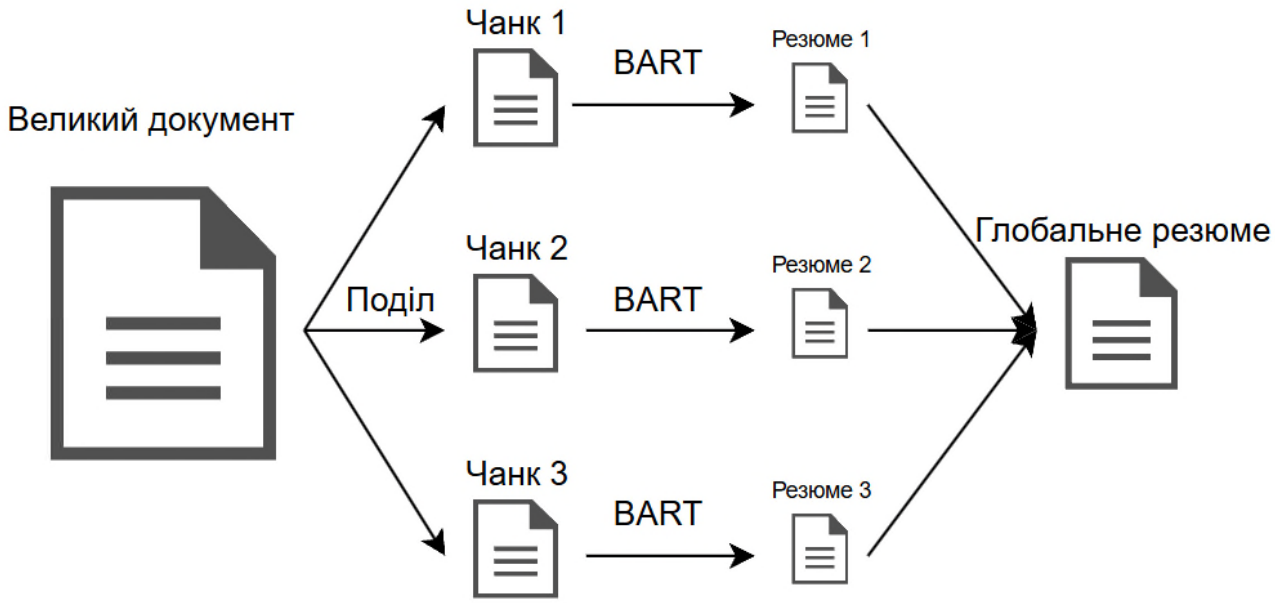


Рисунок 3.2 – Процес резюмування великих файлів

Клас ASREvaluation реалізовує обчислення відсотку помилкових слів для роботи ASR. Цей клас як вхідні дані отримує файл субтитрів та файл транскрипції. Після цього слід форматувати ці дані: призвести весь текст в нижній регістр, прибрати пунктуацію, прибрати великі пропуски і також перетворити вхідні текстові файли в рядок. Ці дії є гарантом якості оцінювання, бо якщо пропустити якийсь з кроків, алгоритм оцінювання може дати збій, наприклад через невидалені знаки пунктуації.

Клас SummaryEvaluation реалізовує оцінку якості створеного резюме. Для оцінки використовується стороння LLM, що виступає в ролі експерта. На вході отримується оригінальний документ і створене резюме. Після цього обчислюються показник G-Eval.

На основі розглянутої архітектури можна дійти висновку, що розроблений метод чітко організований та гнучкий, бо всі його ключові етапі розділені на окремі класи та можуть бути легко замінені у випадку модернізації чи виникнення непередбачуваних ситуацій.

3.2 Тестування методу автоматизованого резюмування навчальних відеоматеріалів

Тестування реалізованого методу автоматизованого резюмування навчальних відеоматеріалів відбуватиметься в два етапи на основі обчислення ключових метрик оцінювання роботи ASR та LLM систем.

Точність роботи моделі Whisper визначається шляхом обрахунку значень WER, що реалізовується шляхом використання бібліотеки JiWER.

Тоді як точність роботи моделі BART визначається через обрахунок показника G-Eval, для визначення якого в якості експерта використовується стороння LLM.

Для повного охоплення функціоналу розглянуто ряд тест-кейсів, кожен з яких має певні відмінності. Поділ відео на наступні категорії:

– Дискусія – відео в якому розмова відбувається між двома і більше людьми;

– Технічний контент – відео в якому йдеться про числові значення характеристик;

– Історичний контент – Відео в якому багато уваги приділяється датам;

– Природничий контент – Відео в якому йдеться про тварин, рослин та інші аспекти природи;

– Виступ – Відео де мовець перебуває на сцені і веде розповідь для широкого кола спостерігачів;

– Акцент мовця – Відео в якому мовець має чіткий іншомовний акцент;

– Багатомовне відео – Відео в якому зустрічаються вставки речень/діалогів які промовляються іншими мовами.

Кожне відео пропонує унікальну інформацію, що не корелюється з іншими тестовими випадками. Загальні характеристики тест-кейсів зображено в таблиці 3.1:

Таблиця 3.1 – характеристики тест-кейсів.

Тест-кейс	Назва	Тема	Категорія	Тривалість
1	Great Horned Owl on the Hunt	Полювання великого рогатого пугача	Природничий контент	3 хвилини 20 секунд
2	Hot Shot Rule	Застосування когнітивних стратегій у сфері лідерства та управління	Виступ	8 хвилин
3	Inside the King Tiger	Дослідження військових технологій середини XX століття	Технічний / Історичний контент	31 хвилина
4	Inside the Easy Eight Sherman Tank	Детальний опис танку M4A3E8 Sherman	Технічний / Історичний контент	18 хвилин 32 секунди
5	THE MORAL SIDE OF MURDER	Вступна лекція до курсу про справедливість	Виступ / Дискусія	54 хвилини 56 секунд
6	How do Graphics Cards Work?	Робота графічних карт	Технічний контент	28 хвилин 29 секунд
7	Japan's population crisis	Демографічна криза в Японії	Багатомовне відео	20 хвилин
8	Can re-freezing Arctic sea ice help save polar bears?	Полярні ведмеді, вплив танення арктичного льоду на їхню популяцію	Природничий контент / Дискусія	12 хвилин 16 секунд

9	Jamila Lyiscott: 3 ways to speak English	Володіння трьома стилями розмови	Виступ / Акцент мовця	4 хвилини 29 секунд
10	Why I keep speaking up, even when people mock my accent	Труднощі пов'язані з акцентом та заїканням. Поняття нормальності	Виступ / Акцент мовця	10 хвилин 48 секунд

Таким чином було створено набір тестових відеоматеріалів, для оцінки якості виконання роботи. Детально розглянуті тест-кейси:

Тест-кейс 1:

Для тестового випадку 1 обрано відеофайл, який містить інформацію велику свої очі рогату сову, що використовує та вуха для безшумних нападів на свою здобич [53]. Кількість слів відповідає значенню: 236. Результат виконання тест-кейсу зображено на рисунку 3.3.

Рисунок 3.3 – Результат виконання тест-кейсу 1

Кількість слів в резюме становить 87, що втричі менше за кількість оригінальному тексті.

Проведення програмного тестування якості роботи моделі Whisper показали результат, що відповідає значенню 5.51%. Це значення знаходяться в межах 0-10% що вказує на високу якість виконання розпізнавання.

Оцінка якості створеного резюме з використанням моделі BART сягає 0.71. Це значення знаходиться в діапазоні, що вказує на високу ефективність, коли зведення значною мірою відповідає визначеним критеріям для показника.

Тест-кейс 2:

Для тестового випадку 2 було обрано відео в якому йде мова про просту, але потужну практику під назвою «правило гарячого удару», яка допомагає перейти до лідерського мислення, звільнитися від інерції та вжити рішучих дій, коли це найважливіше [54]. Кількість слів: 1423. Результат виконання тест-кейсу зображено на рисунку 3.4.

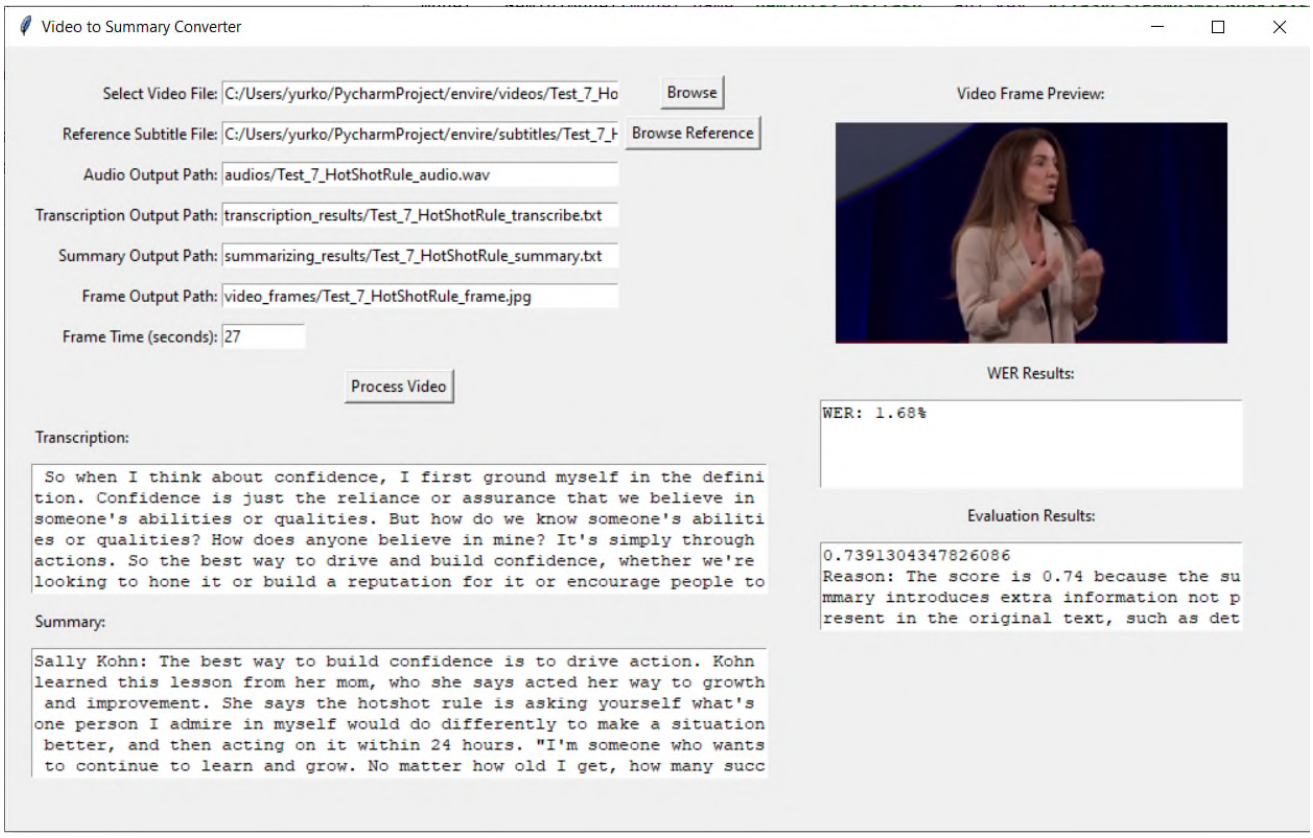


Рисунок 3.4 – Результат виконання тест-кейсу 2

В результаті виконання отримано резюме кількість слів якого відповідає значенню 447.

Після проведення обрахунків отримано значення WER: 1.68% Оскільки значення знаходяться в межах 0-10%, це вказує на високу якість виконання розпізнавання.

Оцінка якості створеного резюме відповідає значенню 0.74. Це значення знаходиться в діапазоні 0.7 – 0.9, і вказує на високу ефективність зведення, яке значною мірою відповідає визначеним критеріям для показника.

Тест-кейс 3:

Для тестового випадку 3 було обрано відео в якому йде мова танк Другої світової війни, Королівський тигр, також відомий як Тигр II. В відео детально розповідається про будову його конструкції, переваги та недоліки озброєння. Мета створення та історія використання [55]. Кількість слів в оригінальному файлі 4351. Результат виконання тест-кейсу зображено на рисунку 3.5.

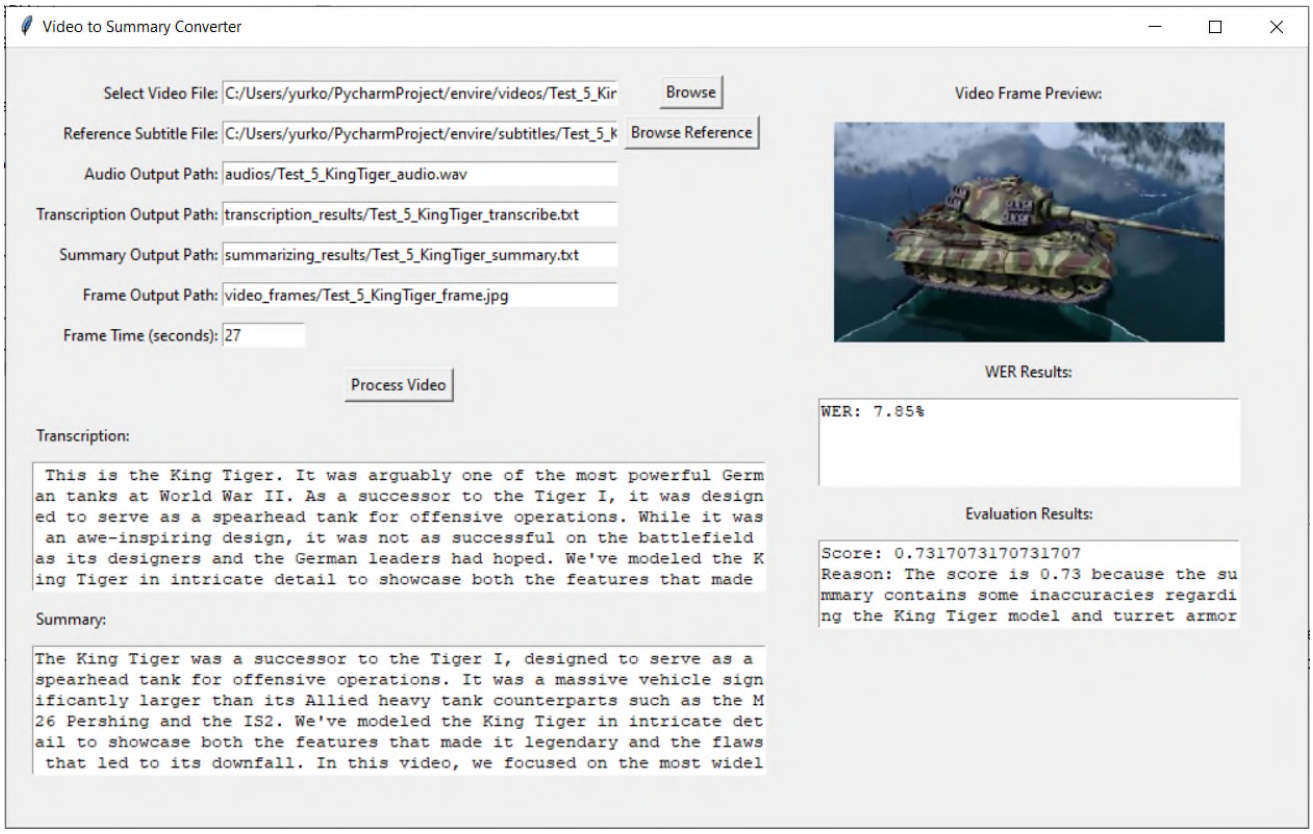


Рисунок 3.5 – Результат виконання тест-кейсу 3

В результаті виконання отримано резюме, кількість слів якого становить 828, що в 5 разів менше за кількість в оригінальному тексті.

Після проведення обрахунків отримано значення WER: 7.85%, що все ще є чудовим результатом бо знаходиться в межах 10% помилок.

Значення 0.73 показнику G-Eval знаходиться в діапазоні 0.7 – 0.9, і вказує на високу ефективність зведення.

Загальна таблиця в якій зображено результати тестування роботи зображено в таблиці 3.2.

Таблиця 3.2 – результати тестування

	Оцінка	
	WER	Показник G-Eval
Тестовий випадок 1	5.51%	0.71
Тестовий випадок 2	1.68 %	0.74
Тестовий випадок 3	7.85%	0.73
Тестовий випадок 4	7.65%	0.8
Тестовий випадок 5	9.50%	0.75
Тестовий випадок 6	4.46%	0.86
Тестовий випадок 7	17.39%	0.69
Тестовий випадок 8	3.35%	0.79
Тестовий випадок 9	11.95%	0.20
Тестовий випадок 10	12.08%	0.40

Після проведення тестування на основі отриманих оцінок створено графіки, що відображають значення оцінок для кожного тестового випадку (Рисунок 3.6 і 3.7).

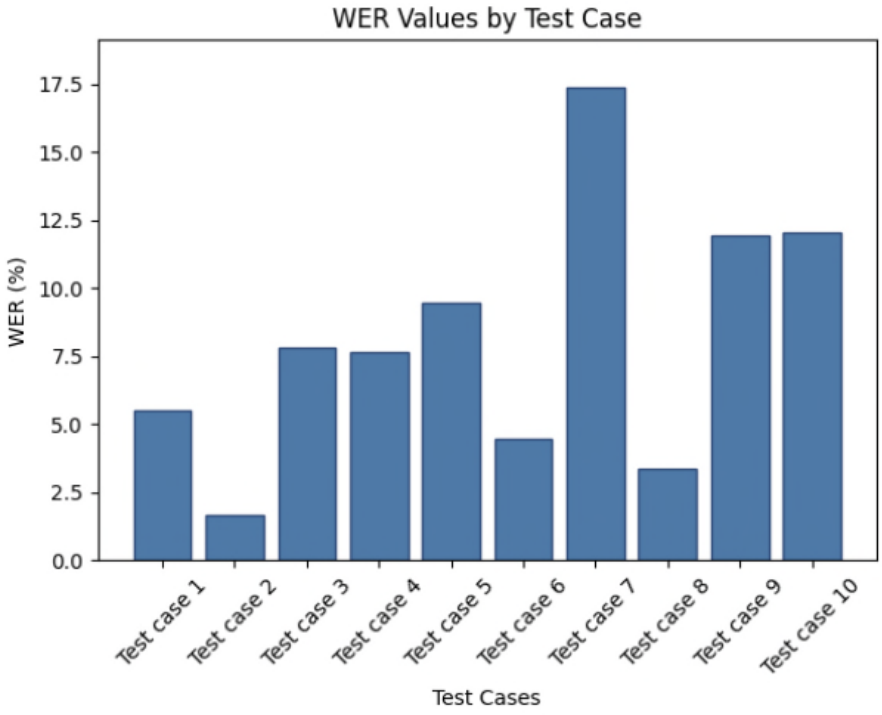


Рисунок 3.6 – Діаграма значень WER для кожного випадку

Опираючись на отримані значення WER, можна дійти висновку, що ASR система працює на хорошому рівні якості, бо значення WER в більшості своїй не перевищує 10% що для сучасних ASR зазначає на якість виконання розпізнавання. Існують також виключення які спостерігаються в тест-кейсах 7, 9 і 10. Ці значення є вищими за інші, оскільки розпізнавання мовлення і відповідно значення WER є дуже чутливим до акценту мовця та багатомовності.

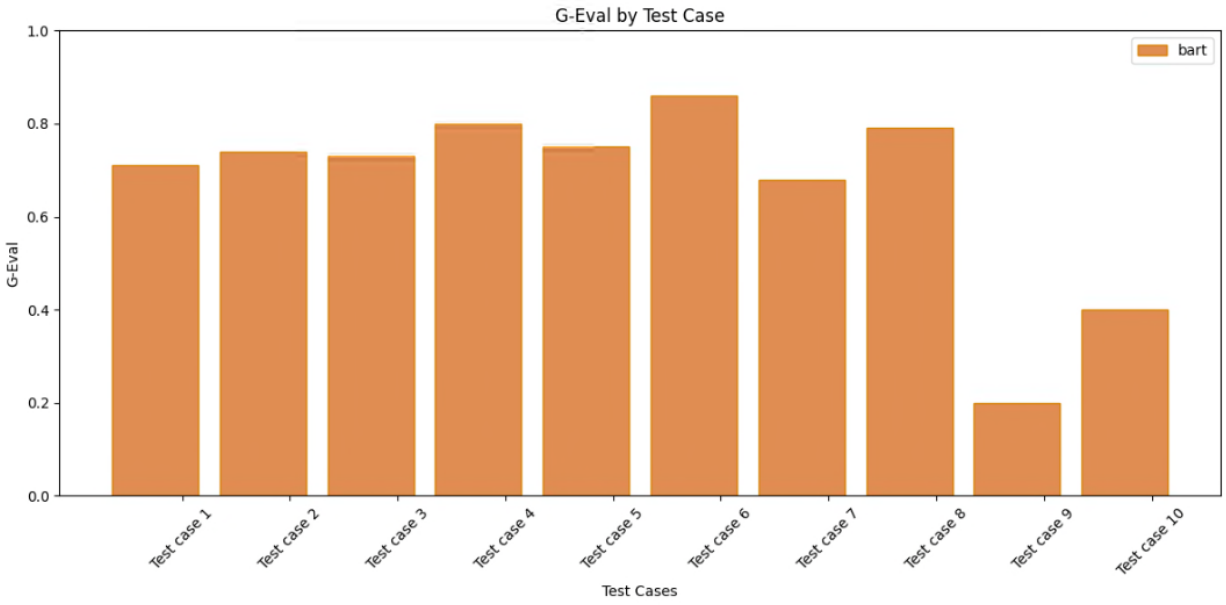


Рисунок 3.7 – Діаграма значень G-Eval

Діаграма, що зображена на рисунку 3.7 зображує значення показників G-Eval для тестових випадків. Опираючись на ці значення можна дійти висновку, що найкращі результати точності показують відео, в який переважно бере участь один мовець. Тип контенту може впливати на результат, оскільки спостерігається значний спад якості для тест кейсів 9 та 10. Це може пояснюватися тим, що в цих тестових відеофайлах значною мірою йшлося про особистий досвід мовця і життєві ситуації з резюмуванням яких модель справляється погано. Разом з тим, в цих відео є чітко виражений акцент що додатково може ускладнити резюмування, оскільки на етапі розпізнавання формується текст для якого є характерним граматичні помилки чи втрата контексту. Отримані значення оцінювання якості резюмування з використанням LLM вказують на високий рівень якості резюмування. Зустрічаються як випадки зі значенням 0.8, що вказують на хороший результат, так і значення що знаходяться нижче порогового значення 0.5, які можна пояснити особливістю вхідного відеофайлу.

На рисунку 3.8 зображено залежність значень WER та показників G-Eval.

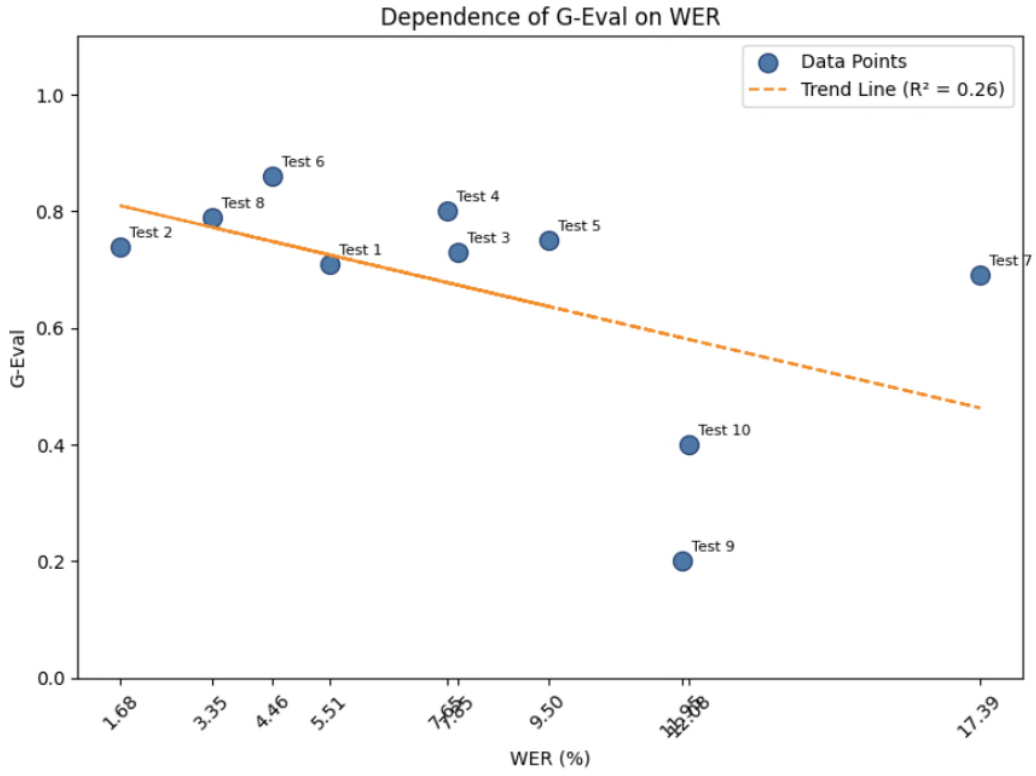


Рисунок 3.8 – Залежність показника G-Eval від значень WER

Графік ілюструє взаємозв'язок між показниками якості резюмування великих мовних моделей (LLM) та частотою помилок у словах (WER), що виникають внаслідок розпізнавання мовлення. Кожна точка на графіку відповідає окремому тесту, де вісь x позначає значення WER у відсотках, а вісь y – відповідні показники G-Eval.

Згідно з графіком, спостерігається дуже слабка негативна лінійна кореляція між WER та показниками G-Eval, про що свідчить коефіцієнт детермінації $R^2 = 0.26$. Це значення вказує, що лише 26% показників G-Eval може бути пояснено варіацією WER. Це вказує на те, що WER не має значного впливу на якість згенерованих резюме.

Зокрема, варто відзначити, що високі значення показників G-Eval досягаються навіть при відносно високих значеннях WER. Наприклад, тест-кейс 1 демонструє показник G-Eval 0.71 при WER близько 5.51%, а тест-кейси 4 та 5 також мають високі показники (понад 0.75) при WER в діапазоні 7-9%. Це підтверджує здатність LLM генерувати якісні резюме навіть за умов значної кількості помилок у вихідному розпізнаному мовленні.

Водночас, існують випадки, де при високому WER (тест кейс 7) показник G-Eval залишається на прийнятному рівні. Разом з тим, тест-кейс 9 та 10 демонструють низькі показники при порівняно високому WER, що може вказувати на вплив інших факторів, окрім WER, таких як чітко виражений акцент мовця чи особливості контенту.

Узагальнюючи, існує незначна тенденція до зниження якості роботи LLM зі збільшенням WER, але ця залежність є слабкою. Це означає, що LLM володіють певною стійкістю до помилок, що можуть зустрічатися у вхідному тексті, і здатні створювати ефективні резюме навіть за неідеальних умов. Це може бути пов'язано з потужними механізмами LLM для розуміння контексту та виправлення помилок, що дозволяє їм компенсувати недоліки вхідних даних.

Опираючись на результати тестування обраховано середні значення WER, що відповідає 8.142% та значення G-Eval для тест-кейсів, які перетнули порогове значення в 0.5 відповідає оцінці 0,75875.

Значення 8.142% для WER свідчить про високу якість роботи системи розпізнавання мовлення, оскільки знаходиться діапазоні 5-10%, що вважається хорошим результатом. Це означає, що система ASR створює транскрипції з достатньо низьким рівнем помилок, щоб бути використаною як вхідні дані для наступних етапів обробки.

Використання G-Eval на базі Gemini 2.0 Flash для оцінки надає глибоку та контекстуальну оцінку, оскільки ця модель здатна враховувати семантичну відповідність, когерентність та релевантність згенерованого резюме. Середнє значення 0,75875 яке отримано в ході обчислення показників G-Eval вказує на високу загальну якість резюмування, виконаного моделлю BART.

Для порівняння якості було проведено тестування шляхом використання інших популярних моделей, що спеціалізуються на резюмуванні тексту: t5-small та pegasus-xsum. Результати тестування зображені на рисунку 3.9.

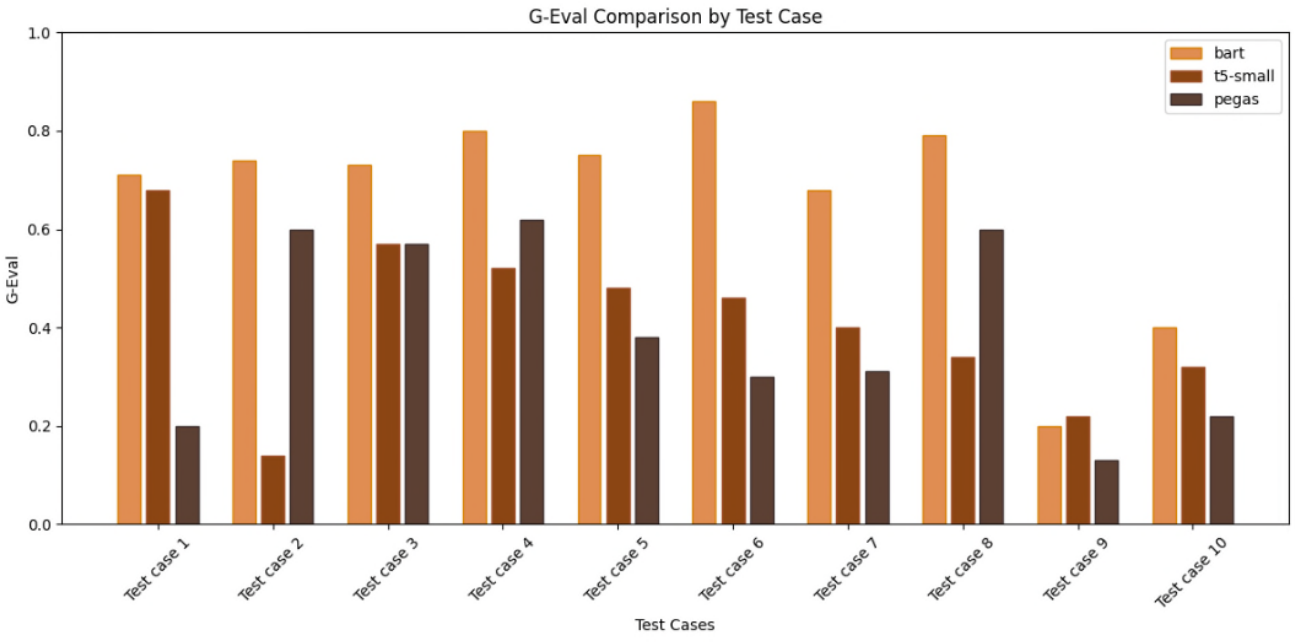


Рисунок 3.9 – Порівняння якості резюмування різних моделей

Діаграма зображує відмінність показників G-Eval для кожного тест-кейсу. Опираючись на ці дані, можна стверджувати, що використана в процесі розробки

методу автоматичного резюмування відеоматеріалів модель BART створює якісніші резюме, оскільки показники G-Eval для переважної більшості тест-кейсів є вищими за показники порівнюваних моделей.

За результатами проведеного тестування, що охоплювало десять тестових випадків, встановлено високу точність автоматичного розпізнавання мовлення (ASR). Переважна більшість отриманих показників ASR знаходяться в межах 10%, що свідчить про високу точність розпізнавання. Одиничні випадки, які перевищують зазначений поріг, пояснюються такими чинниками, як низька якість звуку, наявність іноземного акценту у мовця чи коротких мовних вставок (японською в випадку тест-кейсу 7).

Великої мовна модель (LLM) для резюмування продемонструвала хороший рівень релевантності. Сформовані резюме у більшості випадків адекватно передають смисловий контекст вихідних аудіофайлів, що підтверджує її здатність до ефективного скорочення та збереження ключової інформації.

3.3 Висновки до розділу 3

Для реалізації методу автоматичного резюмування навчальних відеоматеріалів засобами штучного інтелекту було обрано наступний стек технологій: мова програмування Python та інтегроване середовище розробки (IDE) PyCharm. Також для повної реалізації методу було використано модель ASR Whisper, модель LLM для резюмування BART, модель gemini 2.0 flash для оцінювання якості резюме, фреймворк FFmpeg для роботи з медіафайлами, фреймворк DeepEval який спеціалізується на модульному тестуванні великих мовних моделей (LLM) та пакет JiWER для обчислення якості роботи ASR системи.

Реалізований метод має чіткий поділ на класи, кожен з яких відповідає за свою специфічну функцію. Перший клас займається видобуванням аудіо з відеофайлів та їх конвертацією у найбільш придатний для подальшої обробки формат. Наступний клас транскрибує аудіо за допомогою сучасної моделі

автоматичного розпізнавання мовлення, перетворюючи записане мовлення на текстовий документ. Далі, окремий клас відповідає за резюмування отриманих текстових файлів, використовуючи провідну модель для створення стислих, але змістовних витягів, що значно менші за оригінал. Крім того, реалізовано класи, що забезпечують оцінювання якості як автоматичного розпізнавання мовлення, так і створених резюме.

Тестування проведено на базі десяти тестових випадків, кожен з яких є унікальним і відрізняється: тривалістю, контентом, формою віщання (діалог чи монолог). Тестування показало, що мовна модель BART, демонструє низьку схильність до помилок, які виникають на етапі розпізнавання. Середнє значення частоти помилок у словах (WER) на рівні 8.142% свідчить високу точність розпізнавання. Незважаючи на ці помилки, середній показник G-Eva, який використовується для оцінювання роботи LLM, становить 0,75875, що вказує на високу якість кінцевих резюме. Також проведено порівняння роботи інших провідних моделей на основі значень G-Eval для тест-кейсів. Результати показали, що використання моделі BART показує значно кращі результати.

Таким чином, результати тестування свідчать про високий потенціал використання сучасних LLM для автоматичного резюмування аудіо або відеоматеріалів, навіть за відсутності ідеальної якості розпізнавання мовлення.

Загальні висновки

Метою кваліфікаційної роботи бакалавра є підвищення релевантності та точності автоматичного резюмування навчальних відеоматеріалів засобами глибокого навчання.

Реалізований метод дозволяє суттєво економити час на перегляді освітніх відеоматеріалів через створення коротких та лаконічних резюме, які в повній мірі передають оригінальний контекст.

Для коректної реалізації було задано наступний список завдань виконання яких є запорукою успішної роботи:

– дослідити сучасні методи та технології обробки мовлення та автоматичного резюмування тексту;

– розробити метод автоматичного резюмування навчальних відеоматеріалів з використанням нейромережевих засобів;

– створити програмну реалізацію методу автоматичного резюмування навчальних відеоматеріалів для обробки аудіо- та текстових даних;

– оцінити ефективність методу автоматичного резюмування.

Порівняння якості роботи кількох популярних моделей для створення коротких резюме показало, що модель BART створює якісніші резюме, оскільки в переважній більшості тестових випадків показник G-Eval був більшим за значення відповідних показників отриманих для інших моделей.

Тестування системи показало, що розроблений метод в повній мірі виконує своє першочергове завдання. Отримані значення оцінки роботи ASR та LLM систем показали, що в більшості випадків резюме якісно передає контекст оригінального відеофайлу.

Перелік посилань

1. The Rise of Video Content: How to Maximize Engagement in 2024. *Damn Art*. URL: https://damnart.com/the-rise-of-video-content-how-to-maximize-engagement-in-2024/.

2. Understanding Video Content Types. *Quizgecko | AI Quiz Maker*. URL: https://quizgecko.com/learn/understanding-video-content-types-kgp0ml.

3. Kumari N., Singh P. Text Summarization and Its Types: A Literature Review. *Handbook of Research on Natural Language Processing and Smart Service Systems*. 2021. P. 368–378. https://doi.org/10.4018/978-1-7998-4730-4.ch017

4. Types of Text Summarization: Extractive and Abstractive Summarization Basics. Turbolab Technologies. URL: https://turbolab.in/types-of-text-summarization-extractive-and-abstractive-summarization-basics/.

5. Jain D., Borah M. D., Biswas A. Summarization of legal documents: Where are we now and the way forward. Computer Science Review. 2021. Vol. 40. P. 100388. URL: https://doi.org/10.1016/j.cosrev.2021.100388).

6. Abstractive Text Summarization. *Papers With Code*. URL: https://paperswithcode.com/task/abstractive-text-summarization.

7. Speech Recognition Guide: Top Tools & Technologies in 2025. *Transkriptor*. URL: https://transkriptor.com/speech-recognition/.

8. Exploring the Diversity of Speech Recognition Technologies. *Transkriptor*. URL: https://transkriptor.com/speech-recognition-types/.

9. What is speech recognition? A comprehensive guide. *AssemblyAI | AI models to transcribe and understand speech*. URL: https://www.assemblyai.com/blog/speech-recognition.

10. Kirvan P., Lutkevich B., Kiwak K. What is Speech Recognition? | Definition from TechTarget. *Search Customer Experience*. URL: https://www.techtarget.com/searchcustomerexperience/definition/speech-recognition.

11. Introduction to Speech Processing: 2nd Edition / T. Bäckström et al. *Zenodo*. URL: https://doi.org/10.5281/zenodo.6821775.

12. Acoustic Modeling - Microsoft Research. *Microsoft Research*. URL: https://www.microsoft.com/en-us/research/project/acoustic-modeling/.

13. Acoustic Models | Deepgram. *Deepgram*. URL: https://deepgram.com/ai-glossary/acoustic-models.

14. Bento C. Hidden markov models explained with a real life example and python code. *Medium*. URL: https://medium.com/data-science/hidden-markov-models-explained-with-a-real-life-example-and-python-code-2df2a7956d65.

15. What is a deep neural network?. *Botpress | The Complete AI Agent Platform*. URL: https://botpress.com/blog/deep-neural-network.

16. A beginner's guide to language models | built in. *Built In*. URL: https://builtin.com/data-science/beginners-guide-language-models.

17. Speech and language processing. *Stanford University*. URL: https://web.stanford.edu/~jurafsky/slp3/.

18. Activeloop.ai. Maximum Entropy Models. URL: https://www.activeloop.ai/resources/glossary/maximum-entropy-models/

19. AI summarization: how it works and 5 tips for success. *Acorn Labs*. URL: https://www.acorn.io/resources/learning-center/ai-summarization/ (date of access: 29.05.2025).

20. Text Summarization. *Knowledge Zone*. URL: https://knowledgezone.co.in/posts/6394366618f02731f467ecee.

21. Sandu C. Sequence-to-Sequence Models. *Medium*. URL: https://medium.com/@calin.sandu/sequence-to-sequence-models-603920ce9e96.

22. What Is a Transformer Model?. *NVIDIA Blog*. URL: https://blogs.nvidia.com/blog/what-is-a-transformer-model/.

23. Stryker C., Bergmann D. What is a Transformer Model? | IBM. *IBM - United States*. URL: https://www.ibm.com/think/topics/transformer-model.

24. What Is a Large Language Model (LLM)? - DATAVERSITY. *DATAVERSITY*. URL: https://www.dataversity.net/what-is-a-large-language-model-llm/.

25. Otter.ai URL: https://otter.ai/home

26. Speechnotes URL: https://speechnotes.co/

27. SMMRY URL: https://smmry.com/

28. Quillbot.com URL: https://quillbot.com/

29. Zoom URL: https://www.zoom.com/

30. Gladia - How Do ASR Models Work?. *Gladia I Audio Transcription API*. URL: https://www.gladia.io/blog/how-do-speech-recognition-models-work.

31. Introducing Whisper. *OpenAi*. URL: https://openai.com/index/whisper/.

32. Was ist ein großes Sprachmodell?. *Sap.com*. URL: https://www.sap.com/germany/resources/what-is-large-language-model.

33. GeeksforGeeks. LLM Architecture: Exploring the Technical Architecture Behind Large Language Models - GeeksforGeeks. *GeeksforGeeks*. URL: https://www.geeksforgeeks.org/exploring-the-technical-architecture-behind-large-language-models/.

34. Tokenization | Mistral AI Large Language Models. *Bienvenue to Mistral AI Documentation | Mistral AI Large Language Models*. URL: https://docs.mistral.ai/guides/tokenization/.

35. What is Embedding Layer: LLMs Explained. *Chatgptguide.ai*. URL: https://www.chatgptguide.ai/2024/02/29/what-is-embedding-layer-llms-explained/.

36. IBM. What Are Large Language Models (LLMs)? | IBM. *IBM - United States*. URL: https://www.ibm.com/think/topics/large-language-models.

37. Vaswani, Ashish, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin. “Attention is All you Need.” Neural Information Processing Systems (2017).

38. Ultralytics. Attention Mechanism in AI/ML Explained | Ultralytics. *Ultralytics | Revolutionizing the World of Vision AI*. URL: https://www.ultralytics.com/glossary/attention-mechanism.

39. What is Self-attention?. *H2O.ai | Convergence of the World’s Best Predictive and Generative AI for Private, Protected Data*. URL: https://h2o.ai/wiki/self-attention/.

40. Nishad N. Understanding Self-Attention and Multi-Head Attention in Deep Learning. *DEV Community*. URL: https://dev.to/nareshnishad/understanding-self-attention-and-multi-head-attention-in-deep-learning-4jg4.

41. FFmpeg. *FFmpeg*. URL: https://ffmpeg.org/.

42. Gladia - What is WER, or Why Benchmarks Are Misleading. *Gladia I Audio Transcription API*. URL: https://www.gladia.io/blog/what-is-wer.

43. What Is Word Error Rate (WER)? | Rev. *#1 Speech to Text Service For Lawyers + Beyond | Rev*. URL: https://www.rev.com/resources/what-is-wer-what-does-word-error-rate-mean.

44. Accuracy Benchmarking. *Speechmatics*. URL: https://docs.speechmatics.com/tutorials/calculating-wer.

45. LLM Evaluation Metrics: The Ultimate LLM Evaluation Guide - Confident AI. *Confident AI - The DeepEval LLM Evaluation Platform*. URL: https://www.confident-ai.com/blog/llm-evaluation-metrics-everything-you-need-for-llm-evaluation.

46. GitHub - confident-ai/deepeval: The LLM Evaluation Framework. *GitHub*. URL: https://github.com/confident-ai/deepeval.

47. What is Whisper from OpenAI? | Speechify. *Speechify*. URL: https://speechify.com/blog/what-is-whisper-from-openai/.

48. Zain ul Abideen. A Comparative Analysis of LLMs like BERT, BART, and T5. *Medium*. URL: https://medium.com/@zaiinn440/a-comparative-analysis-of-llms-like-bert-bart-and-t5-a4a873251ff.

49. facebook/bart-base · Hugging Face. *Hugging Face – The AI community building the future.* URL: https://huggingface.co/facebook/bart-base.

50. Pichai S. Introducing Gemini 2.0: our new AI model for the agentic era. *Google*. URL: https://blog.google/technology/google-deepmind/google-gemini-ai-update-december-2024.

51. About FFmpeg. *FFmpeg*. URL: https://ffmpeg.org/about.html.

52. GitHub - jitsi/jiwer: Evaluate your speech-to-text system with similarity measures such as word error rate (WER). *GitHub*. URL: https://github.com/jitsi/jiwer.

53. Nat Geo Animals. Great Horned Owl on the Hunt | Nat Geo Wild, 2019. *YouTube*. URL: https://www.youtube.com/watch?v=bt3X8MJgJWo.

54. TED. The “Hot Shot Rule” To Help You Become a Better Leader | Kat Cole | TED, 2025. *YouTube*. URL: https://www.youtube.com/watch?v=lKsvLGdoIH8.

55. Blue Paw Print. Inside the King Tiger, 2025. *YouTube*. URL: https://www.youtube.com/watch?v=vCgdYHatDqw.

ДОДАТКИ

Додаток А

Програмні коди

Програмна реалізація методу автоматичного резюмування відеоматеріалів знаходиться в репозиторії GitHub: https://github.com/YuraAntoniuk/AVS.

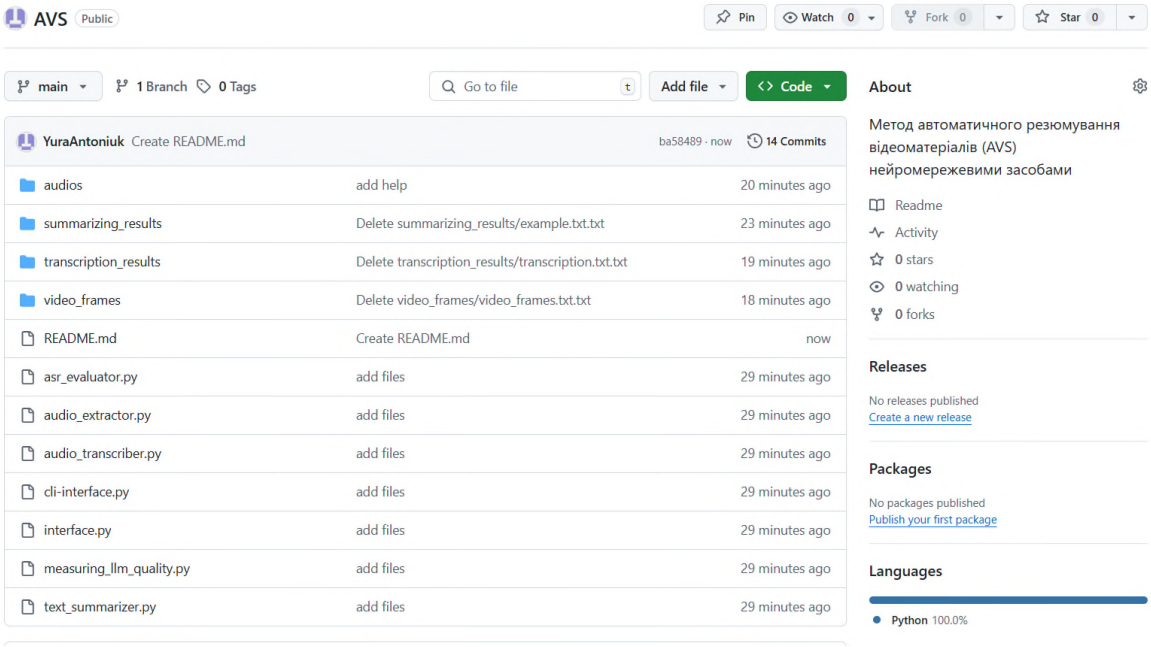


Рисунок А.1 – Головна сторінка репозиторію

Структура репозиторію:

– папки для збереження проміжних і фінальних результатів виконання програмного коду (audios, summarizing_results, transcription_results, video_frames);

– програмний код реалізований в окремих файлах (класах), що відповідає за коректну роботу на кожному етапі автоматичного резюмування.

Додаток Б

Презентація

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

Метод автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами

Виконав: Антонюк Юрій
Студент 4 курсу,
групи КН-21-2

Керівник: Багрій Руслан
Олександрович
Доцент кафедри
Комп'ютерних наук

Актуальність

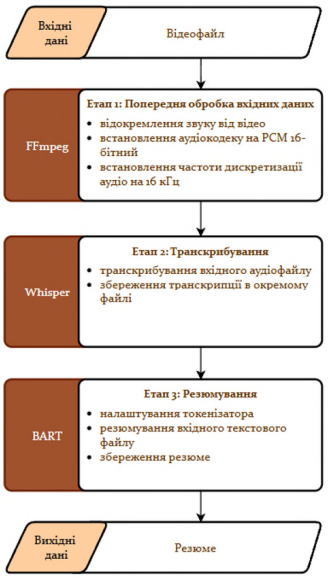
В епоху стрімкого розвитку цифрової освіти та масового поширення онлайн-курсів, вебінарів і відеолекцій, обсяги навчальних відеоматеріалів сильно зростають. Це створює значне навантаження на здобувачів освіти та викладачів, яким доводиться витрачати велику кількість часу на перегляд повного контенту для виявлення ключових концепцій та інформації. Використання методу автоматичного резюмування навчальних відеоматеріалів дозволяє суттєво оптимізувати процес засвоєння знань, забезпечуючи швидкий доступ до суті лекцій та тренінгів.

Мета: підвищення релевантності та точності автоматичного резюмування навчальних відеоматеріалів засобами глибокого навчання.

Завдання:

- дослідити сучасні методи та технології обробки мовлення та автоматичного резюмування тексту;
- розробити метод автоматичного резюмування навчальних відеоматеріалів з використанням нейромережевих засобів;
- створити програмну реалізацію методу автоматичного резюмування навчальних відеоматеріалів для обробки аудіо- та текстових даних;
- оцінити ефективність методу автоматичного резюмування.

Схема методу автоматичного резюмування відеоматеріалів



Характеристики тест-кейсів

Тест-кейс	Назва	Тема	Тривалість
1	Great Horned Owl on the Hunt	Полювання великого рогатого пугача	3 хвилини 20 секунд
2	Hot Shot Rule	Застосування когнітивних стратегій у сфері лідерства та управління	8 хвилин
3	Inside the King Tiger	Дослідження військових технологій середини XX століття	31 хвилина
4	Inside the Easy Eight Sherman Tank	Детальний опис танку M4A3E8 Sherman	18 хвилин 32 секунди
5	THE MORAL SIDE OF MURDER	Вступна лекція до курсу про справедливість	54 хвилини 56 секунд
6	How do Graphics Cards Work?	Робота графічних карт	28 хвилин 29 секунд
7	Japan's population crisis	Демографічна криза в Японії	20 хвилин
8	Can re-freezing Arctic sea ice help save polar bears?	Полярні ведмеді, вплив танення арктичного льоду на їхню популяцію	12 хвилин 16 секунд
9	Jamila Lyiscott: 3 ways to speak English	Володіння трьома стилями розмови	4 хвилини 29 секунд
10	Why I keep speaking up, even when people mock my accent	Труднощі пов'язані з акцентом та заїканням. Поняття нормальності	10 хвилин 48 секунд

Приклад резюмування

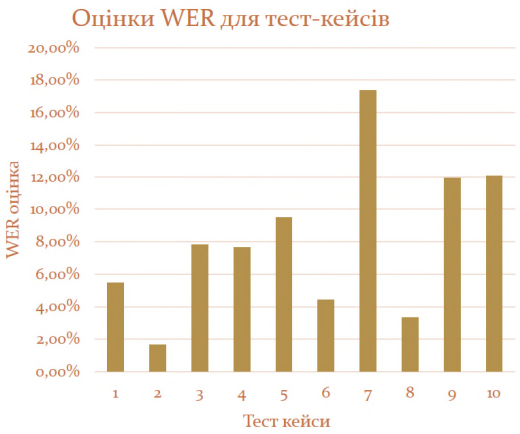
Оригінальний текст

This is the great horned owl's time. He's one of the largest and most powerful owl species in America. As dusk turns to dark, he looks down into a forest clearing. His eyes are tuned for optimal night vision, but more importantly, he listens. His ear-like horns are simply toughs of feathers. No one knows for sure what they're for, but they have nothing to do with hearing. His real ears are hidden under his head feathers. Their positioned asymmetrically, the right one is slightly higher than the left one

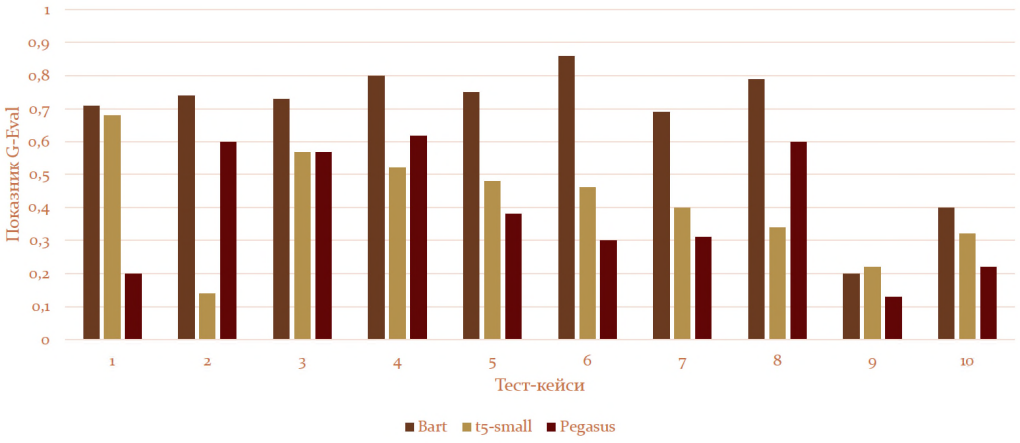
Резюме:

The great horned owl is one of the largest and most powerful owl species in America. His ear-like horns are simply toughs of feathers. His real ears are hidden under his head feathers. Their positioned asymmetrically, the right one is slightly higher than the left one.

Результати тестування



Порівняння якості резюмування різних моделей



Висновок

Реалізований метод дозволяє суттєво економити час на перегляді освітніх відеоматеріалів через створення коротких та лаконічних резюме. Порівняння якості роботи кількох популярних моделей для створення коротких резюме показало, що модель BART створює якісніші резюме, оскільки в переважній більшості тестових випадків показник G-Eval був більшим за значення відповідних показників отриманих для інших моделей. Тестування системи показало, що розроблений метод в повній мірі виконує своє першочергове завдання. Отримані значення оцінки роботи ASR та LLM систем підтверджують, що в більшості випадків резюме якісно передає контекст оригінального відеофайлу.

Дякую за увагу

Thu Jun 19 11:29:03 EEST 2025, Петровський Сергій Степанович, Хмельницький національний університет, ХНУ

Anti-Plagiarism (UA) v-15.281 Educational

The maximum coincidence with one document 3.0%

Dictionaries check: en_US, ru_RU, ua_UA. **Errors in the documents: 12%**

	Document		Sum coincidence on the DB	
ID: 246903 Title: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами Added in a DB: 2025-06-19 Authors: Юрій АНТОНЮК Heads: Руслан БАГРІЙ Consultants: Opponents:	Symbols	Lexemes	Symbols	Lexemes
	57917	886	2645 (5%)	43 (5%)

Plagiarism sources

ID	Description	Plagiarism presence in the document	
		Symbols	Lexemes

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Юрій АНТОНЮК

Співавтор:

Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами

Науковий керівник: Руслан БАГРІЙ, к.т.н., доцент

Підрозділ: Кафедра комп'ютерних наук

Коефіцієнт подібності 1:6.5%

Коефіцієнт подібності 2:3%

Мікропробіли: 0

Заміна букв: 0

Інтервали: 0

Білі знаки: 48

Дата створення звіту: 2025-06-19 11:53:07.0

Після аналізу Звіту подібності констатую наступне:

☒ **Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.**

☐ **Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.**

☐ **Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедурам. Таким чином робота не приймається.**

Обгрунтування:

2025-06-19

Дата

експерт Петровський С. Р.

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК

ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Назва кваліфікаційної роботи *Метод автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами*
Автор *студент групи КН-21-2 Юрій Антонюк*
Освітня програма *Комп'ютерні науки*
Рівень вищої освіти *перший (бакалаврський)*
Спеціальність *122 – Комп'ютерні науки*
Науковий керівник: *к.т.н., доц. каф. комп'ютерних наук Руслан Багрій*

На основі аналізу кваліфікаційної роботи на дотримання вимог академічної доброчесності (у т.ч. відсутності ознак академічного плагіату) з урахуванням результатів перевірки роботи спеціалізованим програмними засобами комісія зробила такий висновок:

№	Висновок	Позначка про відповідність
1	Ознаки академічного плагіату	
1.1	Запозичення, виявлені в роботі, є законними і не є академічним плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних, якщо потрібно). Робота приймається до захисту.	***відповідає***
1.2	Виявлені запозичення не є академічним плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована.	
1.3	Виявлені запозичення не є академічним плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота може бути допущена до захисту після того як буде відкоригована та доопрацьована і успішно пройде повторну перевірку на академічний плагіат.	
1.4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття текстових запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	
2	Інші види порушень академічної доброчесності	***відсутні***

Підтвердження:

Запозичення, виявлені в роботі Юрія Антонюка, не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; до запозичень входять фрагменти, які не мають авторства і містять поширені конструкції та загальновідомі терміни, скорочення. Рівень подібності не перевищує допустимої межі. Таким чином, робота є законною та приймається до захисту.

Обсяг запозичень, визначений системами виявлення збігів/ідентичності/схожості:
- за системою Anti-Plagiarism: 3%;
- за системою StrikePlagiarism КП1: 6.5%, КП2: 3%.

19.06.2025

Завідувач кафедри _____ *Олександр БАРМАК*

Гарант освітньої програми _____ *Олександр МАЗУРЕЦЬ*

Керівник кваліфікаційної роботи _____ *Руслан БАГРІЙ*

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ

Кафедра комп'ютерних наук

ВІДГУК НАУКОВОГО КЕРІВНИКА

на кваліфікаційну роботу бакалавра

студента *гр. КН-21-2 Антонюка Юрія Олександровича*

за темою *Метод автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами*

1. Актуальність теми

Актуальність теми обґрунтована зростанням обсягів навчальних відеоматеріалів у цифровій освіті, що створює потребу в оптимізації їх обробки для ефективного засвоєння знань студентами, викладачами та іншими користувачами. Особливістю теми є застосування нейромережевих засобів, зокрема технологій автоматичного розпізнавання мовлення і великих мовних моделей, для створення стислих і змістовних резюме, що суттєво економить час на перегляд освітнього контенту.

2. Відповідність роботи предметній області Стандарту спеціальності 122 Комп'ютерні науки

Тема кваліфікаційної роботи "Метод автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами" відповідає предметній області спеціальності 122 Комп'ютерні науки та вимогам до кваліфікаційної роботи бакалавра. Результатом роботи є розробка методу, що базується на інтеграції моделей ASR та LLM, для автоматичного резюмування відеоматеріалів. При вирішенні поставлених завдань використано методи обробки аудіо- та текстових даних, нейромережеві технології глибокого навчання, а також аналіз метрик якості для оцінки ефективності.

3. Професійні та особистісні якості бакалавра

Антонюк Ю. О. під час роботи над кваліфікаційною роботою продемонстрував глибоке розуміння теоретичних і практичних аспектів використання нейромережевих технологій для обробки аудіо- та текстових даних, а також здатність до самостійного аналізу та вирішення складних технічних завдань.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Робота виконана самостійно, академічного плагіату не виявлено, усі запозичення оформлено з відповідними посиланнями на джерела.

5. Ступінь оволодіння методами дослідження

При реалізації кваліфікаційної роботи студент проявив високий рівень компетентності та володіння сучасними інструментами, методами й технологіями комп'ютерних наук, зокрема з автоматичного розпізнавання мовлення, генеративних моделей і оцінки їхньої ефективності за допомогою метрик якості.

6. Повнота та якість розкриття теми роботи

Тема роботи повністю розкрита: проведено аналіз актуальності, здійснено огляд сучасних технологій обробки мовлення та резюмування, виконано всі поставлені завдання, а також розроблено концептуальну основу методу, що включає обробку аудіо за допомогою FFmpeg, транскрибування моделлю Whisper і резюмування моделлю BART. Метод підтверджено практичною реалізацією та оцінкою ефективності, що забезпечує його практичну цінність.

7. Логічність, послідовність, аргументованість, літературна грамотність викладення матеріалу

Викладення матеріалу логічне, послідовне та аргументоване. Мова і стиль роботи відповідають стандартам наукових текстів, забезпечуючи чіткість і доступність сприйняття. Структура роботи, що включає огляд технологій, проєктування методу та його програмну реалізацію, відповідає вимогам кваліфікаційних робіт.

8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин

Запропонований метод автоматичного резюмування може бути застосований у системах цифрової освіти, зокрема на веб-платформах університетів, для створення стислих резюме лекцій, вебінарів чи інших навчальних відеоматеріалів. Його компоненти, такі як обробка аудіо та генерація тексту, можуть бути адаптовані для інших інформаційних систем, що потребують автоматизації обробки відеоконтенту.

9. Висновок про можливість допуску кваліфікаційної роботи бакалавра до захисту, на яку оцінку заслуговує робота

*Враховуючи високий рівень виконання, повноту розкриття теми та дотримання всіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка – **«відмінно».***

Керівник _____ *к.т.н., доц. Руслан Багрій*

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ

Кафедра комп'ютерних наук

РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студента *гр. КН-21-2 Антонюка Юрія Олександровича*
за темою: *Метод автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами*

1. Актуальність обраної теми

В епоху стрімкого розвитку цифрової освіти обсяги навчальних відеоматеріалів сильно зростають. Це створює значне навантаження на здобувачів, яким доводиться витрачати велику кількість часу на їх перегляд. Використання методу автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами дозволяє суттєво оптимізувати процес засвоєння знань.

2. Повнота розкриття мети та завдань роботи

Під час виконання кваліфікаційної роботи бакалавра був реалізований метод автоматичного резюмування навчальних відеоматеріалів нейромережевими засобами, що відповідає меті та завданням кваліфікаційної роботи і розкриває їх повною мірою.

3. Зміст кожного розділу роботи

Записка кваліфікаційної роботи складається з трьох розділів. Перший розділ присвячено огляду технологій генеративного штучного інтелекту та аналізу сучасних рішень для розпізнавання мовлення та створення коротких резюме, а також формулює постановку задачі. Другий розділ містить опис проєктування методу автоматичного резюмування відеофайлів. Третій розділ розглядає особливості реалізації методу та результати тестування.

4. Оцінка розробленого методу та його практична цінність

Розроблений метод, що використовує нейромережеві засоби здатен ефективно виконувати свою основну функцію – створювати короткі і лаконічні резюме для відеофайлів, які в повній мірі передають оригінальний контекст. Це дозволяє суттєво зекономити час на перегляді освітніх відеоматеріалів.

5. Якість оформлення кваліфікаційної роботи бакалавра

Записка якісно оформлена відповідно до встановлених вимог, чітко і зрозуміло написана, зі структурованою побудовою розділів та логічною послідовністю викладення матеріалу.

6. Недоліки кваліфікаційної роботи бакалавра

Рекомендовано розглянути можливість автоматичного перекладу на українську мову, оскільки це зробить велику кількість освітніх матеріалів доступною для людей, що не володіють іноземною мовою.

7. Загальний висновок (допускається чи не допускається до захисту), та оцінка на яку заслуговує кваліфікаційна робота.

Враховуючи рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка відмінно.

Рецензент к.т.н., н. зав. кафедри Ю. М.