

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

Факультет інформатики та обчислювальної техніки

Кафедра інформаційних систем та технологій

«На правах рукопису»
УДК 004.932.4:006.9-047.31

До захисту допущено:
Завідувач кафедри
_____ Олександр РОЛІК
«__» _____ 2024 р.

**Магістерська дисертація
на здобуття ступеня магістра
за освітньо-професійною програмою «Інформаційне забезпечення
робототехнічних систем»
зі спеціальності 126 «Інформаційні системи та технології»
на тему: «Система комп'ютерного зору для обробки медіафайлів із
обличчям людини за допомогою нейронних мереж»**

Виконав:
студент 2 курсу, групи ІК-32мп
Басік Сергій Сергійович _____

Керівник:
доцент каф. ІСТ, к.т.н.,
Крилов Євген Володимирович _____

Рецензент:
доцент кафедри ІП, к.т.н., доц.
Лісовиченко Олег Іванович _____

Засвідчую, що у цій магістерській
дисертації немає запозичень з праць
інших авторів без відповідних
посилань.

Студент _____

Київ – 2024 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет інформатики та обчислювальної техніки
Кафедра інформаційних систем та технологій
Рівень вищої освіти – другий (магістерський)
Спеціальність – 126 «Інформаційні системи та технології»
Освітньо-професійна програма «Інформаційне забезпечення
робототехнічних систем»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Олександр РОЛІК

«___» _____ 2024 р.

ЗАВДАННЯ
на магістерську дисертацію студенту
Басіку Сергію Сергійовичу

1. Тема дисертації «Система комп'ютерного зору для обробки медіафайлів із обличчям людини за допомогою нейронних мереж», науковий керівник дисертації Крилов Євген Володимирович, к.т.н., доцент каф. ІСТ, затверджені наказом по університету від «08» 11 2024 р. № 5016-с
2. Термін подання студентом дисертації «09» 01 2024 р.
3. Об'єкт дослідження: система комп'ютерного зору для обробки медіафайлів із обличчям людини за допомогою нейронних мереж.
4. Вихідні дані: персоналізована система, що слідує правилам конфіденційності, аналізує мультимодальні дані на емоції та видає комплексний звіт за період часу у вигляді графіків.
5. Перелік завдань, які потрібно розробити: проаналізувати область об'єкту дослідження; проаналізувати існуючі рішення; сформулювати задачу роботи; спроектувати архітектуру системи; реалізувати інформаційну систему; розробити стартап-проект.
6. Орієнтовний перелік графічного (ілюстративного) матеріалу: діаграма прецедентів, діаграма діяльності, діаграми послідовності, діаграма компонентів, ER-діаграма, даталогічна модель, діаграма розгортання.

7. Дата видачі завдання 02.09.2024 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1.	Огляд та аналіз існуючих рішень	08.09.2024	
2.	Формування вимог до системи	15.09.2024	
3.	Вибір та обґрунтування технологій	29.09.2024	
4.	Розробка системи	06.10.2024	
5.	Тестування системи	21.11.2024	
6.	Розробка стартап-проєкту	29.11.2024	
7.	Подання роботи на передзахист		
8.	Подання роботи на захист		

Студент

Сергій БАСІК

Науковий керівник

Євген КРИЛОВ

РЕФЕРАТ

Система комп'ютерного зору для обробки медіафайлів із обличчям людини за допомогою нейронних мереж: 115 с., 34 табл., 28 рис., 1 дод., 23 джерела.

ШТУЧНИЙ ІНТЕЛЕКТ, НЕЙРОННІ МЕРЕЖІ, МУЛЬТИМОДАЛЬНА ОБРОБКА, XLM-ROBERTA, RESNET, АНАЛІЗ НА ЕМОЦІЇ, АНАЛІЗ ВІДЕО, АНАЛІЗ ЗОБРАЖЕНЬ.

Об'єктом дослідження є система комп'ютерного зору для обробки медіафайлів із обличчям людини за допомогою нейронних мереж.

Метою даної роботи є створення системи комп'ютерного зору, що забезпечує мультимодальний аналіз (міміка, мовлення) з використанням технологій штучного інтелекту. Предметом дослідження є мультимодальний підхід до аналізу людських емоцій в україномовних медіафайлах.

Дисертація включає вивчення існуючих рішень, визначення системних вимог, розробку функціональної моделі, проектування архітектури, реалізацію програмного прототипу, що працює в хмарному середовищі AWS з використанням сервісів для зберігання, обробки та мультимодального аналізу даних, а також тестування системи з метою перевірки її ефективності.

Результати дисертації можуть бути використані для підвищення точності діагностики емоційного стану в медичних установах, оптимізації обслуговування клієнтів у маркетингових компаніях, підвищення якості взаємодії співробітників у професійному середовищі шляхом аналізу їхніх емоцій, а також персоналізації допомоги в психологічних службах.

ABSTRACT

Computer vision system for processing media files with facial expressions using neural networks: 115 p., 34 tab., 28 draw., 1 app., 23 sources.

ARTIFICIAL INTELLIGENCE, NEURAL NETWORKS, MULTIMODAL PROCESSING, XLM-ROBERTA, RESNET, EMOTIONAL ANALYSIS, VIDEO ANALYSIS, IMAGE ANALYSIS.

The object of the study is a computer vision system for processing media files with a human face using neural networks.

The purpose of this work is to create a computer vision system, which provides multimodal analysis (facial expressions, speech) using artificial intelligence technologies. The subject of the study is a multimodal approach to the analysis of human emotions in Ukrainian-language media files.

The dissertation includes the investigation of existing solutions, defining of system requirements, development of a functional model, architecture design, implementation of a software prototype operating in the AWS cloud environment using services for storage, processing, and multimodal data analysis, as well as testing the system to verify its effectiveness.

The results of the dissertation can be used to increase the accuracy of emotional state diagnostics in medical institutions, optimize customer service in marketing companies, improve the quality of employee interaction in a professional environment by analysing their emotions, as well as personalize assistance in psychological services.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ ТА ПОЗНАЧЕНЬ	8
ВСТУП.....	10
1 АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ	12
1.1 Проблема та швидкість оцінювання виразу обличчя людьми	12
1.2 Опис існуючих рішень для відслідковування емоційного стану	14
1.3 Опис існуючих рішень на основі штучного інтелекту.....	18
Висновки до розділу 1	22
2 ПРОЕКТУВАННЯ СИСТЕМИ ОБРОБКИ МЕДІАФАЙЛІВ.....	24
2.1 Формування вимог до системи.....	24
2.2 Сценарії використання системи	25
2.3 Послідовність взаємодії з системою	30
2.4 Опис компонентів системи	32
2.5 Опис проєктованих станів системи.....	33
2.6 Опис сутностей системи	35
Висновки до розділу 2	36
3 ВИБІР ТА ОБҐРУНТУВАННЯ ЗАСОБІВ І ТЕХНОЛОГІЙ.....	37
3.1 Вибір хмарного постачальника	37
3.2 Вибір нейромереж та бібліотек	40
3.3 Вибір сховища даних в хмарному середовищі	43
3.4 Вибір сервісів AWS для забезпечення аналізу медіафайлів.....	47
3.5 Вибір сервісів хмарного середовища для безпеки та доступу з інтернету	50
3.6 Вибір мов програмування та фреймворків для поставленої задачі	52
Висновки до розділу 3	55
4 РЕАЛІЗАЦІЯ СИСТЕМИ ОБРОБКИ МЕДІАФАЙЛІВ.....	57
4.1 Реєстрація та авторизація користувача.....	57
4.2 Завантаження та обробка медіафайлів в систему.....	60
4.3 Навчання ResNet на FER+ наборі даних	63
4.4 Навчання XLM-RoBERTa на обробленому наборі даних GoEmotions	67

4.5 Аналіз медіафайлів використовуючи навчені моделі	71
4.6 Перегляд статистики аналізу обробки медіафайлів	75
Висновки до розділу 4.....	79
5 РОЗРОБЛЕННЯ СТАРТАП-ПРОЄКТУ	80
5.1 Опис ідеї проєкту.....	80
5.2 Технологічний аудит ідеї проєкту	82
5.3 Аналіз ринкових можливостей запуску стартап-проєкту.....	84
5.4 Розроблення ринкової стратегії проєкту	94
5.5 Розроблення маркетингової програми стартап-проєкту.....	98
Висновки до розділу 5.....	102
ВИСНОВКИ	103
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	105

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ ТА ПОЗНАЧЕНЬ

ВООЗ – Всесвітня організація охорони здоров'я;
ЗРЗД – Збори та розрахунки з державною податковою службою;
КПТ – когнітивно-поведінкова терапія;
ІІІ – штучний інтелект;
AdamW – Adam Optimizer with Weight Decay;
API – Application Programming Interface;
AUC – Area Under the Curve;
AWS – Amazon Web Services;
BERT – Bidirectional Encoder Representations from Transformers;
CK+ – Cohn-Kanade;
CNN – Convolutional Neural Networks;
COVID-19 – Coronavirus Disease 2019;
CRM – Customer Relationship Management;
DB – Database;
DynamoDB – Amazon DynamoDB;
EC2 – Elastic Compute Cloud;
EFS – Elastic File System;
FER – Facial Emotion Recognition;
FER – Facial Expression Recognition;
FER+ – Оновлений Facial Expression Recognition 2013 dataset;
FER2013 – Facial Expression Recognition 2013 dataset;
GCP – Google Cloud Platform;
GDPR – General Data Protection Regulation;
GPT-2 – Generative Pretrained Transformer 2;
GPU – Graphics Processing Unit;
HR – Human Resources;
HTTP – Hypertext Transfer Protocol;
IAM – Identity and Access Management;

ISO – International Organization for Standardization;
JAFFE – Japanese Female Facial Expression;
JPG/JPEG – Joint Photographic Experts Group;
JSON – JavaScript Object Notation;
MB – Megabyte;
MFA – Multi-Factor Authentication;
ML-фреймворки – Machine Learning Frameworks;
MP4 – MPEG-4 Part 14;
NN – нейронні мережі;
NoSQL – Not Only Structured Query Language;
NVIDIA T4 – графічний процесор NVIDIA;
OpenCV – Open Source Computer Vision Library;
RDS – Amazon Relational Database Service;
ResNet – Residual Network;
REST – Representational State Transfer;
RNN – Recurrent Neural Networks
S3 – Simple Storage Service;
SDK – Software Development Kit;
SGD – Stochastic Gradient Descent;
SMS – Short Message Service;
SQL Azure – Microsoft Azure Structured Query Language Database;
SWOT – Strengths, Weaknesses, Opportunities, and Threats;
TPU – Tensor Processing Unit;
UI – User Interface;
URL – Uniform Resource Locator;
VGGNet – Visual Geometry Group Network;
XLM-RoBERTa – Cross-lingual Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach.

ВСТУП

Зважаючи на зростаючу залежність від технологій, що призводить до потреби більшої емоційної та психологічної підтримки, потрібно, щоб послуги психічного здоров'я стали більш легкодоступними. На мою думку, це питання є надзвичайно важливим, адже як і підкреслює Всесвітня організація охорони здоров'я: «Немає здоров'я без психічного здоров'я» [1].

Психічне здоров'я турбує дедалі більше людей. За дослідженням 2022 року Всесвітньої організації охорони здоров'я, рівень стресу та тривожності зріс на 25% порівняно з попередніми роками [2]. На зростання цих факторів вплинула як і пандемія COVID-19, так і економічна нестабільність. В даній дисертації було досліджено як можливо позитивно вплинути на ментальне здоров'я та які дії для цього варто вжити.

В роботі було проаналізовано сучасний стан послуг психічного здоров'я, а саме консультації та терапію, як висновок вони часто не є доступними для середньостатистичної людини через високу вартість. На це вказує дослідження Verywell Mind, яке було проведено в 2022, результати показали що 31% респондентів перестали звертатись за психологічною допомогою і направили кошти в інше русло, адже оплата допомоги була занадто високою [3]. А за дослідженням 2019 року, яке опублікував журнал "The Lancet Psychiatry", клеймо та страх осуду сприяють уникненню звертання до спеціалістів, через це більшість людей не отримують своєчасну психологічну допомогу, проживаючи свої переживання наодинці, що впливає на їх подальше життя [4]. Якщо продовжити говорити про вплив на подальше життя, то дослідження журналу "The Lancet Psychiatry", яке велось під наглядом ВООЗ показало, що кожен долар, витрачений на лікування тривожних розладів або ж депресії, повертається у розмірі 4 доларів [5]. Причиною є покращення самопочуття, яке прямо впливає на працездатність.

За роботою "Deep Facial Expression Recognition: A Survey", опублікованою в 2018 році, яка досліджує розпізнавання емоції за виразом обличчя, стало відомо, що глибокі нейронні мережі можуть надати інформацію про емоційний стан

людини з високою точністю (“state-of-the-art recognition”) [6]. На мою думку, їхня робота була інноваційним рішенням, яке містить в собі великий потенціал та дозволить спеціалістам розвиватись у даній сфері глибше.

Нейронні мережі можуть аналізувати не тільки обличчя, але голос і текст, що да нам повну картину емоцій людини. Що з однієї сторони є великим технічним досягненням, а з іншої перед постає етичне питання. Система штучного інтелекту не робить суб'єктивні звіти, натомість емоції відстежуються об'єктивно без потреби в тому, щоб користувач передав дані.

Так як і світ змінюється швидко, так і емоція людини, може за декілька секунд суттєво змінитись. Тому важливо, щоб система працювала так, щоб потрібна підтримка була надана вчасно. З моєї особистої думки, це є особливо важливим в критично стресових або робочих ситуаціях, коли є лише обмежений час на заспокоєння та знаходження рівноваги. Отже, проаналізувавши все вище сказане, можна сказати що існуючі обмеження сучасних інструментів ментальної підтримки можна подолати, а саме шляхом надання персоналізованого зворотного зв'язку. Це буде працювати за допомогою запропонованої системи розпізнавання емоцій на основі штучного інтелекту.

Тобто, поєднання різного виду даних не тільки підвищує точність у визначенні емоцій користувача але й завдяки вбудованим датчикам може зробити так, що проактивна емоційна підтримка посилюється в потрібні моменти, шляхом втручання, ще до того як емоції вийдуть з-під контролю. Але після цього постає питання: як надана архітектура може вирішувати питання конфіденційності користувачів, а конкретно те, як конфіденційні дані, такі як запис голосу або зображення обличчя, отримуються та потім використовуються.

У дисертації буде продемонстровано дизайн та розробку системи з детальними варіантами її використання для імплементації її як і у особистому, так і у професійному житті. Технології штучного інтелекту прокладають шлях для масштабного та персоналізованого емоційного моніторингу. Це допоможе заповнити розрив між емоційною потребою людини та обмеженнями в традиційних сервісах психологічної допомоги.

1 АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ

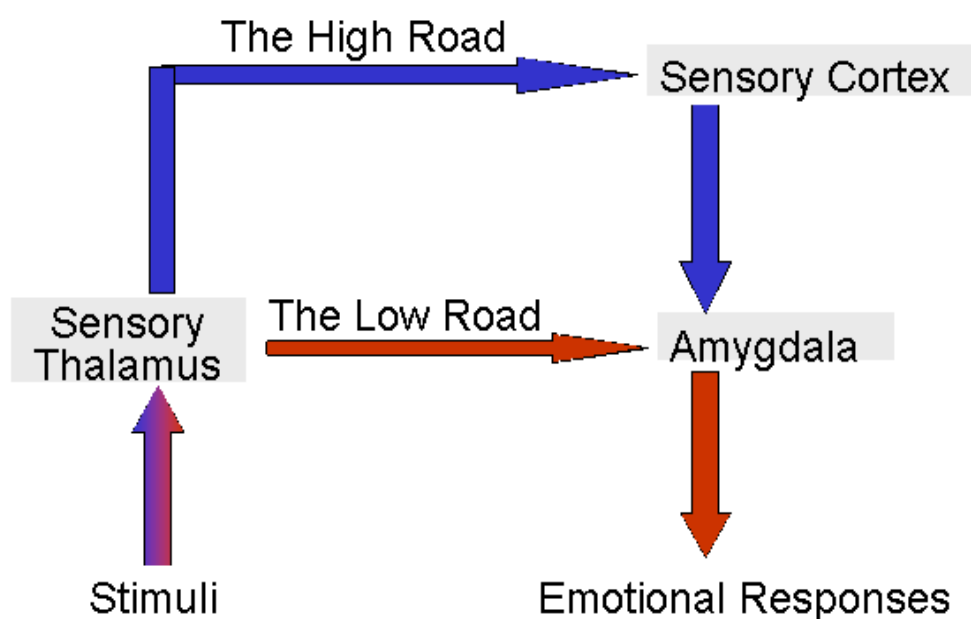
1.1 Проблема та швидкість оцінювання виразу обличчя людьми

Автор книги «Емоційний інтелект» Деніел Гоулман розрізняє два типи мозку: емоційний та раціональний (або, як він його ще називає, інтелектуальний) і описує в своїй книзі, що раціональний мозок це багато шарів навколо найпершого еволюційного людського мозку – емоційного [7]. Емоційний мозок в основному контролюється мигдалиною, яка сприяє швидкій та автоматичній обробці емоційних реакцій, і ця частина мозку реагує швидше, ніж неокортекс, відповідальний за раціональне мислення, що призводить до емоційних реакцій ще до того, як у мозку людини з'являться логічні міркування [8]. Тобто люди часто спочатку відчують емоції, а лише потім розуміють їх.

Розуміння та розпізнавання емоцій багато часу були перешкодами для дослідників, науковців та звичайних пересічних перехожих. Емоції відіграють глибоку роль в тому як люди взаємодіють між собою, приймають рішення і реагують на повсякденні труднощі. Проте навіть точна та коректна ідентифікація, та також подальша відповідь на емоції не завжди є прямолінійною. Однією із ключових особливостей є затримка між відчуттям емоції та як наслідок її обробкою. Як зазначає, психолог-публіцист, Гоулман, мозок людини обробляє емоційні стимули через два різних шляхи: один «короткий шлях», який включає в себе амігдалу (мигдалеподібне тіло в мозку), і «довгий шлях» – це коли емоція в кінцевому випадку оброблюється неокортексом (частиною мозку) [9]. Інколи ця різниця може бути критичною, адже емоції людина відчуває ще до того, як починає їх розуміти. Підтвердження цьому є також у медицині: американський невролог Джозеф Леду прямо займався цим питанням і підтвердив, що емоції (в особливості страх), інколи можуть переважати над раціональним мисленням людини [10]. Деталі зображено на рисунку 1.1.

Проблема усвідомлення емоції криється в швидких автоматичних природних та емоційних відповідях. Амігдала, яка відповідає за емоційні реакції, часто

отримує і обробляє сигнали швидше, аніж неокортекс – частина людського мозку, що відповідає за раціональне мислення та інтерпретацію подій в житті.



LeDoux: Tracing Emotional Pathways (NY Times Nov. 5, 1996)

Рисунок 1.1 – Два шляхи усвідомлення емоцій мозком людини [10]

Результатом є емоційна відповідь організму, яка відбувається ще до того, як людина це усвідомить, а саме, що відчуває це в своєрідний спосіб. Гоулман називає дане поняття емоційним піратством [8] – тобто коли амигдала ініціює миттєву реакцію на стимули, ще навіть до того, як людина сентиментально сприйме ситуацію. Це відставання може ускладнити розуміння своїх емоцій та ускладнити реакцію на ситуацію.

Наведемо приклад: в ситуаціях раптового гніву чи страху, людина може реагувати імпульсивно ще навіть до того, як оцінить контекст події. Емоційна реакція бере над нею гору та приводить її до відповідей, які далекі від раціонального мислення. Тому з'являється потреба в системі, яка б могла ідентифікувати та відслідковувати людські емоції, тим самим допомагаючи людям легше розпізнати та в разі потреби скорегувати свої емоційні відповіді.

Дискусія Гоулмана про роль мигдалеподібного тіла та неокортексу в обробці емоцій проливає нам світло на нейробіологічну основу того, чому люди часто відчують емоції ще до того, як їх усвідомлюють. Мигдалеподібне тіло людини отримує сенсорну інформацію швидше, ніж неокортекс, який дозволяє йому генерувати швидку емоційну реакцію. Цей «обхідний» шлях орієнтований на виживання і функціонує незалежно від свідомого мислення. Це пояснює чому люди відчують такі емоції, як страх, хвилювання або гнів, не відразу розуміючи, чому.

Швидкість реакції мигдалеподібного тіла особливо важлива в ситуаціях, коли на кону стоїть виживання. Однак у сучасному житті така швидка реакція може призвести до неадекватних емоційних реакцій, які не будуть відповідати реальній ситуації. Як приклад, хтось може відчути загрозу під час напруженої розмови на роботі, що викличе захисну емоційну реакцію, і лише згодом, після того, як неокортекс обробить ситуацію, зрозуміти, що небезпеки в дійсності не було.

Невідповідність між швидкістю емоційної та когнітивної обробки підкреслює необхідність втручання зовнішніх систем, таких як штучний інтелект, які можуть допомогти в регулюванні емоцій. Вони можуть виявляти ознаки емоційної реакції до того, як вони переростуть в ескалацію, що забезпечить важливий буфер, який, на жаль, людський мозок не завжди в змозі нам надати. Якщо говорити простими словами, системи штучного інтелекту допоможуть людям отримати кращий контроль над своїми емоціями.

1.2 Опис існуючих рішень для відслідковування емоційного стану

Daylio – це додаток для відстеження настрою, який використовує дані, згенеровані користувачами. Споживачі записують кожен день свої емоції та дії за допомогою попередньо встановлених категорій, створюючи візуальний запис тенденцій настрою з плином часу [11]. Перевагами Daylio є:

- можливість створювати власні мітки щодо настрою та своєї активності, пристосовуючи додаток до свого емоційного стану;

- у додатка простий інтерфейс, який вимагає лише кількох дотиків, щоб ввести запис до журналу, це дуже легко та не вимагає зусиль, тому це сприяє формуванню звички. Інтерфейс додатку зображено на рисунку 1.2;
- немає хмарного аналізу, тому всі дані залишаються на пристрої користувача, що забезпечує контроль над конфіденційністю.

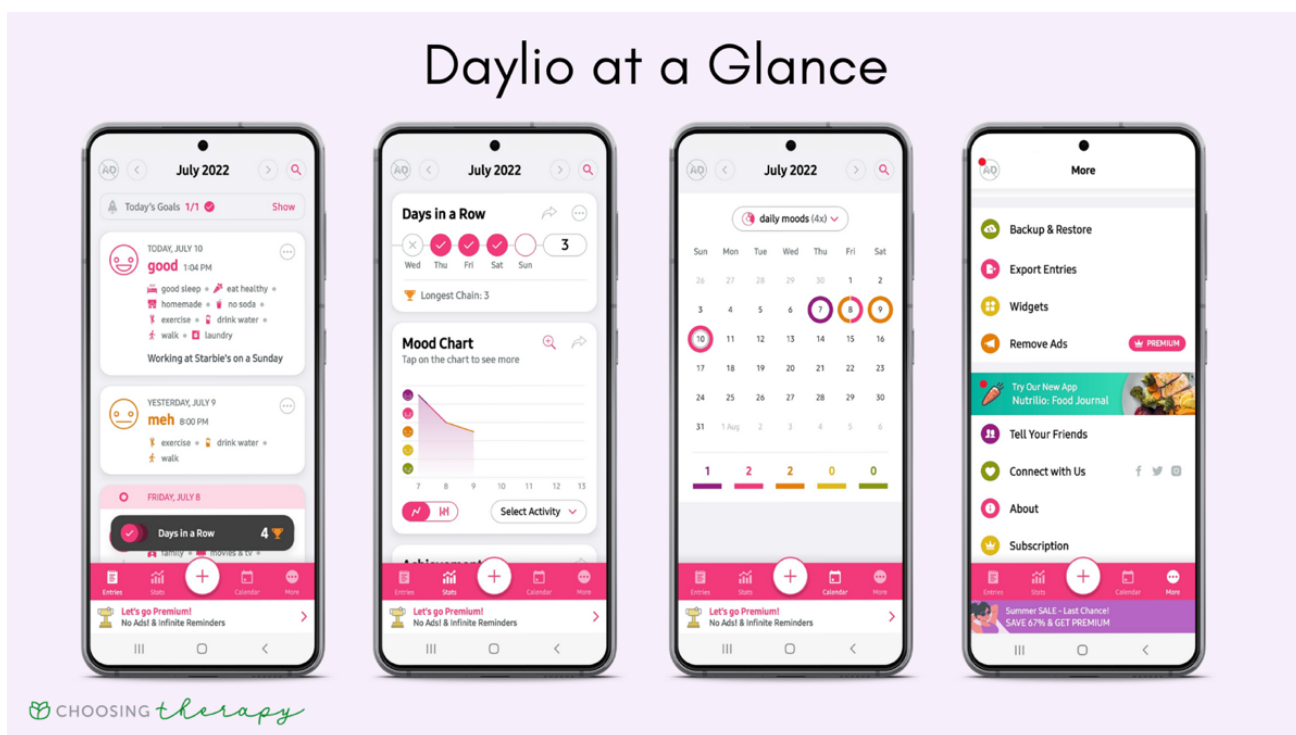


Рисунок 1.2 – Інтерфейс Daylio [11]

В той час як недоліками Daylio є нижче перелічені фактори:

- неточність реєстрування настрою або користувачі можуть забувати це робити, що призводить до неповноти даних;
- додаток повністю покладається на самооцінку користувача, не маючи можливості виявляти глибинні емоційні патерни;
- Daylio не надає рекомендацій чи аналітики, лише сирі тенденції, це обмежує його корисність для довгострокового зростання.

Наступним застосунком-аналогом є Moodnotes, додаток поєднує в собі відстеження настрою з когнітивно-поведінковими техніками. Користувачі додають

свій настрій, роблять записи в щоденнику та отримують поради, як переосмислити негативні думки [12]. Перевагами Moodnotes є:

- він побудований на принципах когнітивно-поведінкової терапії (КПТ), яка використовується спеціалістами в медичній сфері, це заохочує користувачів замислитися над моделями мислення та розпізнавати когнітивні викривлення;
- функція ведення щоденника допомагає користувачам краще зрозуміти свої емоції та тригери через саморефлексію;
- додаток також пропонує персоналізовані підказки, які допомагають користувачеві в кращому розумінні своїх емоцій.

Інтерфейс додатку Moodnotes зображений на рисунку 1.3.

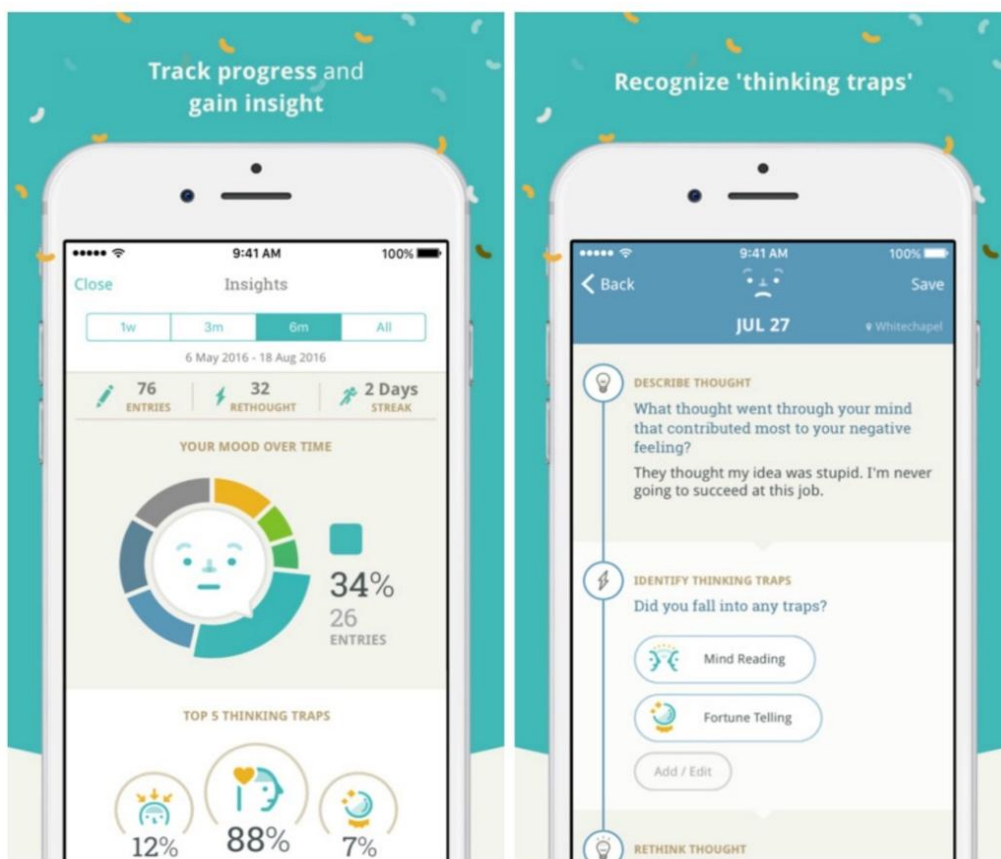


Рисунок 1.3 – Інтерфейс Moodnotes [12]

В той час як недоліками Moodnotes є нижче перелічені фактори:

- додаток вимагає більш активної участі користувачів, адже треба вести журнали, які може бути важко підтримувати;

- як і у випадку з Daylio, точність даних обмежується, а саме здатністю і бажанням користувача послідовно вносити записи до журналу;
- додаток не може визначати зміни в емоціях у реальному часі, він покладається лише на те, що користувач вирішує ввести.

Система комп'ютерного зору для обробки медіафайлів з обличчям людини, яка запропонована в дисертації, має значні переваги над традиційними методами відслідковування емоцій. Однією з головних переваг є здатність об'єктивно виявляти емоції, за допомогою аналізу тонких сигналів від тексту, голосу або зображення, адже на відміну від систем, де користувач вручну вводить дані, система не залежить від свідомості споживача або від його бажання записувати емоції. Також важливо, що за допомогою мультимодального вводу, система надає повнішу картину емоційного стану людини, навіть фіксуючи такі нюанси, які можуть бути пропущені одно модальними крекерами вводу. Окрім того, процес аналізу є автоматизованим, що сприяє наданню дієвих висновків та зворотного зв'язку, які з часом будуть адаптуватися до емоційних патернів. Процес адаптування зменшить когнітивне навантаження на користувача та забезпечить послідовне відстеження емоцій без необхідності вручну вводити дані про свій стан.

Але ці переваги супроводжуються певними викликами, адже збір та обробка особистих даних (міміка та записи голосу), можуть викликати занепокоєння щодо конфіденційності даних та того, як ця інформація зберігається та використовується.

Наступним викликом є висока вартість розробки та розгортання системи на основі штучного інтелекту. Дорожче коштують особливо ті системи, які можуть працювати в режимі реального часу, якщо порівнювати їх з більш простими програмами, в яких використовується ведення журналу. Витрати можна пом'якшити під час тестування, але для масштабних застосунків вони залишаються важливим фактором.

Отже, якщо порівнювати систему комп'ютерного зору з традиційними додатками для відслідковування емоційного стану, то додатки як Daylio і Moodnotes покладаються лише на дані користувачі, що може сприяти упередженості самозвіту. В той час як система на основі штучного інтелекту є

більш об'єктивною та послідовною, адже може одночасно аналізувати і текст, і голос, і зображення, а це допомагає краще зрозуміти емоційний стан людини та провести всебічний моніторинг.

1.3 Опис існуючих рішень на основі штучного інтелекту

З тим, як швидко технології інтегруються у повсякденне життя, зростає потреба у розробці систем на основі штучного інтелекту, які б могли розпізнавати емоції. Людство і так все частіше та частіше взаємодіє зі штучним інтелектом, смарт пристроями або диджиталізованими платформами, тому вкрай необхідно, щоб ці системи могли розуміти емоційні сигнали про стан людини та вміли реагувати на них. Без цього у взаємодії людини з комп'ютером буде бракувати емпатії, а це є важливим аспектом ефективної комунікації.

Якщо говорити про сферу обслуговування, де задоволення клієнта може різко змінюватись, то розпізнавання емоції за допомогою такої системи допомогло б ідентифікувати емоції та можливе незадоволення ще до того, як гість їх відкрито повідомить, що призведе до вищого рівня обслуговування у різних сферах життя.

У сфері охорони здоров'я системи засновані на штучному інтелекті відстежують ментальний стан пацієнтів та надають звіти про зміни та можливі психічні розлади (такі як тривога та депресія) ще до того, як лікар це здатен помітити. Відстеження не обмежуються лише професійною сферою, адже у щоденному спілкуванні люди також схильні відчувати раптові зміни у настрої, що підкреслює важливість швидкого розпізнавання емоційного стану та реакції на нього, адже людина інколи не здатна їх помітити.

Штучний інтелект має потенціал у відслідковуванні і аналізі емоцій постійно, що дозволить швидше ідентифікувати емоційні зрушення. Це є вирішальним в ситуаціях під великим тиском та стресом, під час взаємодій з іншими людьми або навіть при перегляді стрічки соціальних мереж, адже під час цього людські емоції викидаються моментально.

Системи на основі штучні інтелекту навіть здатні вловлювати найдрібніші зміни в міміці, голосі, інтонації та за мовою тіла, надаючи оновлення в реальному часі щодо ментального стану людини. За допомогою ШІ можна передбачати емоційні патерни та пропонувати відповідну інтервенцію, що значно покращить емоційну регуляцію та правильність прийнятих рішень у будь-яких сферах застосування.

Affectiva – це новітня технологія штучного інтелекту, яка аналізує міміку та для аналізу емоційного стану людини. Вона використовується в рекламі, в маркетингових дослідженнях та навіть в автомобільній безпеці [13]. Основні плюси:

- Affectiva одночасно обробляє і міміку, і голос, що надає більш детальну інформацію про емоції користувача;
- використовується в комерційних цілях, а саме для збору зворотного зв'язку від клієнтів, на рисунку 1.4 зображена інформаційна панель для тестування реклами;
- штучний інтелект Affectiva навчається на великому обсязі даних, саме тому вона здатна виявити тонкі емоційні нюанси в різних групах населення.

Недоліки Affectiva описані нижче:

- застосування Affectiva в таких сферах, як маркетинг, фокусується на миттєвих емоційних реакціях, залишаючи поза увагою довгострокові емоційні патерни або глибокий вплив на особистість;
- як і у випадку з будь-яким ШІ, що аналізує персональні дані (особливо відео та голос), користувачі можуть відчувати дискомфорт від збору даних, особливо в корпоративному або публічному середовищі, що може впливати на продуктивність та довіру до компанії в цілому.

Таким чином, Affectiva є ефективним інструментом для детального аналізу емоцій, але важливо враховувати етичні аспекти її використання, зокрема забезпечення конфіденційності та безпеки даних. Баланс між ефективністю, прозорістю та етикою сприятиме підвищенню довіри до таких передових технологій.

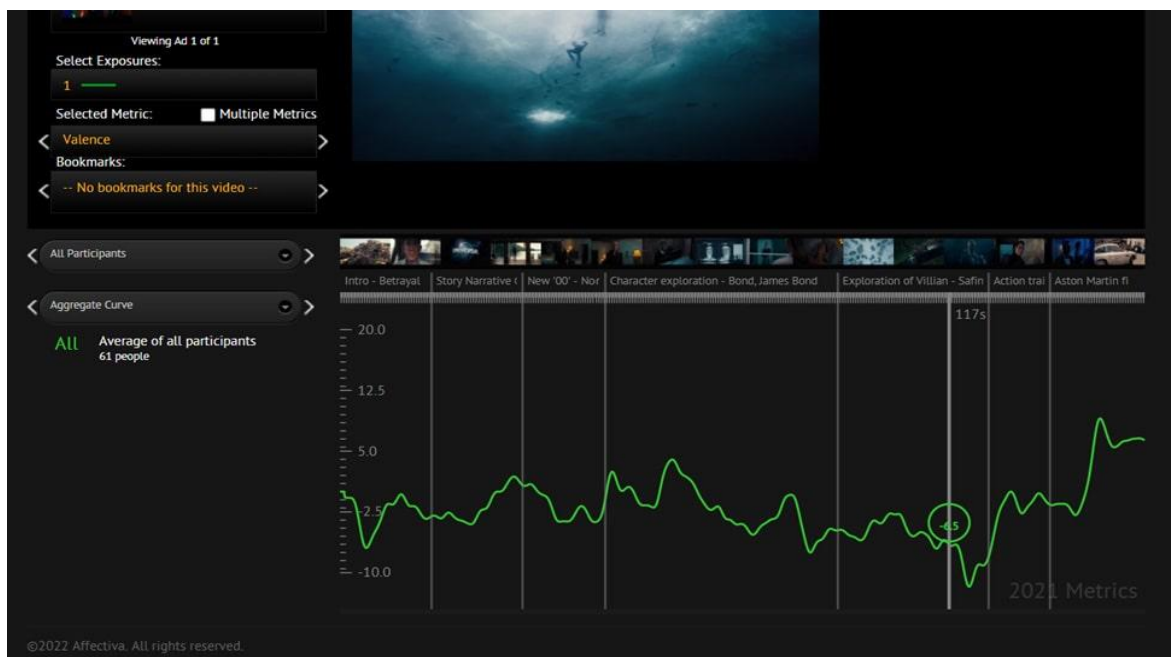


Рисунок 1.4 – Affectiva Media Analytics інформаційна панель для тестування реклами [13]

Replika – це чат-бот із вбудованим штучним інтелектом для надання емоційної підтримки. Цей помічник спілкується текстово з користувачем, розпізнає емоційні сигнали та надає емпатичні відповіді (наприклад співчутливо) [14]. Приклад спілкування з чат-ботом можна побачити на рисунку 1.5.

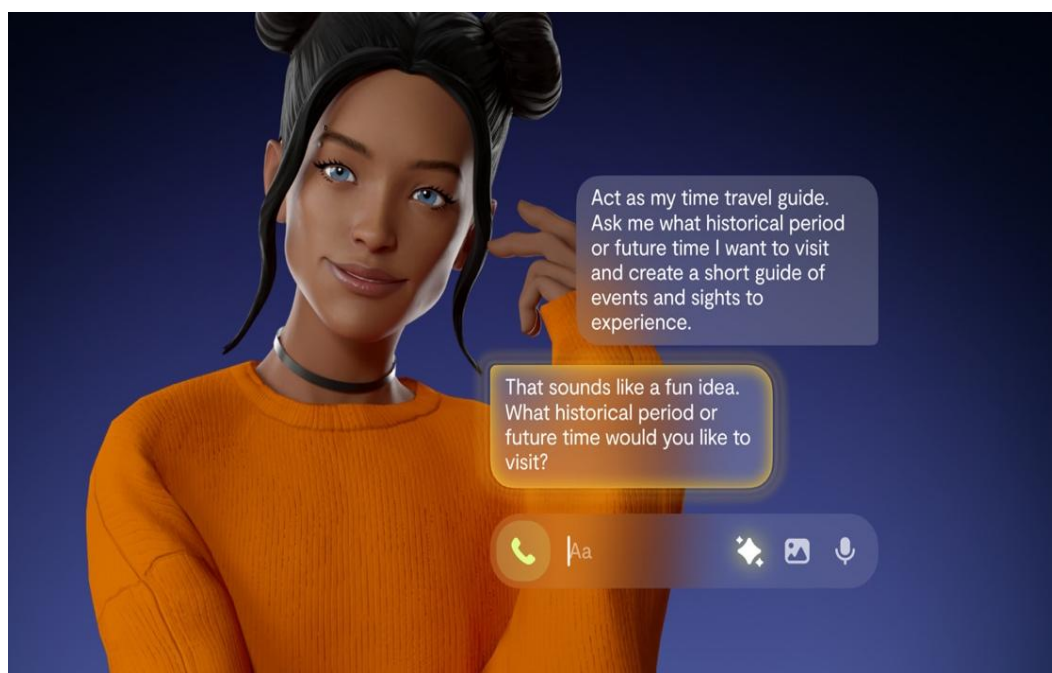


Рисунок 1.5 – Приклад спілкування з чат-ботом Replika [14]

Переваги чат-боту Replika із вбудованим штучним інтелектом:

- Replika може розпізнавати емоційний стан через текст та надавати відповіді, які допомагають користувачеві боротися зі стресом, тривогою та почуттям самотності;
- штучний інтелект покращує свої навички, тобто і емоційні реакції, адаптується до індивідуальних особливостей та стилів спілкування кожного користувача;
- користувач може вільно говорити про свої почуття, адже ШІ не проявляє осуду, що допомагає емоційно розвантажитись.

В той час як недоліками Replika є нижче перелічені фактори:

- Replika може аналізувати лише текст, а не міміку або тон голосу, що суттєво обмежує її здатність глибоко та точно проаналізувати емоційний стан людини;
- хоча Replika може розпізнавати слова та фрази, але не розуміє складні емоційні стани, тому відповіді з часом можуть повторюватись або здаватись поверхневими.

Система відслідковування виразу обличчя людини на основі ШІ, запропонована в цій дисертації, поєднує в собі переваги мультимодальних даних, інтегруючи міміку, голос і текст, щоб запропонувати глибокий огляд емоційного стану людини. На відміну від існуючих додатків, таких як Replika, яка здатна аналізувати лише текст, або Affectiva, яка фокусується на конкретних комерційних випадках, система, запропонована у дисертації, забезпечує більш цілісне, глибоке розуміння емоцій. Ключовою перевагою є надання персоналізованої інформації, а не узагальнений емоційний аналіз. Це дозволяє отримувати більш релевантний, контекстно-специфічний зворотній зв'язок, що покращить точність емоційного відстеження.

Якщо розглядати запропоновану систему з точки зору конфіденційності, то система розроблена з архітектурою, що дає гарантію, що особисті дані обробляються локально та безпечно в хмарі. Це важливо для запобігання проблем, які виникають під час комерційних рішень розпізнавання емоцій. Навіть більше

того, система виходить за рамки простого реактивного аналізу та пропонує проактивну емоційну підтримку. Вона передбачає емоційні патерни та надає контекстні рекомендації (у моменти стресу).

Однак з'являється складність у побудові системи, яка може обробляти кілька модальностей та надавати персоналізований зворотній зв'язок, адже вона є значно більшою, складнішою та ресурсоємнішою ніж одномодальні системи, як Replika наприклад.

Для створення цінних емоційних інсайтів, запропонованій системі потрібен доступ до великої кількості даних та протягом тривалого часу. Це все для того, щоб система могла адаптуватися до індивідуальних емоційних патернів та зрозуміти кожного користувача. Якщо говорити про існуючі на ринку системи, то хоч вони ефективні у своїх сферах, але їхні можливості обмежені. Affectiva здатна розпізнавати емоції в реальному часі для комерційного використання, але в ній немає персоналізації, Replika може проявляти емпатію, але лише в текстових розмовах, тобто вона не здатна аналізувати інші емоційні сигнали. На відміну від цих систем, запропонована система комп'ютерного зору може інтегрувати мультимодальні дані та надавати персоналізовані інсайти. Це робить інструмент, запропонований в дисертації, більш універсальним та кращим у довгостроковому використанні.

Висновки до розділу 1

Отже, розуміння емоцій – це складний процес, адже він вимагає балансу між миттєвими реакціями та довгостроковими рішеннями. Системи на основі штучного інтелекту можуть допомогти це вирішити: забезпечити глибший та кращий емоційний моніторинг та надати своєчасний і персоналізований зворотній зв'язок користувачам. Так як технології штучного інтелекту швидко розвиваються, стає більш можливим використання цих систем щодня у повсякденних справах. Емоційний інтелект потроху стає важливим фактором у особистому та

професійному житті, що робить систему для розпізнавання емоцій та реакції на них не лише актуальною, але й необхідною.

Був проведений аналіз існуючих підходів та рішень – вони мають як і певні переваги, так і певні недоліки. Додатки Daylio і Moodnotes пропонують користувачеві самостійне введення даних про емоційний стан, але саме через ручне введення звіти можуть бути неповними та упередженими. Системи на основі штучного інтелекту (Replika та Affectiva) використовують передові алгоритми та роблять це у реальному часі. Проте Replika здатна аналізувати лише текст, в той час як Affectiva не здатна ефективно враховувати індивідуальні особливості кожної людини.

Тому рішення запропоноване в цій дисертації, направлене на усунення недоліків, а саме шляхом: мультимодальності (здатність аналізувати вираз обличчя, тон голосу та текст), адже це дасть кращу та повнішу картину емоційного стану. Так система усуває потребу в записування емоції та допомагає краще виявляти емоції та в подальшому надає проактивну емоційну підтримку. Наступними перевагами є врахування питань конфіденційності та адаптивність до кожного користувача. Та найважливіше, що система на основі комп'ютерного зору буде корисна як і у особистому, так і у професійному житті.

2 ПРОЕКТУВАННЯ СИСТЕМИ ОБРОБКИ МЕДІАФАЙЛІВ

2.1 Формування вимог до системи

Для забезпечення успіху системи обробки медіафайлів на основі штучного інтелекту дуже важливо визначити конкретні цілі, що будуть впливати на архітектуру та продуктивність системи. Розроблений проєкт має завдання видавати результат аналізу емоцій людей в різних виглядах (графіки чи таблиці), обробляючи численні вхідні дані. Вхідними даними можуть бути зображення з чи відео з обличчям людини. Враховуючи характер застосування, особлива увага приділятиметься балансуванню між точністю і ефективністю системи. Конфіденційність такої системи також виноситься як окрема вимога. Крім того, дизайн системи повинен бути адаптивним та відповідати різним середовищам, як особистим, так і професійним. Нижче наведені функціональні вимоги, що відповідають цим цілям:

- система повинна надавати можливість створювати особистий кабінет з можливістю перегляду та зміни особистої інформації;
- система повинна аналізувати емоційний стан людини по фото;
- система повинна аналізувати емоційний стан людини по відео;
- система повинна мати інформаційну панель по аналізу за обраний період зображуючи інформацію у вигляді графіків та кругових діаграм;
- система повинна мати історію емоційних станів у вигляді табличного списку емоцій з прив'язкою до днів.

Наступні нефункціональні вимоги забезпечують ефективну роботу системи емоційного розпізнавання, зосереджуючи увагу на її продуктивності, масштабованості та безпеці, щоб відповідати вимогам різноманітних реальних застосувань:

- система повинна мати валідацію введених даних та захищеність облікового запису користувача;
- система повинна бути відмовостійка, зберігати інформацію в хмарі балансуючи навантаження між сервісами;

- система повинна мати швидку відповідь від серверу (максимальний час очікування відповіді 3 секунди);
- система повинна мати точність аналізу емоційного стану із зображення з обличчям людини більше як 75%;
- система повинна мати точність аналізу емоційного стану із відео з обличчям і розмовою людини більше як 85%;
- система повинна бути проста у використанні;
- система повинна мати можливість легкого масштабування.

2.2 Сценарії використання системи

У цьому розділі досліджуються різні сценарії використання системи обробки медіафайлів на основі ІІІ, щоб продемонструвати її практичне застосування в різних середовищах. У цих сценаріях буде показано, як систему можна інтегрувати в особисті та професійні налаштування. Аналізуючи конкретні контексти, такі як управління стресом на робочому місці або персоналізований моніторинг психічного здоров'я, буде підкреслено універсальність і адаптивність запропонованого рішення. Візуально розглянути діаграму прецедентів не вдаючись у деталі можна у додатку А. Кожен із цих варіант використання показує як можна покращити процес виявлення емоцій в мультимедійних даних з різних джерел, а згодом це приведе до покращення процесу прийняття рішень, покращення емоційної регуляції та надання цінного зворотного зв'язку користувача. Нижче знаходиться перелік таблиць, які містять детальні описи сценаріїв.

У таблиці 2.1 наведено опис сценарію, коли користувач входить в систему.

Таблиця 2.1 – Сценарій “Вхід до системи”

Назва	Вхід до системи
Актори	Користувач, Система
Передумови	Користувач зареєстрований. Користувач знає свої логін та пароль.
Результат	Користувач успішно входить у систему. Користувач отримує доступ до персоналізованої панелі або функцій застосунку.

Назва	Вхід до системи
Порядок подій	1) Користувач переходить на сторінку входу. 2) Користувач вводить свій логін і пароль. 3) Користувач натискає кнопку 'Вхід'. 4) Система перевіряє введені дані. 5) Система входить у застосунок від імені користувача. 6) Користувач перенаправляється на головну сторінку або панель.
Альтернативні події	1) Якщо логін або пароль введено невірно, система показує повідомлення про помилку. Користувач може повторити спробу входу. 2) Забутий пароль – користувач може зв'язатися з адміністратором.

У таблиці 2.2 наведено опис сценарію, коли користувач реєструється в системі, з порядком подій та альтернативними розвитками подій.

Таблиця 2.2 – Сценарій “Реєстрація”

Назва	Реєстрація
Актори	Користувач, Система
Передумови	Користувач бажає створити новий обліковий запис.
Результат	Користувач успішно зареєстрований у системі. Користувач отримує доступ до функцій застосунку як зареєстрований користувач.
Порядок подій	1) Користувач переходить на сторінку реєстрації. 2) Користувач заповнює форму реєстрації, вводячи свої дані (ім'я, електронна пошта, пароль). 3) Користувач натискає кнопку 'Зареєструватися'. 4) Система перевіряє правильність введених даних. 5) Система створює новий обліковий запис. 6) Користувач отримує підтвердження реєстрації та доступ до застосунку.
Альтернативні події	1) Якщо користувач ввів неправильні або неповні дані, система показує повідомлення про помилку. Користувач виправляє дані і повторно відправляє форму. 2) Якщо введена електронна пошта вже зареєстрована, система показує повідомлення і просить змінити введені дані.

У таблиці 2.3 наведено сценарій перегляду інформаційної панелі користувачем.

Таблиця 2.3 – Сценарій “Перегляд інформаційної панелі”

Назва	Перегляд інформаційної панелі
Актори	Користувач, Система
Передумови	Користувач успішно увійшов у систему. Користувач має зібрані дані про емоційний стан за певний період.
Результат	Користувач переглядає статистику своїх настроїв у вигляді діаграм. Користувач може вибрати період для аналізу даних.
Порядок подій	1) Користувач переходить на інформаційну панель після входу в систему. 2) Користувач вибирає період часу для перегляду даних про емоції. 3) Система відображає графік: вісь X – дні, вісь Y настрої за день. 4) Система відображає кругову діаграму всіх настроїв за обраний період. 5) Система відображає додаткові статистичні дані, такі як середній настрій за період і кількість змін настрою протягом дня.
Альтернативні події	1) Якщо за обраний період немає зібраних даних, система відображає повідомлення про відсутність даних та надає можливість вибрати інший період для перегляду. 2) Якщо інтернет-зв'язок слабкий, система може не завантажити дані вчасно, тоді користувач отримує повідомлення про затримку і може спробувати ще раз. 3) Якщо система недоступна або дані не можуть бути отримані, користувач отримує повідомлення про помилку і її контекст.

У таблиці 2.4 наведено сценарій відображення профілю користувача та доступу до нього.

Таблиця 2.4 – Сценарій “Перегляд профілю користувача”

Назва	Перегляд профілю користувача
Актори	Користувач, Система
Передумови	Користувач успішно увійшов у систему. Користувач має зібрані дані про емоційний стан за певний період.
Результат	Користувач переглядає свій профіль і змінювати особисті дані. Користувач може переглядати свій настрій і видалити всі дані.
Порядок подій	1) Користувач переходить на сторінку профілю після входу в систему.

Назва	Перегляд профілю користувача
	2) Користувач переглядає особисті дані: ім'я, прізвище, номер телефону. 3) Користувач за бажанням змінює ім'я, прізвище або номер телефону, натискаючи кнопку 'Редагувати'. 4) Користувач переглядає свій поточний показник настрою. 5) Система відображає переважаючий настрій за останній місяць. 6) Користувач може натиснути кнопку 'Видалити всі дані', щоб очистити всю інформацію.
Альтернативні події	1) Якщо користувач вводить дані у неправильному форматі, система показує повідомлення про помилку та надає можливість виправити дані та повторно відправити форму. 2) Якщо користувач не бажає змінювати особисті дані, він може просто переглянути та закрити профіль.

У наступній таблиці 2.5 наведено перегляд історії емоційного стану користувача.

Таблиця 2.5 – Сценарій “Перегляд історії обробки медіафайлів”

Назва	Перегляд історії обробки медіафайлів
Актори	Користувач, Система
Передумови	Користувач успішно увійшов у систему. Користувач має зібрані дані про емоційний стан.
Результат	Користувач переглядає історію своїх настроїв у вигляді таблиці. Користувач може видаляти дані за окремі дні.
Порядок подій	1) Користувач переходить на сторінку історії настрою. 2) Користувач бачить таблицю з історією настрою, де кожен рядок містить: дату, оцінку настрою на основі оброблених зображень із відео та тексту. 3) Для детального перегляду інформації, користувач може перейти на окрему сторінку із аналізом для зображення чи відео(на сторінці з відео також можна переглянути аналіз тексту перейшовши у відповідну вкладку). 4) Користувач може натиснути кнопку “Ігнорувати дані під час збору статистики”.
Альтернативні події	1) Якщо за вибраний період немає даних: система відображає повідомлення 'Дані відсутні за вказаний період' та надає можливість змінити період для перегляду. 2) Якщо система має неповні дані, система показує повідомлення про неповноту даних, але все одно відображає наявну інформацію.

У таблиці 2.6 наведений сценарій як користувач завантажує фото або відео для оцінки свого емоційного стану.

Таблиця 2.6 – Сценарій “Завантаження медіафайлу для обробки та аналізу”

Назва	Завантаження медіафайлу для обробки та аналізу
Актори	Користувач, Система
Передумови	Користувач успішно увійшов у систему. Користувач має медіафайл у потрібному форматі.
Результат	Користувач успішно завантажує медіафайл для обробки і аналізу. Система зберігає матеріал та пов'язує його з вибраною датою та користувачем.
Порядок подій	1) Користувач переходить на сторінку завантаження медіафайлу. 2) Користувач обирає дату запису відео чи створення фото. 3) Користувач завантажує медіафайл відповідного формату (відео MP4, фото JPG/JPEG) і розміру (відео до 300 MB, фото до 10 MB). 4) Користувач натискає кнопку 'Завантажити'. 5) Система перевіряє розмір і формат відео (якщо медіафайл відповідає вимогам, система завантажує його для подальшої обробки та аналізу настрою) та відображає підтвердження успішного завантаження. 6) Автоматично після завантаження медіафайлу починається процес обробки, котрий включає в себе: вирізання кадрів із відео, обирання та аналіз кадрів з обличчям людини, транскрипція мовлення та аналіз сегментів на сентиментальність та запис усіх даних до постійного сховища.
Альтернативні події	Якщо користувач завантажує відео у невірному форматі або з перевищенням розміру: – система показує повідомлення про помилку (наприклад, 'Невірний формат відео' або 'Перевищено розмір файлу'); – користувач вибирає правильний файл та повторно завантажує його; – якщо користувач не вибрав дату: система відображає повідомлення 'Виберіть дату перед завантаженням відео'.

Сценарії використання описують систему комп'ютерного зору обробки медіафайлів через призму можливостей взаємодії користувача: реєстрація, вхід, завантаження файлів та перегляд оброблених та проаналізованих даних.

2.3 Послідовність взаємодії з системою

Цей підрозділ висвітлює, які дії та у якій послідовності користувач виконує для виконання поставленого завдання щодо обробки зібраних мультимодальних даних. Нижче буде описано кілька схем взаємодії користувача і системи, а точніше: послідовність при реєстрації чи входу в систему, послідовність дій при бажанні отримати аналіз медіаданих, а також послідовність дій для завантаження медіафайлів до системи для майбутньої обробки.

На схемі (рисунок 2.1) показано процес автентифікації користувача для доступу до системи. Він починається з того, що користувач відкриває сторінку входу. Незареєстрований користувач проходить процес реєстрації, перш ніж повернутися на сторінку входу. Після введення своїх облікових даних система перевіряє їх. Правильні дані дають доступ до інформаційної панелі та ресурсів. Якщо дані не вірні, буде запропоновано повторити спробу. Вичерпання кількості невдалих спроб до повідомлення скинути пароль через службу підтримки.

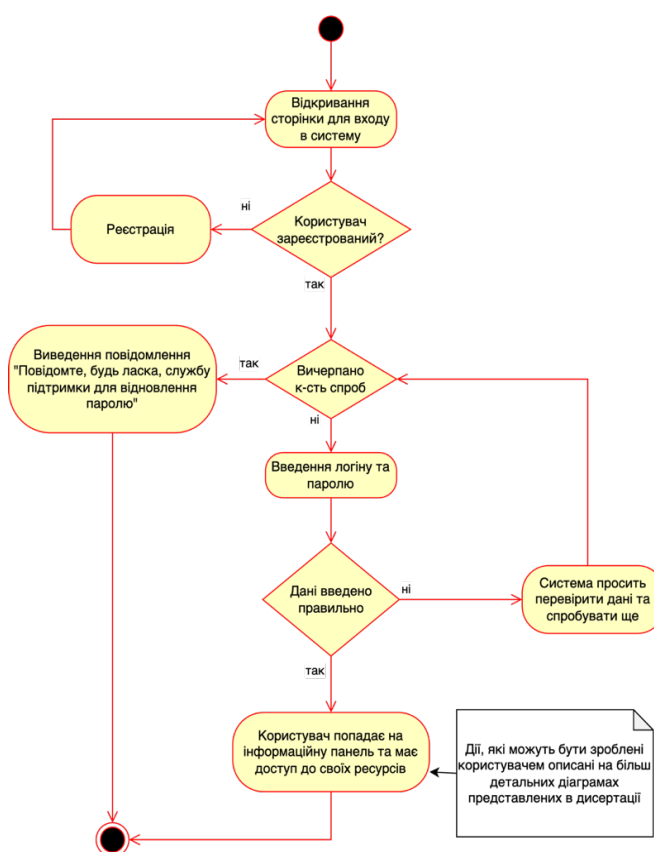


Рисунок 2.1 – Процес автентифікації користувача

Також у додатку Б діаграма активності описує процес взаємодії користувача з інформаційною системою для перегляду інформації та управління медіафайлами. Спочатку користувач обирається спосіб перегляду інформації: інформаційна панель або історія емоційних станів. Якщо обрано інформаційну панель, користувачу надається доступ до відповідної сторінки. Перевіряється, чи достатньо даних для відображення у вигляді графіків. У випадку недостатньої кількості даних, користувачу пропонується перейти на сторінку завантаження медіафайлу, де він може завантажити відео або зображення та обрати дату для асоціації. Після завантаження дані валідуються. Якщо дані вважаються валідними, користувач може переглянути конкретний медіафайл та його аналіз, а також змінити або видалити його за потреби. У разі невдачі валідації, система просить перевірити дані та спробувати ще раз. Якщо обрано перегляд історії станів, перевіряється достатність відображуваних даних. За недостатності даних користувачу пропонується завантажити медіафайл, після чого проводиться валідація. У випадку валідних даних користувач може переглянути та керувати конкретними медіафайлами. Повернення до попереднього інтерфейсу здійснюється після завершення необхідних дій. Завершальні точки процесу відображаються у вигляді кінцевих станів, які завершують взаємодію користувача з системою.

Діаграма послідовності з усіма деталями щодо завантаження файлів у систему зображена у додатку В. Процес починається з ініціації запиту на збереження відеофайлу користувачем. Цей запит передається безсерверній функції, яка відповідає за генерацію посилання для завантаження. Посилання генерується за допомогою сховища об'єктів і надається користувачу. Користувач використовує надане посилання для завантаження відеофайлу безпосередньо до сховища об'єктів. Після успішного завантаження зі сховища об'єктів надсилається сповіщення про завершення процесу завантаження. Це сповіщення ініціює виклик іншої серверної безсерверної функції, яка відповідає за обробку відеофайлу відповідно до вимог завдання, таких як кадрування або кодування. Після обробки оброблене відео зберігається у сховищі об'єктів, а користувач отримує сповіщення про успішне завантаження та обробку відеофайлу.

Ці діаграми необхідні для візуального відображення процесів, робочих циклів і взаємодій системи. Ці діаграми дають краще розуміння системи зацікавленими сторонами. Вони описують реалізацію складних функціональних можливостей і орієнтують для майбутнього обслуговування та вдосконалення.

2.4 Опис компонентів системи

Цей розділ описує основні компоненти та їх взаємодію для забезпечення ефективної роботи всієї системи надаючи користувачу кінцевий результат після обробки медіаданих з різних джерел. Нижче на рисунку 2.2 зображено абстрактні компоненти проєктованої системи.

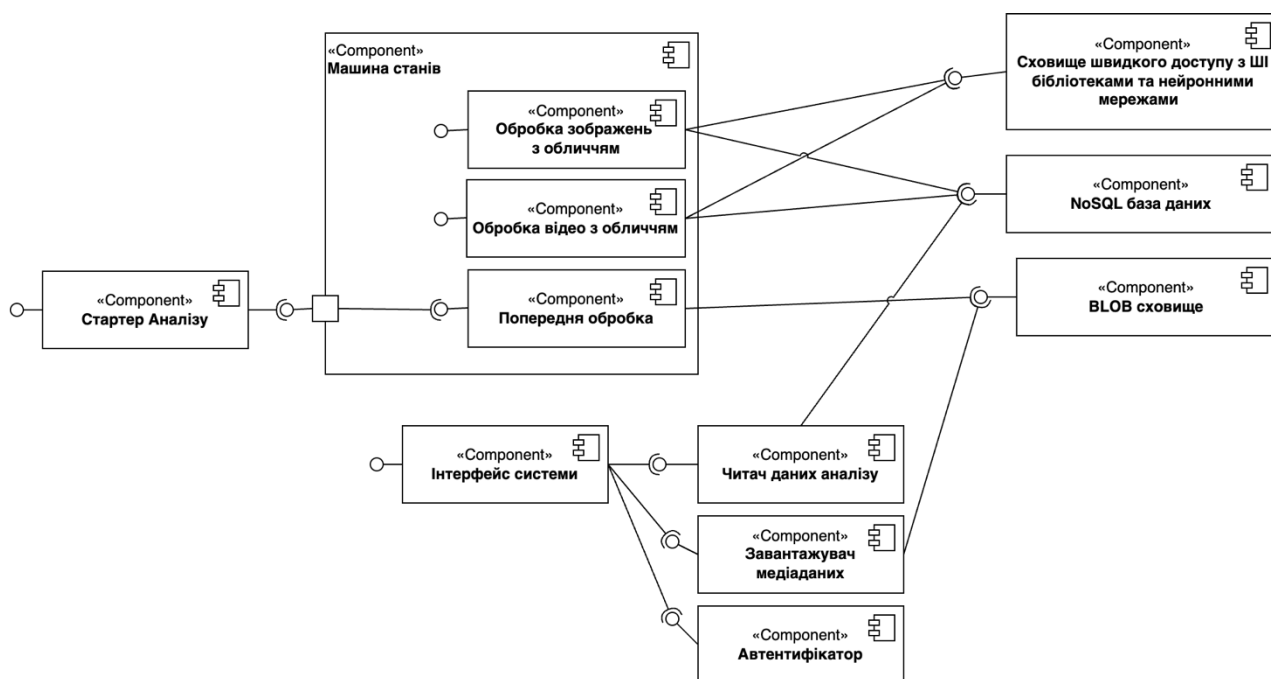


Рисунок 2.2 – Діаграма компонентів проєктованої системи

У центрі діаграми розміщено групу компонентів, що відповідають за обробку вхідних зображень та відео з обличчям. Серед них: «Машина станів» контролює послідовність етапів, «Обробка зображень з обличчям» та «Обробка відео з обличчям» відповідають за розпізнавання облич на вхідних даних, а «Попередня обробка» виконує базову підготовку даних для аналізу.

В окремих групах представлені «Стартер Аналізу» для ініціалізації процесу аналітики, а також «Інтерфейс системи», через який зовнішні користувачі чи процеси взаємодіють із системою. «Читач даних аналізу» забезпечує доступ до результатів аналітичних операцій. А «Завантажувач медіаданих» відправляє автентифіковані запити на збереження медіаданих у спеціальному сховищі великих бінарних даних. Дані, що відносяться до основних сутностей системи, зберігаються у «NoSQL базі даних», «BLOB сховище» для необроблених великих файлів, та «Сховище швидкого доступу з III бібліотеками та нейронними мережами» для прискореного доступу до моделей. «Автентифікатор» керує доступом до цих ресурсів.

Компоненти зв'язані лініями, що відображають потоки даних і викликів. Потоки від обробних компонентів до сховищ та інтерфейсів показують напрям взаємодії та передачі результатів аналізу. Взаємодія між компонентами формує цілісний процес аналізу вхідних медіа з подальшим доступом до збережених результатів. Більш детальна візуалізація описаних вище компонентів розроблюваної системи знаходиться у додатку Г.

2.5 Опис проєктованих станів системи

Стани системи обробки медіафайлів з обличчям людини за допомогою нейронних мереж будуть описані нижче в розділі. У процесі виконання заданих команд користувачем відбувається послідовна зміна станів відповідно до логіки обробки мультимедійних даних. Кожний стан виконує певну функцію та передбачає перехід до наступного стану після завершення відповідних операцій або за ініціативою користувача. Відповідно користувач має вплив на потік станів системи взаємодіючи з нею через інтерфейс безпосередньо задаючи команди для виконання. За логікою, діаграма станів системи, зображена на рисунку 2.3 описує типовий робочий потік проєктованої системи обробки мультимедійних даних, і є узгодженою та послідовною.

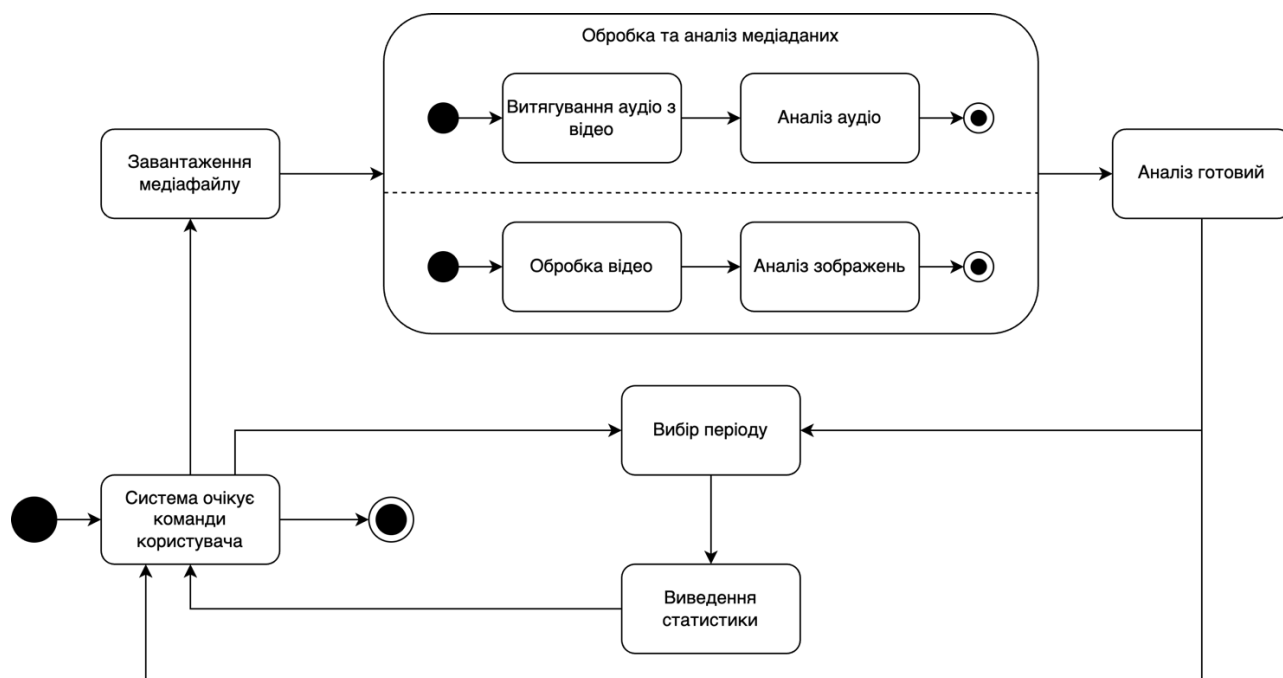


Рисунок 2.3 – Діаграма станів системи

Початковим етапом є завантаження медіафайлу, де система очікує надходження вхідного відео. Завершення цього завантаження ініціює перехід до обробки та аналізу медіаданих. На цьому етапі відбувається розгалуження на паралельні процеси: обробки кадр за кадром відео та витягування аудіо з відео. Паралельно після завершення цих операцій система переходить до наступних двох паралельних станів – аналіз аудіо та аналіз зображень, де виконується оцінка текстового та візуального компонентів. Поєднання результатів із цих гілок веде до стану системи, коли аналіз готовий, що свідчить про узагальнення всієї зібраної інформації. З цього стану, за ініціативою користувача, система переходить до вибору періоду, де користувач може налаштувати часовий діапазон для подальшого огляду. Завершення вибору періоду спричиняє перехід до виклику команди щодо виведення статистики, у якому система презентує агреговані результати за обраний інтервал.

Після демонстрації статистики процес замкнуто повертається до стану Система очікує команди користувача, де система знову готова перейти до інших доступних станів залежно від подальших дій користувача чи зовнішніх подій. Таким чином, кожний перехід зумовлений логічним завершенням операції або

взаємодією з користувачем, а кожен стан чітко відповідає конкретному кроку в процесі обробки та аналізу мультимедіа.

2.6 Опис сутностей системи

У даному підрозділі буде описано основні сутності, що використовуються в системі, що формують основу всієї системи обробки медіафайлів з використанням нейронних мереж. На рисунку 2.4 зображено абстрактну схему зв'язків між корінними сутностями в системі.

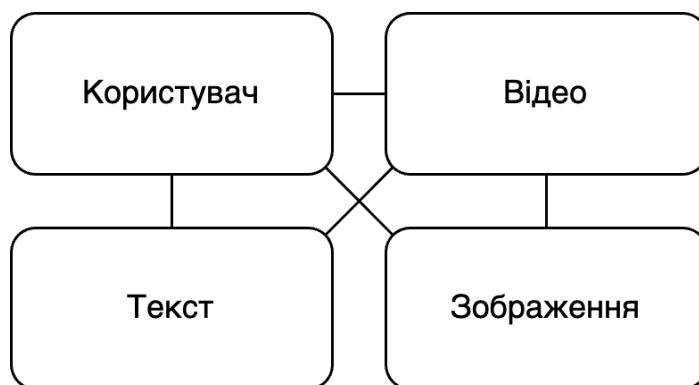


Рисунок 2.4 – Схема зв'язків між корінними сутностями в системі

Зв'язки між сервісами були встановлені таким чином, що Користувачі системи можуть мати кілька Зображень, Відео та Текстів – всі вони є вхідними даними в систему, або результат обробки відео. Зображення і Текст також асоціюються з Відео, тому що вони можуть бути витягнутими звідти в процесі обробки.

У додатку Д зображено детальний аналіз сутностей і їх атрибутів, використаних в роботі системи. Діаграма у додатку зображує сутності та їх атрибути: «Користувач» (UserId, Ім'я, Прізвище, Дата народження, Електронна пошта), «Зображення» (PhotoId, Дата та час завантаження, Джерело, Ключ у сховищі), «Відео» (VideoId, Дата та час завантаження, Ключ у сховищі, Дата та час коли зроблено), «Текст із мовлення» (SpeechTextID, Вміст тексту, Дата та час початку, Дата та час закінчення).

«Користувач» пов'язаний зі своїми медіаресурсами (зображеннями та відео) зв'язками «має» чи «містить» (ромбоподібні вузли). Аналогічно «Зображення» та «Відео» пов'язані з процесами «Аналіз зображення» та «Аналіз мовлення», які зберігають емоції та їх ймовірності. «Текст із мовлення» прив'язаний до користувача та аналізу мовлення. Всі зв'язки та атрибути чітко відображені стрілками та позначками (1, M) позначаючи чисельність зв'язків: зв'язок один-до-багатьох або багато-до-багатьох.

Висновки до розділу 2

У результаті проведеного дослідження було спроектовано систему комп'ютерного зору з обробки медіафайлів з розпізнаванням обличчя людини за допомогою нейронних мереж. Були сформульовані та визначені вимоги: і функціональні, так і нефункціональні. Було збережено баланс між точністю аналізу емоцій, продуктивністю системи та забезпеченням конфіденційності користувацьких даних.

Було описано сценарії використання даної системи для зареєстрованих і незареєстрованих користувачів з детальним описом алгоритму дій для досягнення мети конкретного сценарію. Це надасть універсальність та адаптивність розроблюваної системи в умовах мультимодального виявлення емоцій для покращення повсякденного моніторингу емоційного стану користувачів.

Було описано можливості користувача у системі та послідовність його дій взаємодіючи з системою за допомогою інтуїтивно зрозумілого інтерфейсу. Продемонстровано діаграми різних процесів та схем взаємодій. Всі матеріали надали краще розуміння архітектури системи та її компонентів. Спроектowana система відповідає вимогам описаним в даному розділі, зокрема високій точності, безпеці та можливості масштабування.

Це робить можливим подальший розвиток та інтеграцію у різні сфери: медицина, бізнес, освіта та сфера послуг, що буде сприяти покращенню якості життя та зміцнювати взаємодію між людьми та технологіями.

3 ВИБІР ТА ОБҐРУНТУВАННЯ ЗАСОБІВ І ТЕХНОЛОГІЙ

3.1 Вибір хмарного постачальника

Вибір хмарних технологій для розробки систем ІІІ стає дедалі популярнішим завдяки низці переваг, які забезпечують масштабованість, ефективність та доступність. Однією з головних причин є висока обчислювальна потужність, яку пропонують провайдери хмарних сервісів. Розробка ІІІ вимагає значних ресурсів для тренування моделей, обробки великих обсягів даних та оптимізації алгоритмів. Хмарні сервіси, такі як Amazon Web Services (AWS), Google Cloud Platform (GCP) або Microsoft Azure, надають розробникам доступ до GPU та TPU-інстансів, які можна масштабувати залежно від потреб проєкту. Це дозволяє економити кошти, оскільки ресурси оплачуються лише за фактичне використання, без необхідності інвестицій у дороге обладнання.

Ще одним вагомим аргументом є інтеграція хмарних технологій із сучасними інструментами для розробки та тестування ІІІ. Провайдери хмарних сервісів пропонують готові до використання бібліотеки, платформи для машинного навчання (наприклад, TensorFlow або PyTorch) та сервіси для обробки даних. Це значно скорочує час на розробку та впровадження рішень. Таким чином, використання хмарних платформ сприяє прискоренню інновацій, зменшенню витрат і створенню високопродуктивних систем штучного інтелекту.

Вибір платформи хмарних обчислень є ключовим етапом у розробці системи емоційного моніторингу на основі ІІІ. Платформа повинна забезпечувати виявлення емоцій у реальному часі, масштабовану обробку даних і інтеграцію з сучасними інструментами ІІІ. Після аналізу AWS, Microsoft Azure і Google Cloud Platform [15] було обрано AWS як оптимальне рішення. Ця платформа є лідером у галузі (рисунок 3.1), пропонує широкий спектр послуг, є економічно ефективною та відповідає всім вимогам проєкту. AWS також дозволяє тестувати прототип системи в хмарі з мінімальними початковими витратами, що додає їй додаткових переваг.

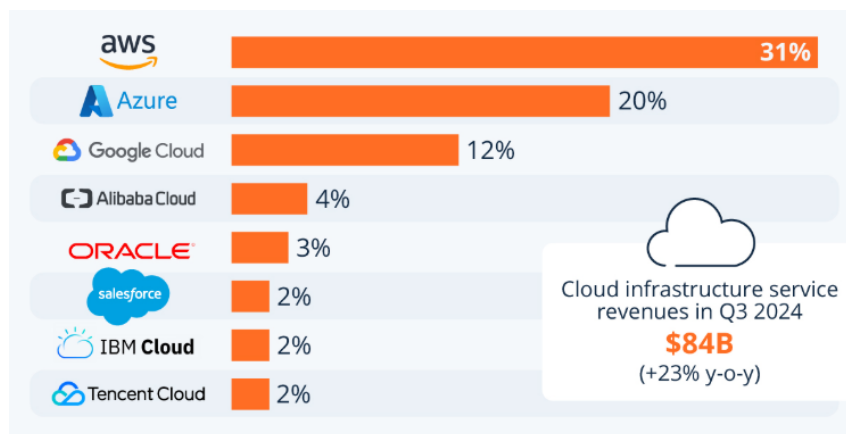


Рисунок 3.1 – Частка світового ринку провідних хмарних послуг [15]

AWS, Azure та GCP порівнювали на основі їхніх служб ІІІ та машинного навчання, можливостей зберігання, масштабованості, економічності та знайомства розробників. Оцінка також включала порівняння з підходом до місцевого середовища для оцінки здійсненності та витрат. Аналіз зображено на таблиці 3.1.

Таблиця 3.1 – Оцінка хмарних платформ: AWS, Azure та GCP

Особливість	AWS	Microsoft Azure	GCP
Сервіси AI та ML	Різнорамні (SageMaker, Transcribe)	Інтеграція з Microsoft AI	Орієнтація на TensorFlow, Vertex AI
Опції сховищ	S3, DynamoDB, RDS	Blob Storage, Cosmos DB, SQL Azure	Bigtable і Firestore
Масштабованість	Висока, автоматична	Помірна, потребує ручного налаштування	Висока, автоматична
Економічність	Низькі початкові витрати	Подібна модель, знижки на продукти Microsoft	Ефективний для ІІІ, високі витрати на мережу
Глобальне покриття	Найширша мережа	Широка мережа, але менша за AWS	Обмежене покриття
Простота розгортання	Просте, з хорошою документацією	Потребує більше зусиль в порівнянні з AWS	Зручний для ІІІ, обмежена підтримка

AWS пропонує найповніший і добре інтегрований набір служб штучного інтелекту та машинного навчання, включаючи Amazon SageMaker для створення та розгортання моделей машинного навчання. Ці послуги безпосередньо узгоджуються з вимогами системи комп'ютерного зору щодо мультимодального виявлення емоцій та аналізу даних. Azure та GCP надають конкурентоспроможні

інструменти штучного інтелекту, але потребують додаткового налаштування (Azure) або більш вузько орієнтовані на штучний інтелект на основі TensorFlow (GCP), що робить їх менш гнучкими для різноманітних потреб цього проєкту. Локальне середовище покладатиметься на фреймворки з відкритим вихідним кодом, такі як TensorFlow або PyTorch, які, хоча й потужні, але не матимуть повної інтеграції та масштабованості на відміну від AWS. Глобальна інфраструктура AWS забезпечує безпрецедентну масштабованість та низьку затримку доступу, що є важливим для виявлення та аналізу емоцій у реальному часі. Покриття Azure є потужним, але не таким широким, тоді як мережа GCP продовжує зростати.

Платна модель AWS дозволяє тестувати і розгортати прототипи системи безпосередньо в хмарі з мінімальними початковими витратами, що ідеально підходить для оцінки доцільності системи комп'ютерного зору. Можливість динамічного масштабування ресурсів також зменшує зайві витрати під час розробки.

Локальне середовище передбачає високі початкові витрати на придбання та обслуговування обладнання, а також поточні операційні витрати, що було б неефективно для проєкту підтвердження концепції.

Мій попередній досвід роботи з інструментами та сервісами AWS ще більше зміцнює мій вибір в якості хмарного провайдера. Ця обізнаність забезпечує швидке розгортання, оптимізовану розробку та скорочення часу навчання, що дозволяє проєкту дотримуватися графіку. Перехід на Azure або GCP потребував би додаткового навчання, що збільшило б час і складність розробки.

AWS був обраний в якості хмарного провайдера завдяки широкому спектру послуг, безшовній інтеграції зі штучним інтелектом та інструментами зберігання даних, а також неперевершеній глобальній інфраструктурі. Можливість розгортання та тестування прототипу безпосередньо в хмарі з мінімальними витратами спрощує оцінку доцільності та прискорює розробку. У порівнянні з локальними середовищами, AWS усуває необхідність попередніх інвестицій в обладнання та обслуговування, пропонуючи масштабоване і економічно ефективне рішення для системи комп'ютерного зору та аналізу емоцій.

3.2 Вибір нейромереж та бібліотек

Вибір моделей нейронних мереж і наборів даних є важливим кроком у розробці системи комп'ютерного зору на основі ШІ. Для задачі розпізнавання виразу обличчя (FER) в якості основних моделей були обрані ResNet і XLM-RoBERTa, а для навчання і оцінки використовувався набір даних FER+. Модель ResNet, будучи поміченою в статті найкращих архітектур CNN, вирішує проблему деградації в глибоких мережах, уможливаючи навчання набагато глибших архітектур шляхом сприяння кращому градієнтному потоку під час зворотного поширення [16]. Цей вибір мереж ґрунтувався на високій продуктивності, відповідності вимогам розпізнавання емоцій та придатності для цілей системи. Рішення було підкріплене обмеженнями, що спостерігалися в альтернативних моделях і наборах даних, які виявилися менш ефективними для досягнення бажаної точності та універсальності, зображення цього можна бачити на рисунку 3.2.

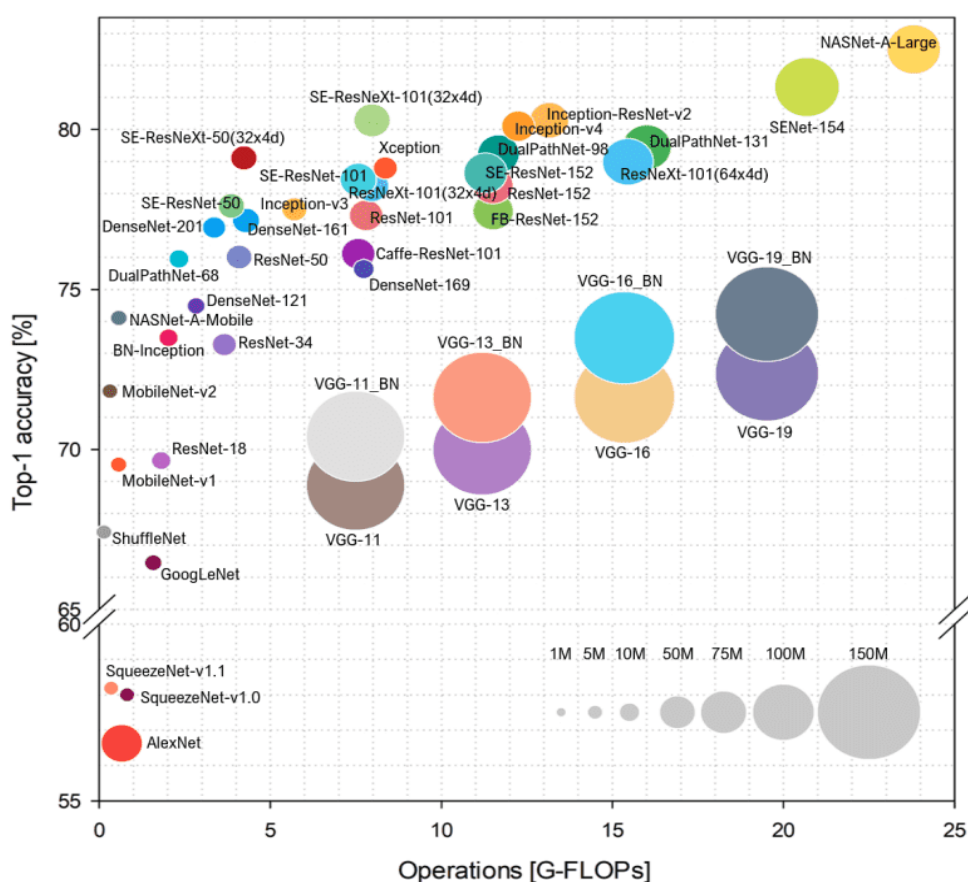


Рисунок 3.2 – Порівняння моделей за точністю [16]

Для компоненту аналізу емоцій на основі тексту було обрано XLM-RoBERTa завдяки її найсучаснішим характеристикам у завданнях розуміння природності на різних мовах. XLM-RoBERTa, багатомовний варіант RoBERTa (Robustly Optimized BERT), використовує архітектуру трансформатора з акцентом на ефективному попередньому навчанні та точному налаштуванні та має розмір словника на 250 тисяч унікальних токенів. Це робить його особливо ефективним для виявлення емоцій у тексті, включаючи нюансовані вирази, які часто присутні в багатомовних наборах даних. Також було розглянуто альтернативні моделі, такі як BERT або GPT-2, але вони виявилися менш придатними з наступних причин:

- BERT та GPT-2 добре працюють із англійськими наборами даних, однак їм бракує попереднього навчання для різних мов, яке забезпечує XLM-RoBERTa, що має вирішальне значення для додатків, які використовують неанглійські текстові вхідні дані;
- удосконалена методологія попереднього навчання XLM-RoBERTa забезпечує краще контекстне розуміння і класифікацію емоцій, особливо при аналізі тонких або складних текстових вхідних даних.

Для FER на основі зображень було обрано ResNet (Residual Network) завдяки його ефективності та надійності у візуальних завданнях. Інноваційні пропуски зв'язки ResNet пом'якшують проблему зникаючого градієнта, що дозволяє навчати більш глибокі мережі, які вловлюють складні риси виразу обличчя. Також були розглянуті альтернативні архітектури, такі як VGGNet або Inception, однак вони були визнані менш оптимальними з наступних причин:

- VGGNet, хоча і є ефективною, але вимагає значних обчислень і є менш ефективною порівняно з ResNet, адже початкові моделі, хоча і потужні, вносять додаткову складність, яка не є необхідною для відносно структурованої задачі FER;
- ResNet стабільно перевершує старіші архітектури в бенчмарках FER, в тому числі на наборі даних FER+ (рисунок 3.3), що робить її надійним вибором для системи.

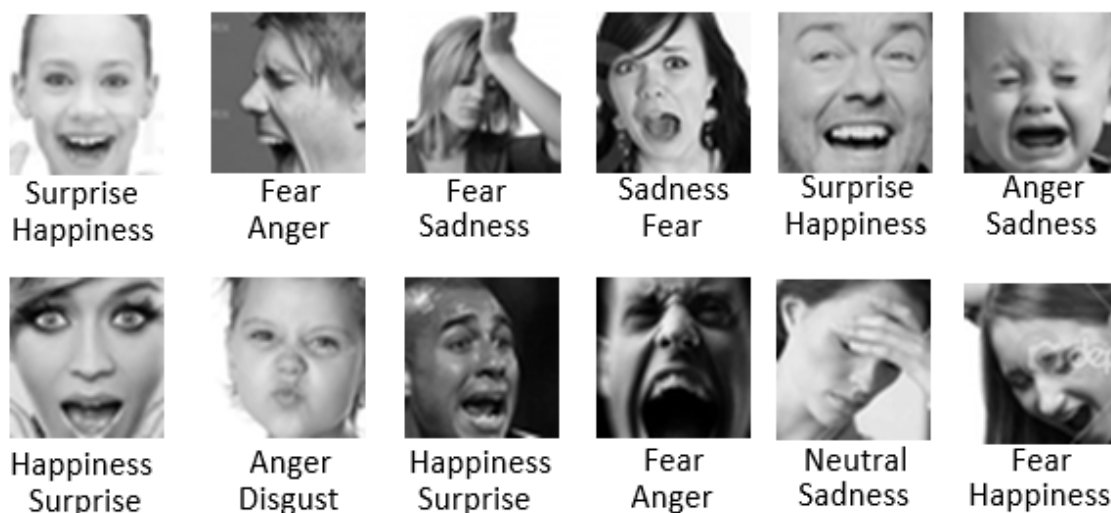


Рисунок 3.3 – Зображення порівняння міток FER2013 та FER+ [17]

Набір даних FER+ обрано як основний набір даних для навчання та оцінки по задачі виявлення виразів на обличчі людини. Цей набір широко використовується та відповідає вимогам проєкту. Набір даних містить майже 36 тисяч зображень виразів обличчя у відтінках сірого, розподілених на сім емоційних станів: гнів, відраза, страх, щастя, сум, здивування та нейтральний стан.

Були розглянуті альтернативні набори даних, такі як CK+, JAFFE або FER2013, але вони виявилися менш придатними:

- FER+ пропонує значно більший набір даних з різноманітними виразами, що покращує здатність моделі до узагальнення, CK+ і JAFFE менші за розміром і не мають такої ж варіативності виразів обличчя, що потенційно обмежує продуктивність у реальних застосуваннях;

- у наборі даних FER+ кожне зображення анотовано 10 різними фахівцями, що покращує якість базових міток істинності для розпізнавання емоцій обличчя порівняно з оригінальним набором даних FER2013 [17], агрегуючи кілька анотацій до зображення, оцінка розподілу ймовірностей емоцій для кожного обличчя мала меншу похибку – це полегшує подальшу розробку алгоритмів, які можуть робити висновки по кількох мітках, а не звичайні з однією;

- FER+ є широко визнаним еталонним набором даних у дослідженнях FER, а тому продуктивність системи можна порівняти з існуючими рішеннями.

GoEmotions датасет було обрано для навчання XLM-RoBERTa, оскільки він розроблений із повним набором із 27 категорій емоцій. Рядки в цьому наборі даних мають по кілька лейблів, що дає детальніше зрозуміти емоції. Набір даних підходить для класифікації за кількома мітками та охоплює широкий спектр повсякденної розмовної мови, а тому його можна застосувати для різноманітних завдань аналізу емоцій, а також для експериментування з співставленням емоцій та навчанням мережі.

Комбіноване використання XLM-RoBERTa для виявлення емоцій на основі тексту із GoEmotions і ResNet із FER гарантує, що система досягає високої точності в різних модальностях. Можливість XLM-RoBERTa розуміти українську мову та надійно вилучати ознаки емоцій у зображення за допомогою ResNet відповідають вимогам системи щодо ефективної обробки різноманітних джерел даних. Набір даних FER+ ще більше підвищує точність та надає стійку основу для різноманітного навчання і підлаштування нейронної мережі. Разом ці вибірки дозволяють системі надавати точні та послідовні результати розпізнавання емоцій, підтримуючи її основну функцію збору та аналізу емоційної статистики.

3.3 Вибір сховища даних в хмарному середовищі

Вибір сховища даних має вирішальне значення для забезпечення ефективної та масштабованої обробки різних типів даних у системі комп'ютерного зору на основі ШІ. Система потребує надійних рішень для керування структурованими даними, такими як профілі користувачів і результати емоційного аналізу, а також неструктурованими даними, такими як відео, зображення та моделі нейронних мереж. Після ретельної оцінки сервісів для зберігання даних які пропонує AWS (рисунк 3.4), DynamoDB, Simple Storage Service (S3) та Elastic File System (EFS) були обрані як найбільш прийнятні рішення. У цьому розділі викладено обґрунтування цих виборів, порівняно їх з альтернативами та підкреслено їхні переваги для цієї системи. Якщо коротко, Amazon EFS пропонує масштабоване спільне сховище файлів, доступне для кількох екземплярів EC2 із автоматичним

масштабуванням, Amazon S3 забезпечує масштабоване сховище об'єктів, доступне через Інтернет із різними класами сховища для різних потреб доступу, а Amazon EBS забезпечує блочне сховище з низькою затримкою для окремих екземплярів EC2 з стабільною продуктивністю [18].



Рисунок 3.4 – AWS сервіси для зберігання даних [18]

Сховище даних для цієї системи має відповідати декільком ключовим вимогам:

- система має працювати зі схемами даних, що змінюються, наприклад, динамічними емоційними метриками або розширенням записів користувачів;
- у міру зростання системи сховище даних повинно обробляти зростаючі обсяги даних без погіршення продуктивності;
- необхідно досягти балансу між операційними витратами і функціональністю, особливо для зберігання мультимедійних даних.

Amazon DynamoDB було обрано для зберігання даних пов'язаних з користувачем, результатів аналізу та метаданих завдяки її гнучкості та продуктивності. Безсхемний дизайн DynamoDB легко адаптується до змінної структури даних користувача та емоційних показників, уникаючи жорсткості традиційних реляційних баз даних. Ця гнучкість особливо цінна для обробки динамічних наборів даних. Також завдяки операціям читання та запису з низькою затримкою DynamoDB забезпечує майже миттєвий доступ до даних і, на відміну

від реляційних баз даних, які вимагають масштабування вручну, DynamoDB автоматично масштабується для обробки різних робочих навантажень, забезпечуючи стабільну продуктивність у міру зростання системи. DynamoDB працює за моделлю pay-as-you-go, що робить його економічно ефективним варіантом, особливо на початкових етапах розробки та розгортання системи.

Amazon S3 було вибрано для зберігання великих неструктурованих даних, таких як завантажені користувачами відео та зображення. S3 спеціально розроблено для зберігання та отримання великих об'єктів, що робить його ідеальним вибором для керування мультимедійними файлами. S3 автоматично масштабує сховище відповідно до зростаючих обсягів даних, гарантуючи, що система може обробляти зростаючі завантаження користувачів без додаткового налаштування. Більше того, завдяки надійності і високій доступності S3 гарантує користувачеві безпечне завантаження файлів і зберігання та легку доступність. Також є параметри багаторівневого зберігання, які дозволяють файлам, до яких часто звертаються, залишатися на стандартному рівні, тоді як старіші або менш важливі файли можна перемістити на дешевші рівні, такі як S3 Glacier, без втрати продуктивності.

Amazon Elastic File System (EFS) було обрано для зберігання великих попередньо навчених моделей нейронних мереж, таких як ResNet50 і XLM-RoBERTa, а також бібліотек Python. AWS Lambda має ліміт пам'яті 512 МБ, чого недостатньо для великих моделей нейронних мереж, але EFS долає це обмеження, надаючи практично необмежену пам'ять. Кілька функцій Lambda можуть отримати доступ до однієї файлової системи EFS, усуваючи дублювання та спрощуючи керування ресурсами. Разом із тим, EFS дозволяє оновлювати моделі та бібліотеки NN без повторного розгортання функцій Lambda, гарантуючи, що останні версії завжди доступні. EFS має високу пропускну здатність та доступ із низькою затримкою, це підходить для роботи з нейронними мережами та великими бібліотеками Python. У таблиці 3.2 представлено порівняння основних служб зберігання даних AWS, яке охоплює їхні ключові функції.

Таблиця 3.2 – Порівняння AWS сховищ даних за характеристиками

Функція	DynamoDB	S3	EFS	RDS
Тип даних	Структуровані	Напівструктуровані	Неструктуровані	На основі файлів
Масштабованість	Автоматична	Висока	Автоматична	Висока
Продуктивність	Читання/запис із низькою затримкою	Оптимізовано для великих об'єктів	Низька затримка для частого доступу	Висока для реляційних запитів
Вартість	Оплата за використання	Рівневий підхід до оптимізації вартості	Висока	Вищі операційні витрати
Інтеграція з AWS	Безшовна	Безшовна	Безшовна	Обмежена
Гнучкість	Без схеми	Об'єктно-орієнтована	Файлова основа	Фіксовані схеми
Випадок використання в системі	Метадані та емоційний аналіз	Відео та зображення	Моделі та бібліотеки машинного навчання	Реляційні набори даних

Хоча RDS підходить для реляційних даних, він вимагає попередньо визначених схем, ручного масштабування та додаткових витрат для великомасштабних операцій. Його жорсткість і складність роблять його менш придатним для динамічних наборів даних, таких як результати емоційного аналізу. Локальні рішення пропонують повний контроль, але потребують великих початкових витрат і поточного обслуговування. Їм не вистачає масштабованості та інтеграції, які надають сервіси AWS, що робить їх непрактичними для цієї хмарної системи.

Вибрані компоненти сховища даних – DynamoDB, S3 і EFS – працюють разом, щоб ефективно справлятися з різноманітними потребами системи в даних:

- DynamoDB зберігає метадані, результати емоційного аналізу та профілі користувачів для швидкого надсилання запитів і візуалізації;
- S3 обробляє завантаження відео та зображень, пропонуючи довговічність і масштабованість для неструктурованих даних;

– EFS забезпечує централізоване зберігання моделей і бібліотек нейронних мереж, забезпечуючи ефективно та масштабоване виконання робочих процесів машинного навчання в Lambda.

Комбіноване використання Amazon DynamoDB, Amazon S3 і Amazon EFS надає системі ефективність обробки різних типів даних (рисунок 3.6), залишаючись при цьому масштабованою та економічно ефективною. DynamoDB пропонує доступ до структурованих даних із низькою затримкою, S3 оптимізує зберігання для великих мультимедійних файлів, а EFS усуває обмеження Lambda щодо зберігання, надаючи спільний доступ до нейронних мереж. Цю архітектуру було вибрано, щоб відповідати цілям системи, забезпечуючи надійну продуктивність і повну інтеграцію всіх компонентів.

3.4 Вибір сервісів AWS для забезпечення аналізу медіафайлів

Розробка компонентів аналізу медіа для системи комп'ютерного зору на основі ШІ вимагає ретельного вибору сервісів AWS, щоб забезпечити масштабованість, економічну ефективність та простоту створення прототипів. Обрана архітектура використовує безсерверні можливості AWS та інтегрує кілька сервісів для ефективної обробки мультимедійних файлів, підтримуючи при цьому функціональні вимоги системи. У цьому розділі наведено детальне обґрунтування обраних сервісів та їх ролі у запропонованому робочому процесі.

Для реалізації завдання FER було обрано AWS Lambda у поєднанні з Elastic File System (EFS) для прототипування. Основними перевагами Lambda функцій є:

- функції автоматично масштабуються відповідно до різних робочих навантажень, що є критично важливим для динамічної обробки медіа;
- плата стягується лише за використаний обчислювальний час, що ідеально підходить для обробки спорадичних медіа-подій;
- Lambda легко інтегрується з іншими сервісами AWS, що дозволяє робочому процесу залишатися модульним і гнучким;

– використання AWS Lambda з Amazon Elastic File System (EFS), для роботи з великими залежностями та моделями машинного навчання, забезпечує ефективне розгортання значних пакетів, адаптуючи для завдань у реальному часі без обмежень, пов'язаних із обмеженнями зберігання Lambda [20][21].

В даній системі вона забезпечує безсерверне середовище для виконання коду у відповідь на певні тригери, усуваючи потребу в управлінні базовою інфраструктурою (рисунок 3.5). Завдяки інтеграції Lambda з EFS файли моделі нейронної мережі ResNet завантажуються безпосередньо у функцію, що забезпечує ефективне виконання моделі FER. Такий підхід має кілька переваг:

- Lambda дозволяє виставляти рахунки за використання, що робить його економічно вигідним на етапі створення прототипу, коли робочі навантаження є непостійними;
- код і моделі, розроблені для Lambda, можуть бути легко перенесені на екземпляри EC2 в майбутньому, у випадку використання більших наборів даних;
- EFS надає рішення для спільного зберігання даних, яке усуває необхідність повторного завантаження файлів моделей, забезпечуючи ефективний доступ до даних під час виконання Lambda.

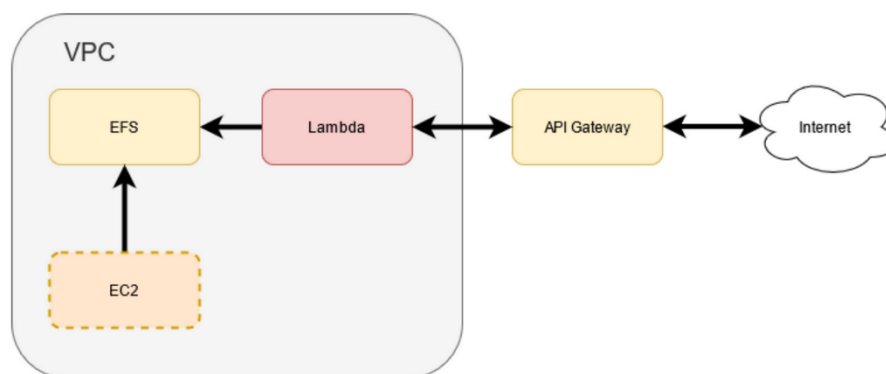


Рисунок 3.5 – Високорівнева схема архітектури для FER

Для розпізнавання емоцій на основі тексту за допомогою моделі XLM-RoBERTa було обрано аналогічну конфігурацію Lambda + EFS. Ця конфігурація особливо добре підходить для прототипу з використанням невеликих файлів mp4, де аудіо може бути витягнуто і транскрибовано паралельно з обробкою перед

аналізом емоцій. Здатність Lambda обробляти робочі навантаження, керовані подіями, EFS забезпечує швидкий доступ до попередньо навчених файлів моделей XLM-RoBERTa та ResNet. Тому виявлення емоцій на основі тексту виконується ефективно і не вимагає спеціальної інфраструктури.

Також під час розробки даної системи були створені додаткові лямбда-функції AWS, що використовуються для запуску моделей, обробки результатів і керування метаданими. Процес керування обробки медіа виконується за допомогою AWS Step Functions. Цей сервіс дозволяє створити машину станів, яка автоматизує виконання багатьох завдань, таких як попередня обробка медіа, виконання моделі та після-обробка. Покрокові функції мають такі переваги:

- спрощене управління робочими процесами: складні робочі процеси розбиваються на дискретні стани, що дозволяє чітко візуалізувати і контролювати шляхи виконання;
- обробка помилок: вбудовані механізми повторних спроб гарантують, що завдання плавно відновлюються після збоїв, підвищуючи надійність системи;
- інтеграція на основі подій: покрокові функції безперешкодно працюють з Lambda, MediaConvert та іншими сервісами AWS, забезпечуючи плавні переходи між етапами обробки.

Amazon EventBridge був обраний для полегшення комунікації між сервісами AWS та автоматизації робочих процесів. Цей сервіс прослуховує події, що генеруються різними компонентами робочого процесу, і запускає відповідні дії. Наприклад, коли відео завантажується в S3-бакет, EventBridge гарантує, що відповідні лямбда-функції будуть викликані для початку обробки. Даний сервіс забезпечує відокремлену комунікацію сервісів, зменшуючи складність системи. Налаштувавши правила подій можна гарантувати, що робочі процеси адаптуються до конкретних тригерів та умов.

Сповіщення Amazon S3 відіграють вирішальну роль в ініціюванні процесу аналізу медіаматеріалів. Коли нове відео або зображення завантажується в S3, механізм сповіщень запускає відповідні лямбда-функції. Аналіз медіа починається автоматично після завантаження файлу, що забезпечує мінімальну затримку. А

також, для майбутніх розширень системи без порушення існуючого робочого процесу, даний сервіс дозволяє налаштувати додаткові тригери.

Для попередньої обробки мультимедійних файлів було обрано AWS Elemental MediaConvert завдяки його здатності працювати з різними форматами та роздільною здатністю, а також легкою інтеграцією із іншими AWS сервісами. MediaConvert спрощує підготовку відео для аналізу, витягуючи аудіопотоки, змінюючи розмір зображень і конвертуючи файли в стандартні формати. Сервіс підтримує розширені функції обробки мультимедійних даних, гарантуючи, що майбутні вимоги, наприклад, більш складні типи мультимедійних даних, можуть бути реалізовані. Також MediaConvert, як керований сервіс, автоматично масштабується для обробки численних запитів на обробку.

Поєднання Lambda, EFS, сповіщень S3, EventBridge, Step Functions та MediaConvert гарантує, що робочий процес аналізу медіа залишається економічно ефективним, масштабованим і простим в управлінні. Безсерверна модель Lambda знижує витрати на інфраструктуру під час створення прототипів, зберігаючи при цьому гнучкість для майбутнього масштабування. Step Functions та EventBridge забезпечують автоматизоване управління робочим процесом, зменшуючи ручне втручання та оптимізуючи конвеєр обробки. Сповіщення S3 та MediaConvert ще більше покращують робочий процес, дозволяючи запускати тригери в реальному часі та спрощуючи попередню обробку медіа.

Ця архітектура відповідає вимогам системи до виявлення та аналізу емоційних станів на основі мультимедійних даних, одночасно підтримуючи ефективні процеси розробки та тестування. Завдяки використанню інтегрованих сервісів AWS, технології аналізу медіа для поточних потреб і майбутніх розширень проєкт є хорошим рішенням для системи комп'ютерного зору.

3.5 Вибір сервісів хмарного середовища для безпеки та доступу з інтернету

Безпечний і масштабований веб-доступ має важливе значення для комп'ютерного зору на основі ІІІ, щоб користувачі могли надійно взаємодіяти з її

функціями, зберігаючи при цьому цілісність конфіденційних емоційних даних. Для цього було обрано Amazon Cognito та Amazon API Gateway. Ці сервіси забезпечують надійну автентифікацію користувачів і безпечний зв'язок між фронтендом і бекендом системи, що відповідає вимогам системи щодо масштабованості та безпеки.

Amazon Cognito забезпечує автентифікацію користувачів та управління обліковими записами. Він пропонує бездоганну інтеграцію зі службами AWS, такими як Lambda (рисунок 3.6), і підтримує широкий спектр методів автентифікації, включаючи федеративні постачальники ідентифікаційних даних (наприклад, Google, Facebook) і багатофакторну автентифікацію (MFA). Такий вибір гарантує надійне управління даними користувачів, що відповідає найважливішим вимогам безпеки та відповідності.

AWS API Gateway можна інтегрувати з Amazon Cognito для автентифікації користувачів, налаштовуючи авторизацію в API Gateway і захист кінцевих точок API, так, щоб лише авторизовані користувачі отримали доступ до послуги [22].

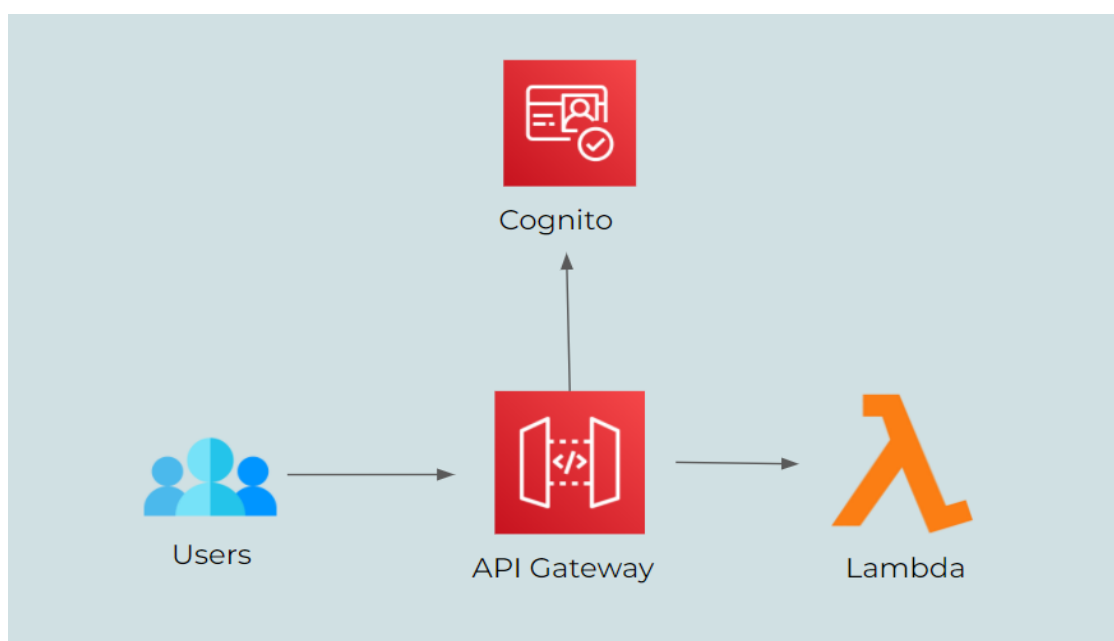


Рисунок 3.6 – Автентифікація користувачів за допомогою Amazon Cognito [22]

Масштабованість Cognito є ще одним ключовим фактором, оскільки він може працювати з великими базами користувачів без ручного втручання. Його здатність

інтегруватися з внутрішніми сервісами, такими як AWS Lambda і DynamoDB, забезпечує безперебійний потік даних між процесами автентифікації та основними функціями системи. Хоча у дисертації були розглянуті сторонні постачальники ідентифікаційних даних та користувацькі системи автентифікації, їм не вистачає економічної ефективності та простоти інтеграції, яку забезпечує Cognito.

Amazon API Gateway слугує контрольованою точкою входу для взаємодії користувачів з системою. Керуючи зв'язком між веб-інтерфейсом і внутрішніми сервісами за допомогою REST правил API, він забезпечує безпеку даних і масштабованість.

Вбудовані функції безпеки API Gateway, такі як шифрування, перевірка запитів та обмеження швидкості, захищають систему від несанкціонованого доступу та потенційного перевантаження. Крім того, API Gateway інтегрується безпосередньо з AWS Lambda для створення безсерверної архітектури, яка знижує вимоги до управління інфраструктурою, зберігаючи при цьому високу продуктивність. Цінова модель з оплатою за фактом використання також робить його економічним вибором як на етапі створення прототипу, так і на етапі масштабування.

Поєднання Amazon Cognito та API Gateway було обрано для забезпечення безпечної, масштабованої та економічно ефективної архітектури веб-доступу. Cognito спрощує автентифікацію користувачів, забезпечуючи безпечний вхід та управління обліковими записами, в той час як API Gateway забезпечує безперебійний зв'язок між користувачами та бекендом. Разом ці сервіси відповідають цілям системи – забезпечити надійний аналіз медіафайлів, зберігаючи при цьому конфіденційність даних та операційну ефективність.

3.6 Вибір мов програмування та фреймворків для поставленої задачі

Відповідно до звіту Datadog «The State of Serverless», Python і Node.js є найбільш використовуваними середовищами виконання програмного коду серед користувачів AWS Lambda, на них припадає більше половини всіх викликів

функцій завдяки їх підтримці та широким спільнотам розробників [23]. На рисунку 3.7 продемонстрована статистика по кількості викликів проти типів середовищ, що використовуються в AWS Lambda-функціях.

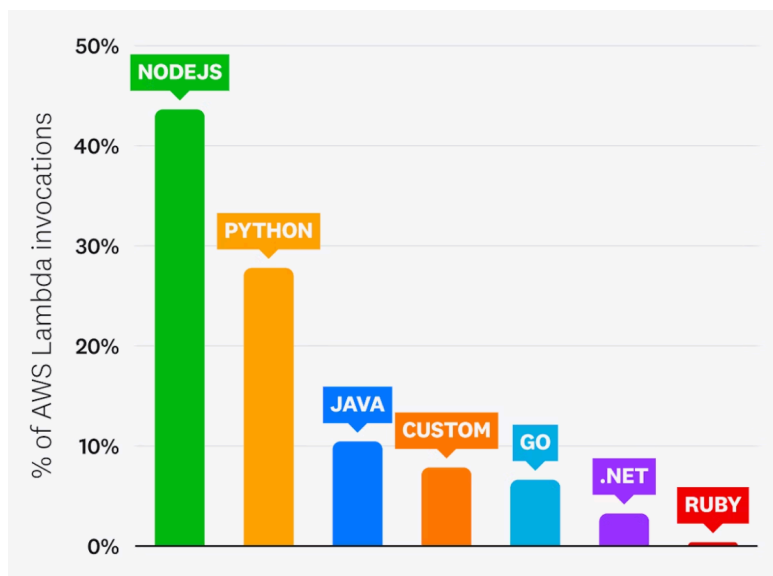


Рисунок 3.7 – Основні середовища виконання програмного коду, що використовуються в Lambda-функціях [23]

Реалізація спирається на відомі бібліотеки глибокого навчання, такі як PyTorch та Hugging Face Transformers. PyTorch було обрано за його гнучкість та ефективне використання графічного процесора, тоді як Hugging Face Transformers забезпечує безперешкодний доступ до попередньо навчених моделей, таких як XLM-RoBERTa [23].

Для ResNet попередньо навчені ваги, доступні в PyTorch, спрощують точне налаштування і скорочують час навчання. Ці бібліотеки було обрано замість таких альтернатив, як TensorFlow або Keras, через їхні кращі можливості налагодження та легкість інтеграції з робочими процесами проєкту.

Вибір відповідних мов програмування та фреймворків є ключовим аспектом розробки системи комп'ютерного зору на основі ІІ. Вибрані технології мають відповідати різноманітним вимогам системи, включаючи ефективну інтеграцію ІІ, маніпулювання даними та безперебійну взаємодію з сервісами AWS.

Java, Python і Node.js були обрані на основі їх придатності для конкретних завдань, масштабованості в безсерверних середовищах і надійної підтримки бібліотек.

Java, будучи мовою, якою я володію найкраще, була обрана для завдань, які вимагають безпечної та надійної взаємодії зі службами AWS, наприклад генерування посилань для завантаження медіаданих і налаштування MediaConvert. Інтеграція Java з AWS SDK дозволяє ефективно генерувати безпечний та обмежений за часом доступ до об'єктів, що зберігаються в S3. Його надійна обробка помилок і багатопотокова архітектура забезпечують надійну роботу навіть у великих масштабах.

Python був обраний для завдань, пов'язаних із розпізнаванням виразу обличчя, інтеграцією нейронної мережі та попередньою обробкою зображень. Його домінування в робочих процесах ШІ та машинного навчання забезпечує сумісність із такими фреймворками, як TensorFlow і PyTorch. Такі бібліотеки, як OpenCV і Pillow, забезпечують ефективне розпізнавання облич і обрізання зображень, що є критично важливим для підготовки даних для розпізнавання емоцій. Python спрощує інтеграцію таких моделей, як ResNet і XLM-RoBERTa, використовуючи попередньо підготовлені ваги та інструменти для тонкого налаштування. Велика екосистема Python і бібліотеки з багатьма засобами роботи з нейронними мережами є ідеальними для реалізації розпізнання емоцій.

Node.js було обрано для завдань, які вимагають керування даними та статистичних обчислень через його асинхронну архітектуру, керовану подіями. Node.js бездоганно інтегрується з AWS SDK для обробки даних, оновлень і маніпуляцій, забезпечуючи ефективну взаємодію зі сховищем і базами даних. Node.js є найбільш популярним, тому буде можливість знайти підтримку зі складними ситуаціями. Ефективність Node.js у обробці одночасних завдань забезпечує економію коштів і масштабованість безсерверних робочих процесів.

Поєднання Java, Python і Node.js дозволяє кожному компоненту системи використовувати мову, яка найкраще підходить для його завдань, тоді як AWS Lambda полегшує їх інтеграцію в безсерверну архітектуру. Наприклад, попередньо

підписані URL-адреси, згенеровані Java, запускають завантаження медіафайлів, Python обробляє ці файли за допомогою моделей AI, а Node.js керує оновленнями даних і статистикою.

Висновок до розділу 3

У результаті проведеного аналізу було здійснено обґрунтований вибір засобів і технологій, необхідних для розробки системи комп'ютерного зору на основі штучного інтелекту. Було обрано хмарний постачальник Amazon Web Services (AWS) завдяки його високій масштабованості, економічній ефективності та широкому спектру сервісів, що відповідають вимогам проєкту. Порівняльний аналіз між AWS, Microsoft Azure та Google Cloud Platform підтвердив переваги AWS щодо інтеграції з інструментами штучного інтелекту, глобального покриття та простоти розгортання.

Для реалізації задач розпізнавання емоцій було обрано моделі ResNet та XLM-RoBERTa, а також набір даних FER+, що забезпечують високу точність та універсальність аналізу як візуальних, так і текстових даних. Ці вибори були підкріплені оцінкою альтернативних моделей та наборів даних, які не відповідали вимогам щодо продуктивності та масштабованості.

У виборі сховищ даних було надано перевагу Amazon DynamoDB, S3 та Elastic File System (EFS) завдяки їхній гнучкості, автоматичній масштабованості та економічній ефективності. Ці сервіси забезпечують ефективне управління як структурованими, так і неструктурованими даними, що є критично важливим для системи, що обробляє великі обсяги мультимедійної інформації. Для забезпечення аналізу медіафайлів було інтегровано сервіси AWS Lambda, Step Functions, EventBridge та MediaConvert. Ця комбінація сервісів дозволяє автоматизувати робочі процеси, забезпечуючи масштабованість, економічну ефективність та простоту управління системою. Використання безсерверних технологій Lambda у поєднанні з EFS дозволяє ефективно обробляти великі моделі нейронних мереж та забезпечувати швидкий доступ до даних.

Безпека та доступність системи були забезпечені за допомогою сервісів Amazon Cognito та API Gateway. Amazon Cognito надає надійну автентифікацію користувачів, управління обліковими записами та зручну інтеграцію з іншими сервісами, тоді як API Gateway забезпечує безпечний зв'язок між серверною і клієнтською частинами, захищаючи дані від несанкціонованого доступу та забезпечуючи високий рівень захисту інформації.

Для реалізації функціональності системи було обрано мови програмування Java, Python та Node.js, що дозволяє ефективно інтегрувати різні компоненти системи та забезпечує високу продуктивність і гнучкість розробки. Використання відповідних фреймворків та бібліотек, таких як PyTorch та Hugging Face Transformers, сприяло ефективній інтеграції моделей штучного інтелекту та забезпеченню точності аналізу емоцій.

Таким чином, вибір засобів і технологій було здійснено з урахуванням вимог до масштабованості, ефективності, безпеки та інтеграції. Система комп'ютерного зору обробки медіафайлів з зображенням обличчя має надійну основу. Обрана архітектура та використані технології відповідають сучасним стандартам і забезпечують можливість адаптації системи до змінних потреб ринку.

4 РЕАЛІЗАЦІЯ СИСТЕМИ ОБРОБКИ МЕДІАФАЙЛІВ

4.1 Реєстрація та авторизація користувача

Тепер опишемо автентифікацію та авторизацію як один із важливих аспектів системи комп'ютерного зору для обробки медіафайлів із обличчям людини за допомогою нейронних мереж в розробленому проєкті. Amazon Cognito, керований сервіс у хмарі, підтримує функції входу та реєстрації користувачів. Автентифікація користувачів з Cognito обрана за надійність набору функцій для автентифікації з токенами, багатофакторною автентифікацією, а також безшовною інтеграцію з іншими сервісами AWS (наприклад, сервіси API Gateway та Lambda). У цьому розділі детально описано реалізацію процесів входу та реєстрації, зокрема розроблений програмний код для створення HTTP-запитів та застосування заходів безпеки щодо доступу до прикладних програмних інтерфейсів.

AWS Cognito виступаючи в ролі менеджера користувацьких пулів полегшує розробку автентифікації для даної роботи та слугує каталогом для зберігання облікових даних і метаданих користувачів. Після проходження процесу реєстрації надана користувачами інформація починає надійно зберігатися в корзині користувачів AWS Cognito. Момент входу в систему, як зазвичай, має перевірку введених облікових даних користувачем та поверненням сервером згенерованих в Cognito токенів для авторизації в систему для доступу до ресурсів, доступних для користувача.

Здатність Cognito виконувати завдання автентифікації та авторизації робить процес розробки простішим, дозволяючи зосередитися на головній функціональності системи – виявленні та аналізі емоцій. Крім того, вбудовані функції безпеки роблять так, що система відповідає сучасним стандартам і захищає конфіденційні користувацькі дані.

Процес реєстрації робить так, що нові користувачі можуть безпечно створювати облікові записи. AWS Cognito перевіряє введені користувачами дані (адреси електронної пошти та паролі), щоб вони відповідали попередньо визначеним політикам безпеки. Після успішної реєстрації ми дізнаємось, що дані

користувача зберігаються в пулі користувачів Cognito. Нижче наведено приклад коду (рисунок 4.1), що генерує HTTP POST запит, який використовується для процесу реєстрації.

```

88      /**
89       * Registers a new user in the Cognito User Pool.
90       * @param username - The username (email or custom username) of the new user.
91       * @param password - The password for the new user.
92       * @param userAttributes - Additional attributes (e.g., email, phone number) for the user.
93       * @returns A promise resolving to the registration response.
94       */
95      1 usage
96      async registerUser(
97          username: string,
98          password: string,
99          userAttributes: { Name: string; Value: string }[]
100      ): Promise<any> {
101          const secretHash : string = await this.getSecretHash(username);
102
103          const commandInput: SignUpCommandInput = {
104              ClientId: this.clientId,
105              SecretHash: secretHash,
106              Username: username,
107              Password: password,
108              UserAttributes: userAttributes,
109          };
110
111          const command : SignUpCommand = new SignUpCommand(commandInput);
112
113          try {
114              return await this.client.send(command);
115          } catch (error) {
116              console.error("Registration failed", error);
117              throw error;
118          }
119      }

```

Рисунок 4.1 – Код для створення HTTP POST запиту під час реєстрації

Опис основних параметрів у вищезгаданому запиті:

- ClientId, обов'язковий, ідентифікує програму, що взаємодіє з групою користувачів Cognito;
- ім'я користувача (username), обов'язкове, надається користувачем і в даному застосунку є його електронною адресою;
- надійний пароль, обов'язковий, надається користувачем слідуючи правилам заданим у AWS Cognito Application, використовується політика паролів, наприклад мінімальна довжина та вимоги до символів для підвищення безпеки;

– атрибути (UserAttributes), за замовчуванням додаткові метадані, які можна налаштувати для конкретного застосунку і навіть зробити обов’язковими, приклад атрибутів: адреса електронної пошти чи номер телефону користувача – тобто маємо місце для персоналізації.

Після успішної реєстрації Cognito надішле повідомлення з підтвердженням на вказану електронну адресу, залежно від налаштувань пулу користувачів. Завжди можна налаштувати і SMS підтвердження, що робить це зручнішим для різних користувачів.

Було проаналізовано процес входу, який дозволяє зареєстрованим користувачам пройти аутентифікацію та отримати токени доступу для безперервної взаємодії з системою. З цього випливає, що процес входу забезпечує високий рівень безпеки та зручності. Це сприяє захисту даних та покращує користувацький досвід.

Нижче наведено приклад відповіді з токенами від AWS Cognito (рисунок 4.2), який використовується для доступу до ресурсів.

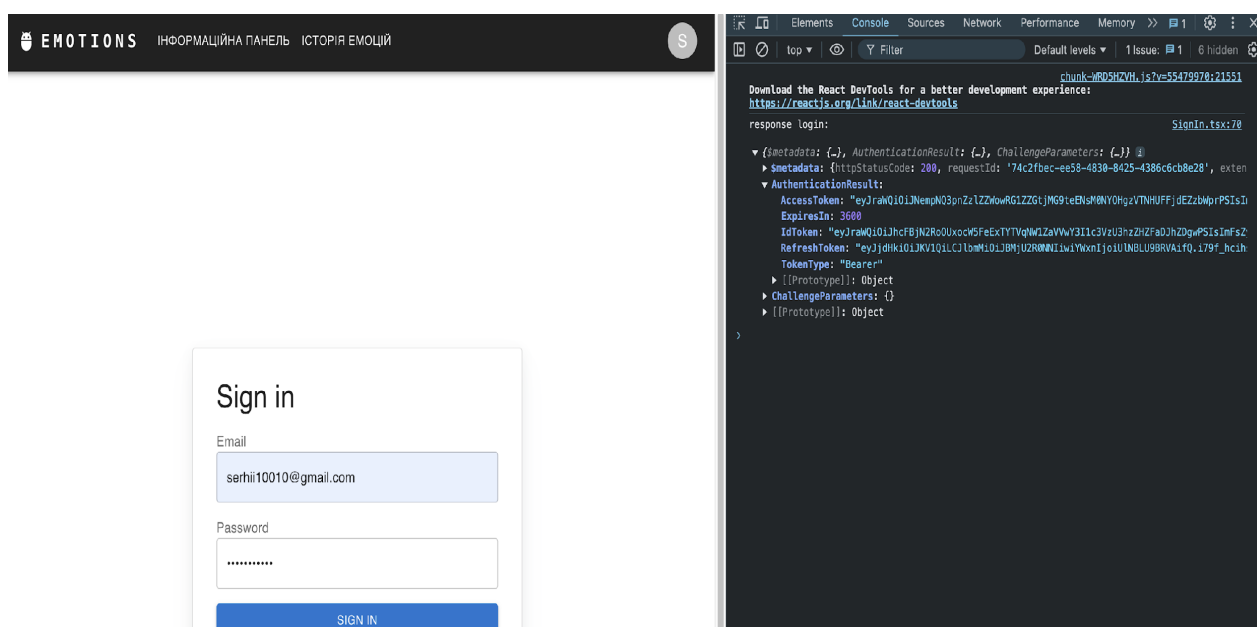


Рисунок 4.2 – Відповідь з токенами AWS Cognito

Захищаючи конфіденційність даних користувача, зв'язок між клієнтами та бекендом шифрується за допомогою HTTPS, запобігаючи перехопленню даних.

Токени, випущені Cognito, перевіряються під час кожного запиту для забезпечення їх автентичності та дійсності. Багатофакторну автентифікацію можна ввімкнути для додаткового рівня безпеки, вимагаючи від користувачів підтвердити свою особу за допомогою кодів електронної пошти.

Використання AWS Cognito для автентифікації користувачів забезпечує безпечний і ефективний процес входу і реєстрації. Cognito надає групу функцій, за допомогою якої система може безпечно керувати обліковими записами користувачів, зберігаючи відповідність сучасним стандартам безпеки.

4.2 Завантаження та обробка медіафайлів в системі

Ефективна та безпечна обробка медіаданих є важливим компонентом системи комп'ютерного зору на основі штучного інтелекту. Відео, завантажені користувачами, обробляються для вилучення кадрів для подальшого аналізу, забезпечуючи сумісність з конвеєром розпізнавання виразу обличчя системи. У цьому розділі описано реалізацію робочого процесу завантаження та обробки медіаданих з використанням AWS S3, AWS Lambda та AWS MediaConvert. Використання даних сервісів дає можливості для масштабованості, автоматизації та економічної ефективності у майбутньому. Деталізацію розгортання всієї системи в хмарі AWS можна розглянути в додатку Е.

Завантаження починається з генерації попередньо підписаного автентифікованого посилання. Це посилання відкрите для PUT запитів безпосередньо до S3-бакета. PreSignedLinkFunction, лямбда-функція відповідає за створення цих попередньо підписаних посилань. Функція витягує з вхідного HTTP-запиту ідентифікатор користувача і час зйомки відео, перевіряє авторизацію, а тоді будує ключ S3 для відео і створює попередньо підписане посилання на попередньо визначене місце у S3. Також це посилання має певний час дії, обмежуючи несанкціонований доступ. Код на Java, що реалізує описаний робочий процес, показано нижче (рисунок 4.3).

```

27      @Override
28      public APIGatewayProxyResponseEvent handleRequest(APIGatewayProxyRequestEvent event,
29                                                         Context context) {
30          Map<String, String> queryParams = event.getQueryStringParameters();
31          String userId = queryParams.get("UserID");
32          Instant timeVideoTaken = Instant.ofEpochSecond(Long.parseLong(queryParams.get(
33              "TimeVideoTaken")));
34          String videoId = UUID.randomUUID().toString();
35          String s3Key = String.format("videos/%s/%s_%s.mp4", userId, videoId, timeVideoTaken);
36
37          context.getLogger().log(
38              String.format(
39                  "Input data: userId: %s, timeVideoTaken: %s, videoId: %s, s3Key: %s",
40                  userId,
41                  timeVideoTaken, videoId, s3Key),
42              LogLevel.INFO);
43
44          PresignedPutObjectRequest preSignedUrl = generatePreSignedUrl(s3Key);
45          ResponseBodyDto responseBody =
46              ResponseBodyDto.builder().preSignedLink(preSignedUrl.url().toString()).build();
47
48          return new APIGatewayProxyResponseEvent()
49              .withStatusCode(200)
50              .withBody(GSON.toJson(responseBody));
51      }
52
53      public PresignedPutObjectRequest generatePreSignedUrl(String s3Key) {
54          PutObjectPresignRequest preSignRequest = PutObjectPresignRequest.builder()
55              .signatureDuration(Duration.ofSeconds(LINK_ACTIVE_PERIOD))
56              .putObjectRequest(r -> r.bucket(System.getenv( name: "BUCKET_NAME")).key(s3Key))
57              .build();
58          return s3Presigner.presignPutObject(preSignRequest);
59      }

```

Рисунок 4.3 – Приклад коду на Java для робочого процесу

Цей процес гарантує безпечне завантаження відео до S3 з мінімальною затримкою і без впливу на сервер. Після генерації попередньо підписаної URL-адреси користувач завантажує своє відео безпосередньо до вказаного S3-бакета. Посилання дійсне лише протягом обмеженого періоду, після чого стає недійсним. Такий підхід мінімізує ризик несанкціонованого завантаження та спрощує процес завантаження медіафайлів для користувачів. Коли відео успішно збережеться в S3, генерується сповіщення і відправляється у StartMediaConvertJobFunction – лямбда функцію, яка відповідає за створення завдання MediaConvert для обробки завантаженого відео за наданою конфігурацією. Сповіщення містить метадані, які потрібні функції, для ідентифікації місця завантаженого відео. Грануюючи систему на завантаження та обробку, ми надіємося на майбутню масштабованість і надійність. Функція налаштовує завдання MediaConvert на нарізання кадрів з відео створюючи 1 кадр/секунду. Оброблені кадри зберігаються в окремій директорії S3.

Конфігурація завдання MediaConvert визначає вхідні та вихідні параметри, необхідні для обробки відео. Нижче наведено спрощене представлення конфігурації вхідного файлу (рисунок 4.4).



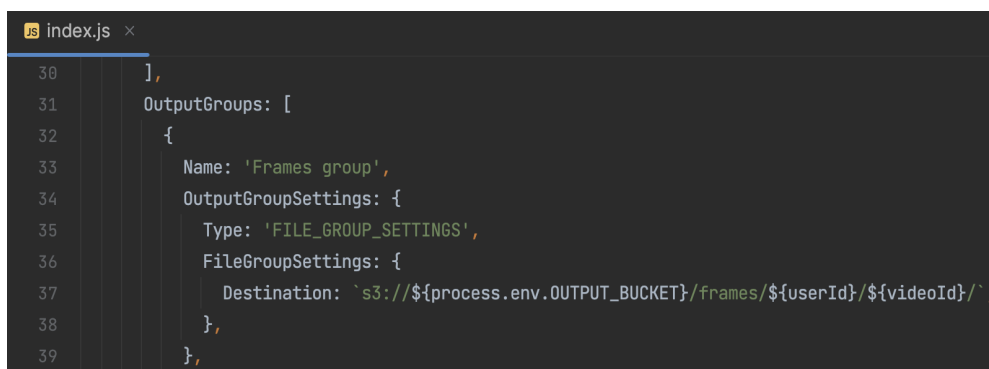
```

17  return {
18    Role: process.env.MEDIA_CONVERT_ROLE,
19    Settings: {
20      Inputs: [
21        {
22          FileInput: `s3://${bucket}/${key}`,
23          AudioSelectors: {
24            'Audio Selector 1': {
25              SelectorType: 'TRACK',
26              Tracks: [1]
27            }
28          }
29        },
30      ],

```

Рисунок 4.4 – Приклад конфігурації вхідного файлу в MediaConvert

Конфігурація регулює витягування та зберігання кадрів в структурованому вигляді, що робить їх легкодоступними для подальшої обробки за допомогою нейронних мереж. Згідно отриманого завдання MediaConvert сервіс обробляє відео відповідно до заданої конфігурації. Нижче на рисунку 4.5 наведено спрощене представлення конфігурації вихідного файлу.



```

30  ],
31  OutputGroups: [
32    {
33      Name: 'Frames group',
34      OutputGroupSettings: {
35        Type: 'FILE_GROUP_SETTINGS',
36        FileGroupSettings: {
37          Destination: `s3://${process.env.OUTPUT_BUCKET}/frames/${userId}/${videoId}/`,
38        },
39      },

```

Рисунок 4.5 – Конфігурація вихідного файлу для обробки в MediaConvert

Вищеописані завантаження та обробка медіафайлів, використані в розроблюваній системі комп'ютерного зору для обробки медіафайлів із обличчям

людини за допомогою ШІ, ефективно підготовлюють дані для подальшого використання. Попередньо підписане посилення є достатньо надійним заходом безпеки. Інтегровані S3-сповіщень і MediaConvert автоматизує конвеєр обробки підготовлюючи відео до аналізу з мінімальним ручним втручанням. Дана архітектура є ефективною та надійною та є міцною основою для виявлення емоцій та статистичного аналізу. Вона дозволяє в майбутньому змінювати частини та розміри застосунку.

4.3 Навчання ResNet на FER+ наборі даних

У розділі 3 було описано архітектури ResNet50 та ResNet101 чудовим вибором для задачі аналізу емоційного стану, використовуючи набір даних FER+. Ці моделі ефективно працюють із завданнями класифікації. А також, вони мають оптимальну архітектуру для обробки великого обсягу структурованої інформації. На даному етапі, буде описано аспекти навчання: підготовка даних, архітектурні рішення, основні труднощі та основні спостереження під час навчання, будуть використані стратегії оптимізації,.

FER+ містить у собі зображення розміром 48x48 пікселів у градаціях сірого. Ці дані розділено на сім емоцій: гнів, відраза, страх, щастя, смуток, здивування та нейтральність. Категорія «Огида» має недостатню кількість прикладів, створюючи ризик неточності моделі.

Нижче наведено кругову діаграму для візуалізації розподілу лейблів набору даних з відсотками кожної емоційної категорії (рисунок 4.6).

ResNet, скорочення від Residual Network, вирішує проблему зникнення градієнта, запроваджуючи швидкі з'єднання, що дозволяє глибоким архітектурам ефективно навчатися. Використовувалися обидві моделі ResNet50 і ResNet101, які відрізнялися насамперед глибиною та кількістю залишкових блоків.

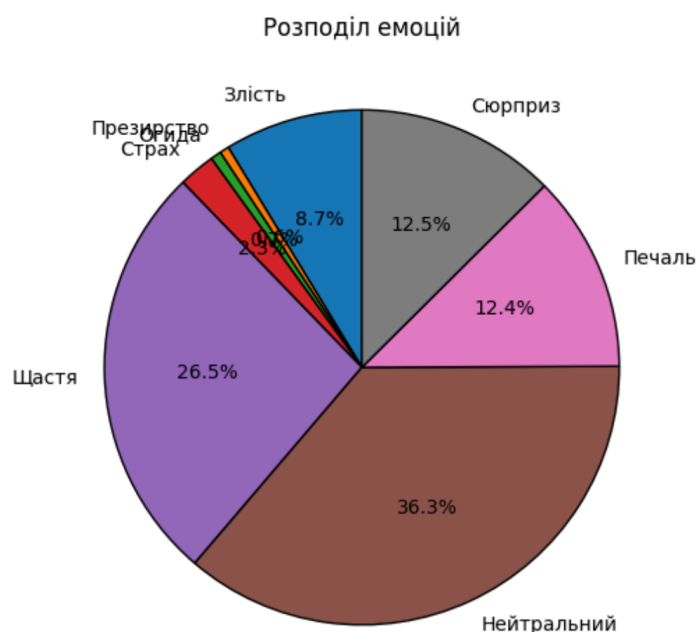


Рисунок 4.6 – Розподіл міток емоцій у FER+

В ResNet50 архітектура містить 50 шарів, що складаються з згорткових блоків (зі зменшенням дискретизації) та ідентифікаційних блоків (збереження розмірів). У таблиці 4.1 наведено порівняння показників ResNet50 та ResNet101.

Таблиця 4.1 – Порівняння характеристик ResNet50 та ResNet101

Характеристика	ResNet50	ResNet101
Час навчання на епоху	42 секунди	66 секунд
Остаточна точність навчання	95,73 %	93,13 %
Остаточна точність перевірки	96,34 %	90,04 %
Перевірка AUC (площа під кривою)	97,98 %	96,57 %

Обчислення для навчання на одну епоху при використанні ResNet50 швидші на 20 секунд ніж для ResNet101, де складність архітектури трішки підвищена. Модель ResNet50 досягла трішки кращої продуктивності з ResNet101 із зменшеними обчислювальними витратами. Це робить її ідеальною для розроблюваного проєкту, обраним через його ефективність у зближенні до глобального мінімуму, уникаючи локальних мінімумів. Обидві моделі були оптимізовані за допомогою стохастичного градієнтного спуску (SGD) з імпульсом 0,9 та швидкістю навчання 0,001. Функція втрат – CategoricalCrossentropy для

прискорення конвергенції шляхом використання градієнтів із попередніх ітерацій. Щоб забезпечити ранню зупинку та запобігти переобладнанню, навчання припинялося, якщо втрати точності не покращувалися після 10 епох. Моделі ResNet50 і ResNet101 навчалися протягом 50 епох. Моделі не були перенавчені, адже валідаційні точність, втрати та AUC (площа під кривою) поступово покращувались.

Модель ResNet50 продемонструвала швидшу конвергенцію та досягла дещо вищих показників перевірки, що підтверджує її придатність для поставленої задачі (рисунок 4.7).

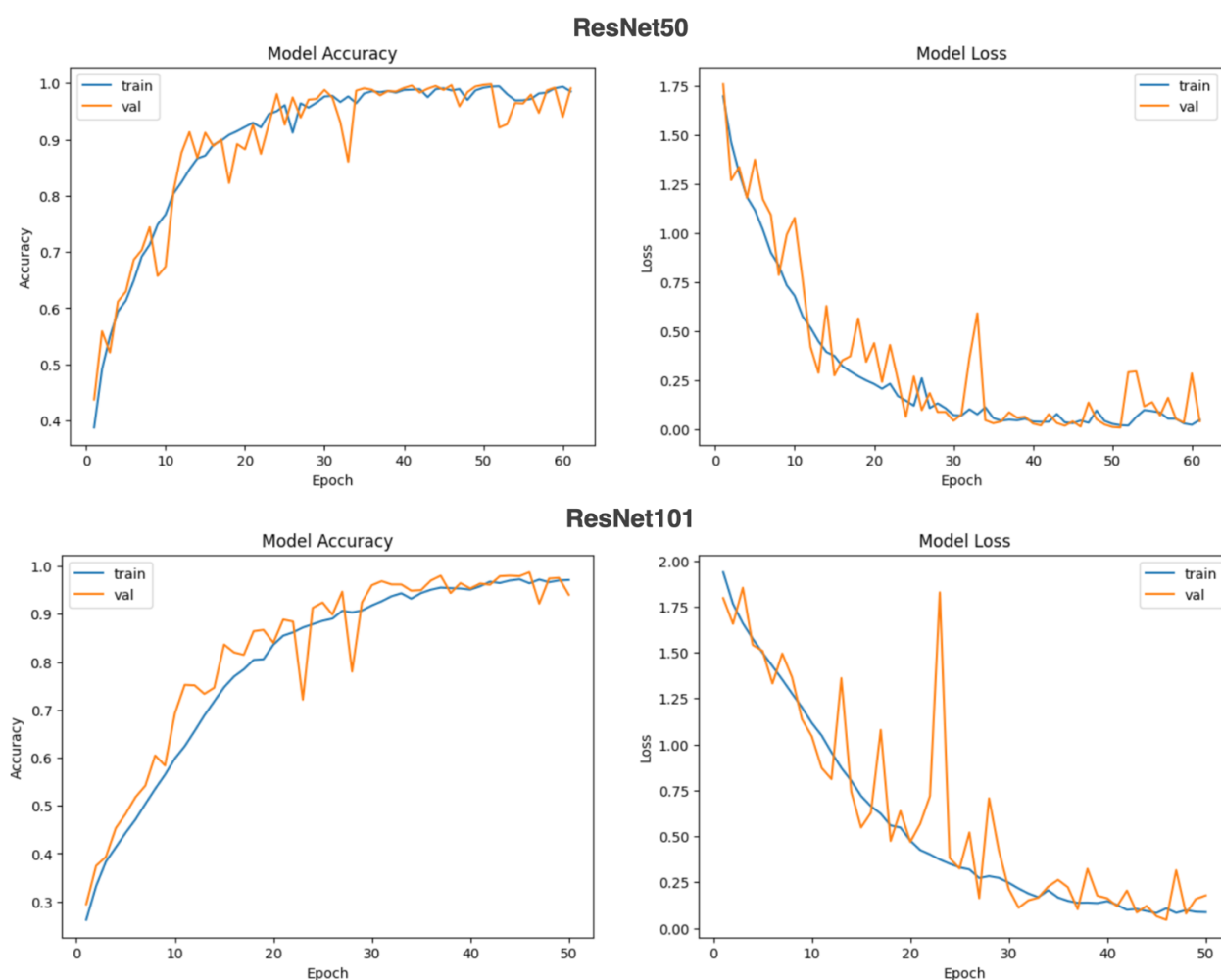


Рисунок 4.7 – Графіки точності та втрат моделей при навчанні та перевірці

У той час як більшість емоційних категорій досягли високої точності та запам'ятовування, категорія «Огида» та «Страх» стабільно мали меншу

ефективність через менший розмір вибірки. Такі методи, як передискретизація або зважування класів, можуть покращити продуктивність у майбутніх ітераціях.

Точність перевірки уважно відслідковувала точність навчання, що свідчить про те, що переобладнання було ефективно пом'якшено за допомогою методів ранньої зупинки та регуляризації.

AUC стабільно покращується протягом епох, перевищуючи 99,9% для обох моделей, що вказує на чудову здатність розпізнавання. Точність постійно перевищувала пригадування рідкісних емоцій, узгоджуючи з дисбалансом набору даних.

Матриці невідповідності (рисунок 4.8) були створені для аналізу ефективності моделі в емоційних категоріях. Модель ResNet50 продемонструвала чітку класифікацію більшості емоцій з незначними помилками в тісно пов'язаних категоріях, таких як «нейтральні» та «печаль».

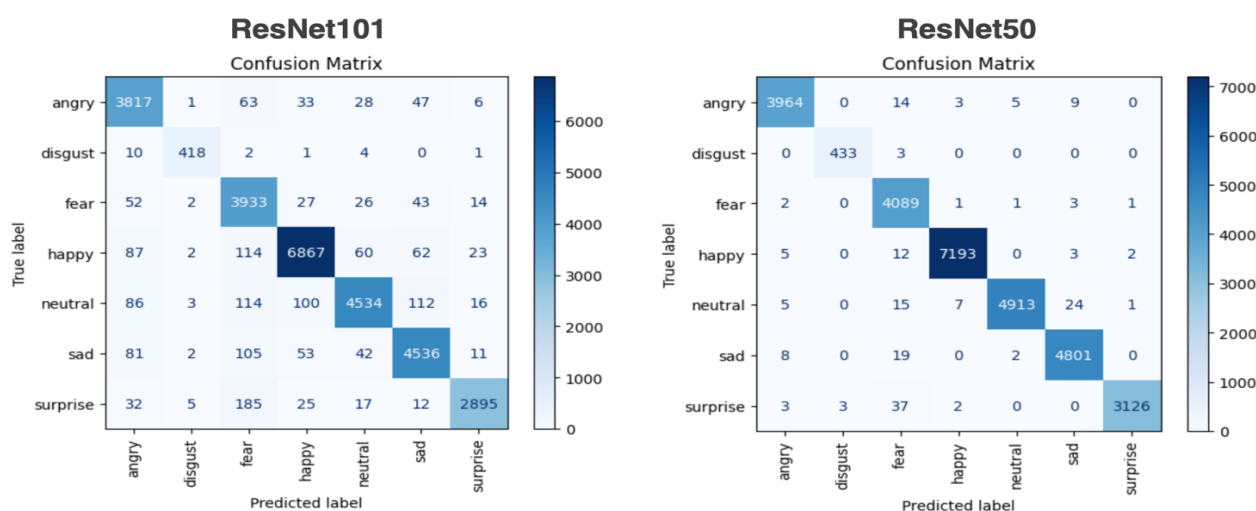


Рисунок 4.8 – Матриці невідповідності моделей ResNet101 та ResNet50

Моделі ResNet50 і ResNet101 пройшли успішне навчання на наборі даних FER+, продемонструвавши їхню ефективність у виявленні емоційних станів. ResNet50 було визнано більш придатним для практичних застосувань завдяки швидшому навчанню та порівняним показникам продуктивності. Майбутні вдосконалення можуть включати розширення набору даних і тонке налаштування гіперпараметрів для подальшого підвищення точності класифікації.

4.4 Навчання XLM-RoBERTa на обробленому наборі даних GoEmotions

Процес навчання моделі XLM-RoBERTa класифікації емоцій на основі українського тексту включав кілька ключових кроків: від підготовки та обробки набору даних до впровадження ефективних стратегій навчання. У цьому розділі описано методологію, використану для адаптації набору даних GoEmotions, узгодження його лейблів із FER+ для мультимодального розпізнавання емоцій та навчання моделі з високою точністю на тексті перекладеному українською мовою.

Набір даних GoEmotions, еталон для класифікації емоцій, був обраний через широкий спектр категорій емоцій і включення звичайного розмовного тексту. Набір даних містить 27 категорій емоцій разом із нейтральною міткою та містить текстовий вміст із емодзі, сленгом та нецензурною лексикою. Ці елементи були збережені під час попередньої обробки, щоб зберегти контекст природної мови.

27 міток емоцій у GoEmotions були зіставлені з 8 мітками, які використовувалися у FER+. Такі мітки, як «радість» і «хвилювання», були згруповані під «ЩАСЛИВО», а «гнів» і «лють» були вирівняні з «ЗЛІСТЬ». Це відображує, що обидва набори даних можуть використовуватися узгоджено в системі для єдиного мультимодального аналізу емоцій (рисунок 4.9).

```

29 emotion_mapping = {
30     'Happiness': ['admiration', 'amusement', 'caring', 'desire', 'excitement',
31                 'gratitude', 'joy', 'love', 'optimism', 'pride', 'relief'],
32     'Anger': ['anger', 'annoyance'],
33     'Neutral': ['approval', 'realization', 'neutral'],
34     'Surprise': ['confusion', 'curiosity', 'surprise'],
35     'Sadness': ['disappointment', 'embarrassment', 'grief', 'remorse', 'sadness'],
36     'Disgust': ['disapproval', 'disgust'],
37     'Fear': ['fear', 'nervousness']

```

Рисунок 4.9 – Співставлення міток емоцій з GoEmotions а FER+

Щоб адаптувати набір даних для розпізнавання емоцій українською мовою, зразки тексту було автоматично перекладено за допомогою Microsoft Azure Translator. Хоча автоматичний переклад може створювати шум, він пропонує

масштабований спосіб підготувати дані такою мовою з обмеженими ресурсами, як українська. Основні міркування включали забезпечення точного перекладу емоційно насичених фраз і збереження контекстуальних Емоїї (рисунок 4.10).

```

52 const translate = async index => {
53   try {
54     const content = fs.readFileSync(`data/data-${index}.txt`, 'utf-8')
55
56     if (!content) return;
57
58     const contentLength = splitData(content).length
59
60     const response = await axios({
61       baseURL: TRANSLATOR_URL,
62       url: '/translate',
63       method: 'post',
64       headers: {
65         'Ocp-Apim-Subscription-Key': API_KEY,
66         'Ocp-Apim-Subscription-Region': REGION,
67         'Content-type': 'application/json',
68         'X-ClientTraceId': uuidv4().toString()
69       },
70       params: { 'api-version': '3.0', 'from': 'en', 'to': 'uk' },
71       data: [{ 'text': content }],
72       responseType: 'json'
73     })
74
75     const responseData = response.data[0].translations[0].text
76     const responseDataLength = splitData(responseData).length
77
78     fs.writeFileSync(`data/output-${index}.txt`, responseData)
79
80     if (contentLength !== responseDataLength) {
81       console.log(`${contentLength} !== ${responseDataLength}`);
82       process.exit(1)
83     }
84   } catch (e) {
85     console.log(e);
86   }
87 }

```

Рисунок 4.10 – Приклад асинхронного перекладу тексту

Текст було токенізовано за допомогою спеціального для XLM-RoBERTa токенайзера `RobertaTokenizerFast`, який обирається автоматично з бібліотеки трансформерів `HuggingFace` за програмним кодом нижче (рисунок 4.11).

```

237 def init_model(self):
238     try:
239         self.tokenizer = AutoTokenizer.from_pretrained(self.model_name)
240

```

Рисунок 4.11 – Код ініціалізації токенайзера для XLM-RoBERTa

Емоїї були збережені в тексті та закодовані як спеціальні токени, щоб модель могла дізнатися про їхнє емоційне значення. Нецензурні та розмовні терміни залишилися без змін, що відображає реальне використання та емоційну насиченість. Отриманий набір даних забезпечив міцну основу для навчання моделі XLM-RoBERTa (рисунок 4.12) з вирівняними мітками та мовним різноманіттям, яке відображає реальні сценарії.

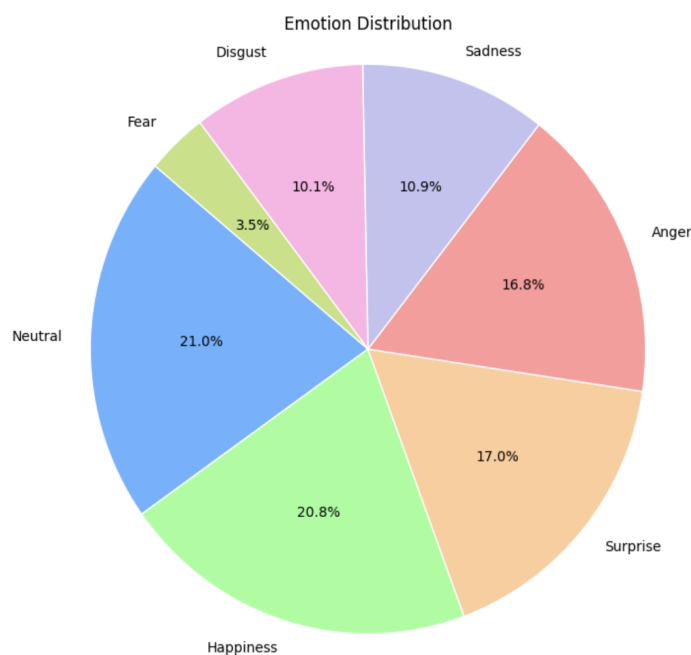


Рисунок 4.12 – Діаграма розподілу емоційних класів

Було обрано механізм балансування дистриб'юції рядків Oversampling, що допомагає псевдовипадково обрати рядок і дублювати його, щоб підняти до вищого рівня рядків певної емоції. Саме тому запропонований датасет мультилейбловий, адже розповсюдження далеко не ідеальне навіть після повторної вибірки (resampling).

XLM-RoBERTa, багатомовний варіант RoBERTa, було обрано через його здатність ефективно вирішувати міжмовні завдання. Архітектура складається з трансформатора кодування з 24 шарами, 16 головками уваги та розміром вбудовування 1024. Ця конфігурація дозволяє моделі фіксувати складні лінгвістичні нюанси та ідентифікувати контекстуальні зв'язки в тексті.

Оптимізатор AdamW використовувався під час навчання через його здатність ефективно обробляти великомасштабні моделі, такі як XLM-RoBERTa. AdamW поєднує адаптивні темпи навчання з регуляцією зниження ваги, що запобігає переобладнанню та забезпечує стабільну конвергенцію. Були обрані такі параметри:

- швидкість навчання: $1e-5$;
- кроки розминки: 500 (для поступового збільшення темпу навчання);
- розмір батчу: 32;
- значення `weight_decay=0.02` є компромісом між достатньою

регуляризацією та збереженням здатності моделі вивчати складні залежності в даних, занадто високе значення може обмежити модель занадто сильно, перешкоджаючи їй вивченню важливих патернів, занадто низьке значення може бути недостатнім для запобігання перенавчанню, особливо на невеликих або шумних датасетах.

Навчання проводилося на графічному процесорі NVIDIA T4 з 16 ГБ пам'яті, що дозволило ефективно працювати з великою моделлю та набором даних. Накопичення градієнта (`gradient_accumulation_steps=2`) використовувалося для підтримки керованого розміру партії при максимальному використанні GPU.

Функцію `CrossEntropyLoss` було обрано для обчислення помилки для завдання багатокласової класифікації емоцій.

Ефективність моделі оцінювалася за допомогою точності, оцінки втрат, `Gradient Norm` запам'ятовування (для оцінки стабільності навчання).

Щоб запобігти переобладнанню, було застосовано ранню зупинку, відстежуючи втрати перевірки з порогом 5 епох.

Під час навчання модель досягла послідовного покращення продуктивності в різні епохи. Після точного налаштування спостерігалася точність перевірки приблизно 59%, з точністю 60%. Хоча текст з автоматичним перекладом може вносити деякі неточності, висока багатомовність XLM-RoBERTa пом'якшила ці наслідки.

Завдяки узгодженню міток GoEmotions із FER+ система забезпечує інтегрований аналіз тексту та зображень. Таке узгодження має вирішальне значення для поєднання текстового аналізу настроїв із розпізнаванням виразу обличчя та для формування цілісної картини емоційних станів.

Навчання XLM-RoBERTa на українському наборі даних GoEmotions демонструє адаптивність і надійність моделі для завдань розпізнавання емоцій. Використовуючи засоби перекладу, передові методи токенізації та високопродуктивний оптимізатор, було створено надійну основу для класифікації емоцій в українському тексті. Ця інтеграція з FER+ закладає основу для синергічного мультимодального аналізу емоцій.

4.5 Аналіз медіафайлів використовуючи навчені моделі

Нижче описано процес обробки відеофайлів з використанням сервісів AWS для виявлення емоцій в аудіо та відеокадрах, зберігання результатів та надсилання сповіщень. Візуалізація послідовності подій зображена у додатку Ж.

Відеофайл зберігається в AWS S3, після чого завдання обробки відео запускається за допомогою AWS Elemental MediaConvert. Відео читається з S3 та передається до MediaConvert для подальшої обробки, де воно розбивається на окремі кадри за допомогою AWS Lambda. Кожен кадр аналізується для виявлення емоцій за допомогою функції AWS Lambda "Detect Emotions In Frames", а аудіотрек відео аналізується функцією "Detect Emotions In Audio". Результати аналізу зберігаються в AWS DynamoDB. Координація процесів здійснюється через AWS Step Functions, а події управляються за допомогою AWS EventBridge. Після успішного завершення обробки, сповіщення надсилаються відповідним отримувачам. Для обробки кількох кадрів використовується механізм циклу, який повторює процес аналізу для кожного кадру.

Після успішного завершення завдання MediaConvert подія надсилається до Amazon EventBridge, який запускає машину станів. Ця інтеграція гарантує, що

аналіз емоційного стану починається автоматично, не вимагаючи ручного втручання, підтримуючи безперебійний та ефективний робочий процес.

Аналіз емоційних станів є ключовим елементом системи комп'ютерного зору на основі штучного інтелекту. Цей процес включає в себе аналіз вхідних візуальних і звукових даних для вишулення та оцінки емоційного змісту. Для цього машина станів AWS Step Functions організовує робочий процес, підтримуваний лямбда-функціями, EFS і DynamoDB для ефективної обробки, зберігання та аналізу даних (рисунок 4.13). Робочий процес запускається EventBridge після завершення завдання MediaConvert, гарантуючи, що весь конвеєр залишається масштабованим та керованим подіями.

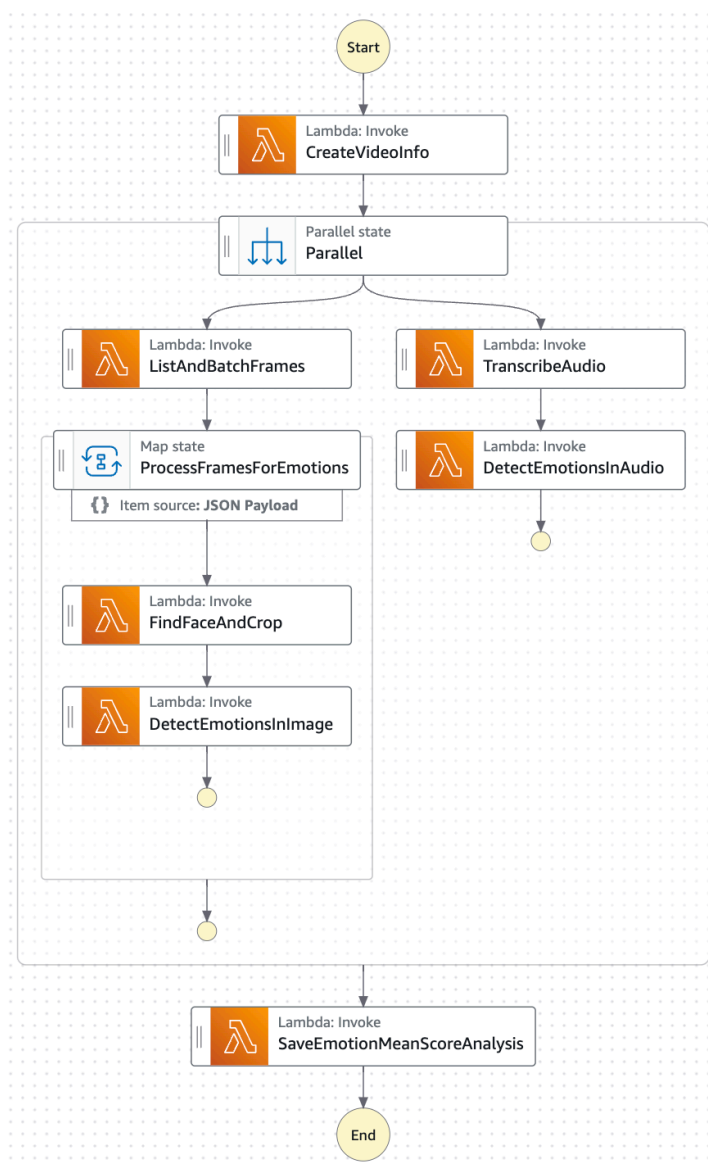


Рисунок 4.13 – Діаграма станів згенерована сервісом AWS Step Function

У таблиці 4.2 наведено основні компоненти автомата станів, які використовуються для аналізу емоційних станів.

Таблиця 4.2 – Ключові компоненти автомата станів

Компонент	Опис дій
CreateVideoInfo	Ініціює новий документ для відео даних у таблиці Video DynamoDB та додає наявну інформацію, наприклад час створення відео.
ListAndBatchFrames	Читає кадри з S3, згенеровані завданням MediaConvert, і організовує їх у керовані пакети використовуючи шлях, вказаний у виході MediaConvert. Ці кадри групуються в пакети, що полегшує обробку і зменшує навантаження на пам'ять.
ProcessFramesForEmotions	Використовує стан Map для паралельної обробки пакетів. Паралельне виконання надає одночасний аналіз декількох кадрів, покращуючи продуктивність.
FindFaceAndCrop	Застосовується для кожного кадру в батчі. Використовуються методи комп'ютерного зору для виявлення облич на зображеннях, їх обрізання та збереження результатів у S3. Тому лише релевантні частини кадрів будуть проаналізовані на наявність емоцій, зменшуючи обчислювальну складність.
DetectEmotionsInImage	Обробляє обрізані зображення за допомогою моделі ResNet50 для виявлення та аналізу емоційних виразів.
TranscribeAudio	Транскрибує аудіо, витягнуте з відео, використовуючи модель Whisper OpenAI
DetectEmotionsInAudio	Виявляє емоційний зміст у транскрипції з мовлення за допомогою XLM-RoBERTa
SaveEmotionMeanScoreAnalysis	Агрегує виявлені емоції у всіх кадрах відео для обчислення середнього значення кожної емоції. Ці агреговані дані зберігаються в базі даних DynamoDB. Звідки потім є доступ для створення статистичних висновків.

EFS використовується як рішення для спільного зберігання тяжких для AWS Lambda моделей нейронних мереж (понад 300-400 МБ, максимум 500 МБ для усіх підключених бібліотек та моделей).

DynamoDB зберігає остаточні результати емоційного аналізу, включаючи середні оцінки емоцій для кожного відео. Її безсхемний дизайн і здатність працювати з високою пропускнуою здатністю читання/запису роблять її ідеальною для зберігання і запиту динамічних емоційних даних.

Завдання аналізу емоційного стану ефективно поєднує сервіси AWS, включаючи Step Functions, Lambda, MediaConvert, EFS і DynamoDB, для обробки мультимедійних даних у масштабований, автоматизований спосіб. Використання паралельної обробки, ефективного зберігання даних та моделей машинного навчання гарантує, що візуальні та аудіо емоційні сигнали аналізуються всебічно. Така реалізація відповідає меті системи – забезпечити точний і глибокий емоційний моніторинг для користувачів. Для розпізнання тексту в аудіодоріжці використовується Whisper від OpenAI. Модель, яка використовується є або large, або medium в залежності від вимоги обробки (рисунок 4.3).

Таблиця 4.3 – Опис та порівняння якості роботи моделей

Розмір моделі	Час обробки	Текст
large	2схв	Наш онлайн записувач є зручним і простим онлайн застосунком, який може бути використаний прямо у вашому браузері. Це дозволяє вам записувати звук, використовуючи мікрофон, і зберігати це як mp3 файл.
medium	30 сек	Наш онлайн записувач є зручним і простим онлайн застосунком, який може бути використаний прямо у вашому браузері. Це дозволяє вам записувати звук використовуючи мікрофон і зберігати це як mp3 файл.
small	20 хв	Наш онлайн записувач є срочним і простим онлайн застосунком, який може бути використаним прямо у вашому браузері. Це дозволяє вам записувати звук використовуючи мікрофон і зберігати це як mp3file.
tiny	10 сек	Наш онлайн записувач є срочним і простим онлайн застосонку мікій може бути використеним праву у вуважому бралозері. Тодо зваляє вам записувати звук, використовувачими крофон і збрігадеця як MP3File.

Аналіз емоційних станів у розробленій системі опирається на кінцевий автомат, який керує двома паралельними робочими процесами: розпізнавання емоцій на основі тексту з аудіо та відео, та розпізнавання емоцій на основі

зображень з кадрів відео. Подвійний процес улавлює та аналізує і мовлення, і візуальні емоційні сигнали, забезпечуючи повне розуміння емоційного стану користувача.

Результати аналізу емоційного стану організовано в окремій таблиці DynamoDB. Кожна з них оптимізована для типу даних, що зберігаються. Ознайомитися із візуалізацією даталогічної моделі можна у додатку И. Такий поділ забезпечує модульність, масштабованість і ефективне надсилання запитів:

- текстові результати, включаючи транскрибований текст і виявлені емоції, зберігаються в таблиці SpeechText. Кожен запис містить: витягнутий текст зі звуку відео та позначки емоцій і відповідні оцінки впевненості;
- результати на основі зображень зберігаються в таблиці Image. Кожен запис містить: шлях до обробленого зображення в S3 та виявлені емоції;
- після аналізу тексту та зображень середній емоційний стан для відео обчислюється шляхом поєднання результатів із таблиць SpeechText і Image. Цей сукупний результат зберігається в таблиці «Відео» в атрибуті «Аналіз», надаючи підсумок загального емоційного вмісту відео. Оцінки аудіо та зображення об'єднуються, беручи їхнє середньозважене значення, надаючи пріоритет аналізу на основі зображень через сильнішу кореляцію виразів обличчя з конкретними емоційними станами. Остаточні результати потім зберігаються в таблиці Video, що робить їх легко доступними для подальшого аналізу або візуалізації.

Використання XLM-RoBERTa для аналізу на основі тексту та ResNet для аналізу на основі зображень у поєднанні зі структурованою організацією даних DynamoDB забезпечує надійний і комплексний аналіз емоційного стану. Агрегуючи результати на рівні відео, система забезпечує інтегрований емоційний огляд, деталізований і масштабований. Такий підхід узгоджується з метою дисертації – забезпечити точне мультимодальне емоційне розуміння.

4.6 Перегляд статистики аналізу обробки медіафайлів

Надання користувачам доступної та детальної інформації про їхній емоційний стан є критично важливим компонентом системи комп'ютерного зору

на основі штучного інтелекту. Реалізація передбачає зчитування даних, що зберігаються в DynamoDB, та їх візуалізацію за допомогою зручного інтерфейсу інформаційної панелі. Інтерфейс відображає агреговану емоційну статистику, отриману з відео, мови та зображень за вибраний період часу. Крім того, необроблені дані аналізу емоцій обробляються і візуалізуються, щоб запропонувати комплексне розуміння емоційних патернів користувача.

Всі результати емоційного аналізу зберігаються в Amazon DynamoDB. Кожен запис у базі даних містить такі деталі, як тип проаналізованого медіа (відео, мова або зображення), часові мітки, емоційні оцінки та пов'язані з ними метадані. Коли користувач обирає певний діапазон дат на інформаційній панелі, виконується запит до DynamoDB, щоб отримати дані, які відповідають ідентифікатору користувача та зазначеному періоду.

Безсхемна архітектура DynamoDB полегшує зберігання структурованих даних, забезпечуючи при цьому високу пропускну здатність і низьку затримку для запитів. Результати запитів повертаються у форматі JSON, який включає детальні емоційні оцінки та їхні часові мітки. Приклад структури відповіді від серверу наведено нижче на рисунку 4.14.

```

1  {
2  "videos": [
3  {
4  "Analysis": {...},
88 "VideoKey": {"S": "videos/serhiyo/2008ed50-e339-44e3-b018-e6a4df053d3e_2024-11-23T10:32:25Z.mp4"...},
91 "TakenTimestamp": {"N": "1732357945"...},
94 "VideoId": {"S": "9c13aef3-5f97-4823-b021-c3df0ae96ac"...}
97 },
98 {...},
193 {...},
288 {...},
383 {...}
478 ],
479 "images": [
480 {"imageContent": "/9j/4AAQSkZJRgABAQAAQABAAD/2wBDAAgGBgcGBQgHBwcJCQgKDBQNDAsLDBkSEw8UHRofHh0aHBwgJC4nI
576 {"imageContent": "/9j/4AAQSkZJRgABAQAAQABAAD/2wBDAAgGBgcGBQgHBwcJCQgKDBQNDAsLDBkSEw8UHRofHh0aHBwgJC4nI
672 {"imageContent": "/9j/4AAQSkZJRgABAQAAQABAAD/2wBDAAgGBgcGBQgHBwcJCQgKDBQNDAsLDBkSEw8UHRofHh0aHBwgJC4nI
768 {"imageContent": "/9j/4AAQSkZJRgABAQAAQABAAD/2wBDAAgGBgcGBQgHBwcJCQgKDBQNDAsLDBkSEw8UHRofHh0aHBwgJC4nI
864 {"imageContent": "/9j/4AAQSkZJRgABAQAAQABAAD/2wBDAAgGBgcGBQgHBwcJCQgKDBQNDAsLDBkSEw8UHRofHh0aHBwgJC4nI
960 {"imageContent": "/9j/4AAQSkZJRgABAQAAQABAAD/2wBDAAgGBgcGBQgHBwcJCQgKDBQNDAsLDBkSEw8UHRofHh0aHBwgJC4nI
1056 ],
1057 "userId": "serhiyo",
1058 "startDate": 1732350000,
1059 "endDate": 1732359259
1060 }

```

Рисунок 4.14 – Приклад відповіді від серверу з аналізом емоційного стану

Потім ці дані обробляються на сервері для обчислення середніх значень для кожної емоції та підготовки їх до візуалізації на інформаційній панелі. Інформаційна панель призначена для представлення результатів емоційного аналізу в зрозумілій та цікавій формі, її вигляд можна побачити на рисунку 4.15.



Рисунок 4.15 – Приклад візуалізації Інформаційної панелі

Вона складається з наступних компонентів:

- полігони для аналізу медіа: на інформаційній панелі відображаються три окремі полігони, що представляють емоційні стани, отримані на основі: Кожен полігон відкладає емоції (наприклад, СПОКІЙ, ЩАСТЯ, СУМ) уздовж своїх осей, а відповідні оцінки інтенсивності представлені як відстані від центру, це дає швидкий візуальний підсумок емоційних патернів користувача в різних типах медіа;
- агреговані емоції, виявлені в кадрах, витягнутих із завантажених користувачем відео;
- емоційні інсайти, отримані з транскрипцій мовлення;
- результати аналізу завантажених або витягнутих зображень;

— за вибраний період часу на інформаційній панелі відображається середня оцінка для кожної емоції, це дозволяє користувачам зрозуміти свої домінуючі емоційні стани з плином часу, надаючи інформацію, яка може вплинути на поведінкові патерни або зусилля з емоційної регуляції;

— користувачі можуть вибрати певний діапазон дат для фільтрації результатів аналізу, дані з DynamoDB запитуються відповідно, а інформаційна панель оновлюється, щоб відобразити обраний період часу.

Сирі JSON-дані, отримані з DynamoDB, обробляються у формат, придатний для візуалізації. Емоційні оцінки зображень усереднюються за всіма проаналізованими кадрами протягом періоду. Аналіз емоційності мовлення об'єднує результати декількох транскрипцій. Аналіз відео об'єднує результати на рівні кадрів та транскрипції, щоб отримати загальну емоційну оцінку для кожного відео. Ці агреговані результати відображаються у вигляді графіків або ж як показано на рисунку 4.16 у вигляді історії емоційних станів з можливістю сортування, пошуку тощо.

EMOTIONS ІНФОРМАЦІЙНА ПАНЕЛЬ ІСТОРІЯ ЕМОЦІЙ						
Час створення	Джерело	Тип медіа	Емоції	Вміст	Дії	
2024-11-23 12:32:25	video	Зображення	Спокій: 99.09% Спонтанічність: 0.68% Сум: 0.19% Здивування: 0.02% Злість: 0.01% Відраза: 0.00% Щастя: 0.00% Страх: 0.00%		 	
2024-11-23 12:32:26	video	Зображення	Здивування: 99.83% Спонтанічність: 0.14% Щастя: 0.03% Спокій: 0.00% Страх: 0.00% Відраза: 0.00% Злість: 0.00% Сум: 0.00%		 	
2024-11-23 12:32:27	video	Зображення	Спонтанічність: 93.76% Спокій: 2.93% Злість: 2.46% Сум: 0.69% Відраза: 0.07% Здивування: 0.07% Щастя: 0.00% Страх: 0.00%		 	
2024-11-23 12:32:28	video	Зображення	Спонтанічність: 76.39% Спокій: 22.07% Здивування: 0.92% Злість: 0.37% Сум: 0.22% Відраза: 0.02% Щастя: 0.01% Страх: 0.00%		 	

Рисунок 4.16 – Приклад візуалізації історії аналізу емоційних станів

Використання чітких та інтерактивних графіків для візуалізації емоційного аналізу медіафайлів, а також агреговані середні показники для кожної емоції роблять систему інтуїтивно зрозумілою з таким способом інтерпретації.

Висновки до розділу 4

У результаті проведеної реалізації було розроблено систему обробки медіафайлів із обличчям людини за допомогою нейронних мереж. Це відповідає поставленим вимогам у розділі 2: щодо автентифікації користувачів, завантаження та обробки медіаданих, навчання моделей розпізнавання емоцій, а також надання користувачам доступу до їхніх емоційних статистик.

Перш за все, було впроваджено процес реєстрації та авторизації користувачів, беручи за основу сервіс Amazon Cognito, який забезпечує надійну та безпечну автентифікацію.

Далі було реалізовано механізм завантаження та обробки медіафайлів з використанням сервісів AWS S3, Lambda та MediaConvert. Користувачам було надано можливість завантажувати відео та зображення через попередньо автентифіковані посилання, а тому їх медіадані могли швидше зберегтись у S3.

Для задач розпізнавання емоцій були навчені моделі ResNet50 та XLM-RoBERTa на відповідних наборах даних FER+ та GoEmotions. Модель ResNet50 використана для класифікації виразів обличчя, а XLM-RoBERTa – для ефективного аналізу емоційного змісту тексту українською мовою. Завдяки цьому було досягнуто високу точність розпізнавання емоцій на основі візуальних та текстових даних. Аналіз емоційного стану користувачів здійснювався за допомогою інтегрованих сервісів AWS, таких як Step Functions та EventBridge. Результати аналізу зберігалися в базі даних DynamoDB, що забезпечувало швидкий доступ до даних та їх ефективне управління.

Нарешті, було розроблено інформаційну панель, яка надає користувачам доступ до детальної статистики їхніх емоційних станів. Інтерфейс панелі включає візуалізації у вигляді графіків, що дозволяють легко інтерпретувати емоційні патерни та відстежувати зміни емоцій з часом.

Таким чином, реалізована система обробки медіафайлів із обличчям людини за допомогою нейронних мереж реалізувалась ефективною, безпечною та масштабованою мультимодальною системою.

5 РОЗРОБЛЕННЯ СТАРТАП-ПРОЄКТУ

5.1 Опис ідеї проєкту

Ідеєю стартап-проєкту є розробка інноваційної система комп'ютерного зору для обробки медіафайлів із обличчям людини, яка базується на технологіях штучного інтелекту.

Ця система покликана об'єднати аналіз мультимодальних даних – міміки та тексту, – з метою надання комплексного аналізу емоційного стану користувача. Система має на меті вирішити проблему мультимодального розпізнавання емоцій, що робить її корисною у різноманітних середовищах: від особистого користування до професійної діяльності. Детальний опис ідеї можна побачити у таблиці 5.1.

Таблиця 5.1 – Опис ідеї стартап-проєкту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Розробка системи комп'ютерного зору для обробки медіафайлів із обличчям людини за допомогою нейронних мереж	Моніторинг психоемоційного стану у сфері охорони здоров'я	Своєчасне виявлення емоційних проблем, покращення доступу до послуг психічного здоров'я, автоматизація аналізу емоційного стану
	Використання у професійній сфері для аналізу емоційної атмосфери на робочому місці	Підвищення емоційного інтелекту працівників, зниження рівня стресу в команді, оптимізація колективної роботи завдяки виявленню емоційної динаміки
	Інтеграція в системи клієнтського сервісу для персоналізованого підходу до клієнтів	Підвищення задоволеності клієнтів, швидке реагування на незадоволення, підвищення ефективності роботи операторів
	Інструмент для індивідуального	Саморегуляції емоцій, підтримка при прийнятті рішень, врахування власного емоційного стану для

Зміст ідеї	Напрямки застосування	Вигоди для користувача
	моніторингу емоцій у повсякденному житті	розуміння та впливу на особисті та професійні ситуації

Система вирізняється мультимодальністю, яка дозволяє аналізувати міміку і текст одночасно, забезпечуючи точний емоційний профіль. Вона автоматизує процес без залежності від самозвітів, гарантує об'єктивність та адаптується до індивідуальних емоційних патернів користувачів.

Таблиця 5.2 містить порівняння ідеї проєкту з уже існуючими рішеннями потенційних конкурентів.

Таблиця 5.2 – Визначення сильних, слабких та нейтральних характеристик ідеї проєкту

№ п/п	Техніко-економічні характеристики ідеї	Товари та концепції потенційних конкурентів			W	N	S
		Мій проєкт	Affectiva	Replika			
1.	Мультимодальний аналіз даних	Аналіз міміки і тексту	Аналіз міміки та голосу, без тексту	Аналіз лише тексту			+
2.	Робота в реальному часі	Частково, із затримкою	Частково, із затримкою	Частково		+	
3.	Персоналізація для кожного користувача	Індивідуальні профілі та адаптація до користувача	Мінімальна адаптація	Часткова адаптація		+	
4.	Автоматизація збору даних	Автоматизований процес, не	Залежність від ручного завантаження	Залежність від			+

№ п/ п	Техніко- економічні характеристики ідеї	Товари та концепції потенційних конкурентів			W	N	S
		Мій проєкт	Afectiva	Replika			
		залежить від користувача		текстового вводу			
5.	Конфіденційність даних	Високий рівень захисту даних (AWS, відповідність GDPR)	Середній рівень захисту	Середня конфіденційність			+
6.	Доступність для масового використання	Середній рівень доступності	Високий рівень доступності	Високий рівень доступності	+		

5.2 Технологічний аудит ідеї проєкту

Для реалізації ідеї проєкту важливим кроком є проведення технологічного аудиту, щоб оцінити наявні ресурси, визначити технічні аспекти та порівняти ідею з існуючими рішеннями. Виявлені сильні та слабкі сторони проєкту, обґрунтування конкурентних переваг продемонстровані у таблиці 5.3.

Таблиця 5.3 – Технологічна здійсненність проєкту

№ п/ п	Ідея проєкту	Технології її реалізації	Наявність технологій	Доступність технологій
1.	Мультимодальний аналіз емоцій (міміка, текст)	Нейронні мережі (напр., CNN для міміки, трансформери для тексту)	Існують, широко використовуються	Доступність бібліотек для машинного навчання

№ п/ п	Ідея проєкту	Технології реалізації	Наявність технологій	Доступність технологій
2.	Робота в реальному часі	Інтеграція з хмарними сервісами (AWS Lambda) та оптимізовані моделі ШІ	Існують, доступні на ринку	Доступні через підписку на хмарні сервіси AWS
3.	Автоматизація збору та аналізу даних	Автоматизовані скрипти для збору даних, API хмарних сервісів для обробки мультимодальних даних	Існують частково, потребують налаштування	Доступні, але вимагають налаштування та інтеграції
4.	Забезпечення конфіденційності та безпеки даних	AWS Identity and Access Management (IAM), шифрування даних на рівні хмарних сховищ	Існують, стандарт у хмарній індустрії	Доступні через AWS, відповідно до стандартів безпеки (GDPR)
5.	Персоналізація аналізу для користувача	Моделі машинного навчання на даних користувача	Існують, але потребують тренування	Доступні з використанням відкритих ML- фреймворків

Всі потрібні технології, мультиплатформні, з високою продуктивністю при роботі з мультимодальними даними та підтримкою з боку спільноти розробників, можуть на ринку можуть бути використані для реалізації проєкту.

5.3 Аналіз ринкових можливостей запуску стартап-проєкту

Оскільки на ринку вже існують інші проєкти, що надають подібні послуги, важливо провести аналіз ринкових можливостей і потенційних загроз, які можуть виникнути під час запуску стартапу.

Першим етапом цього аналізу є дослідження попиту, результати якого подано в таблиці 5.4.

Таблиця 5.4 – Попередня характеристика потенційного ринку стартап-проєкту

№ п/п	Показники стану ринку (найменування)	Характеристика
1.	Кількість головних гравців, од	5
2.	Загальний обсяг продаж, грн/ум.од	Близько 10 млрд грн/рік у сфері технологій емоційного інтелекту
3.	Динаміка ринку (якісна оцінка)	Зростає
4.	Наявність обмежень для входу (вказати характер обмежень)	Висока конкуренція, необхідність значних інвестицій у маркетинг та технології
5.	Специфічні вимоги до стандартизації та сертифікації	Відповідність GDPR (захист даних), сертифікація для медичних застосувань
6.	Середня норма рентабельності в галузі (або по ринку), %	12-15

За результатами аналізу ринок виглядає перспективним для входження, але потребує чіткої стратегії позиціонування та виконання стандартів.

Наступним етапом є ідентифікація потенційних груп клієнтів, їх основних характеристик, а також формування орієнтовного списку вимог до сервісу для кожної з цих груп. Результати аналізу подано в таблиці 5.5.

Таблиця 5.5 – Характеристика потенційних клієнтів стартап-проєкту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1.	Потреба у моніторингу психоемоційного стану для здоров'я	Лікарі, психологи, клініки, пацієнти з тривожними розладами	Дотримання медичних стандартів (GDPR, сертифікація). Фокус на простоті інтеграції з іншими медичними інструментами	Висока точність аналізу, простота використання. Відповідність стандартам, надійна підтримка
2.	Покращення емоційного інтелекту у професійній сфері	HR-менеджери, корпоративні тренери, керівники команд	Оцінка емоційної атмосфери команди. Важлива інтеграція з корпоративними платформами (наприклад, Slack, CRM)	Масштабованість, легка інтеграція. Доступність консультацій та навчання користувачів
3.	Потреба у персоналізован	Індивідуальні користувачі, які цікавляться	Орієнтація на доступність та конфіденційність	Персоналізація, доступність на всіх пристроях

№ п/ п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
	ій емоційній підтримці	особистісним розвитком	Увага до ціни та зручності використання	Оперативна підтримка, зручність підписки
4.	Підвищення якості обслуговуванн я клієнтів у приватному бізнесі	Сервісні компанії, кол- центри, відділи підтримки клієнтів	Швидке реагування на емоції клієнтів. Фокус на оптимізації процесів і зменшенні конфліктів	Робота в реальному часі, інтеграція з CRM. Гарантія надійності, технічна підтримка 24/7

Третім етапом є аналіз ринкового середовища, який включає визначення факторів загроз (табл. 5.6) та факторів можливостей (табл. 5.7). Ці фактори можуть як негативно, так і позитивно впливати на впровадження проєкту на ринку.

Таблиця 5.6 – Фактори загроз

№ п/ п	Фактор	Зміст загрози	Можлива реакція компанії
1.	Висока конкуренція на ринку	Наявність сильних конкурентів, таких як Afectiva та Replika	Розробка унікальних функцій і акцент на персоналізацію продукту

№ п/ п	Фактор	Зміст загрози	Можлива реакція компанії
2.	Обмежений доступ до інвестицій	Високі витрати на розробку та маркетинг	Демонстрація рентабельності та перспективності проєкту для залучення інвесторів
3.	Стандартизація та сертифікація	Вимоги до відповідності стандартам GDPR та сертифікації у сферах	Забезпечення відповідності стандартам і залучення експертів для сертифікації
4.	Технологічні бар'єри	Складність інтеграції з іншими платформами та потреба у значних ресурсах	Використання хмарних сервісів для масштабування та спрощення інтеграції
5.	Нестабільність попиту	Невизначеність у масштабах і стабільності попиту	Проведення маркетингових досліджень і тестування продукту
6.	Високі витрати на підтримку клієнтів	Потреба у постійній технічній підтримці та оновленнях	Автоматизація підтримки та оптимізація внутрішніх процесів

Таблиця 5.7 – Фактори можливостей

№ п/ п	Фактор	Зміст можливості	Можлива реакція компанії
1.	Зростання попиту на Emotion AI	Збільшення інтересу до рішень на основі емоційного інтелекту у різних галузях	Активне просування продукту у сфері медицини, бізнесу та освіти

№ п/ п	Фактор	Зміст можливості	Можлива реакція компанії
2.	Розвиток хмарних технологій	Доступність інструментів для масштабованого обчислення і зберігання даних	Використання AWS для зменшення витрат і забезпечення продуктивності
3.	Висока потреба в персоналізації	Зростаючий попит на персоналізовані сервіси для аналізу емоцій користувачів	Інтеграція адаптивних моделей, які враховують індивідуальні потреби
4.	Розширення ринку AI-рішень	Активне впровадження штучного інтелекту в нові сфери(охорона здоров'я)	Орієнтація на нові ринкові сегменти та партнерство з ключовими гравцями
5.	Зростання інтересу до здорового самопочуття	Популярність сервісів для підтримки психоемоційного здоров'я	Позиціонування продукту як інструменту для покращення якості життя

Для глибшого розуміння ринкового середовища наступним кроком є ступеневий аналіз конкуренції, результати якого подано у таблиці 5.8.

Таблиця 5.8 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
Тип конкуренції – монополістична	На ринку домінують кілька великих гравців та новачки	Розробка унікальних функцій для виділення серед конкурентів

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
Рівень конкурентної боротьби – міжнародний	Головні конкуренти працюють на глобальному ринку	Адаптація продукту для різних ринків, участь у міжнародних виставках
За галузевою ознакою – внутрішньогалузева	Конкуренція серед рішень для аналізу емоцій на основі AI	Фокус на точності мультимодального аналізу та швидкості обробки даних
За видом товарів – товарно-видова	Продукти надають схожі функції, але різні якості	Забезпечення високої якості аналізу та доступності
За характером конкурентних переваг – нецінова	Більшість конкурентів акцентуються на технологічних інноваціях	Впровадження унікальних технологій (мультимодальний аналіз, персоналізація)
За інтенсивністю – марочна	Конкуренти активно просувають бренди	Інвестування в маркетинг і формування бренду з акцентом на унікальні переваги

Далі виконуємо глибший аналіз конкурентного середовища галузі з використанням моделі п'яти сил М. Портера, результати якого представлені у таблиці 5.9.

Таблиця 5.9 – Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
	Affectiva, Replika,	Бар'єри вступу	Можуть диктувати	Висока сила впливу на	Загроза замінників

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
	Hume, Retorio, Realeyes, Eyeris	через технології та інвестиції	ціни на технологічні компоненти	ціноутворення та функціонал	низька через унікальність мультимодального аналізу
Висновки	Інтенсивність конкуренції висока	Бар'єри входу існують, але є можливості	Диктують умови у частині цін	Клієнти вимагають якість та інновації	Замінники не створюють значного тиску на ринку

Аналіз конкуренції, представлений у таблицях 5.8 та 5.9, показує, що Ринок емоційного аналізу є перспективним для виходу, оскільки попит на інноваційні рішення стабільно зростає, а бар'єри входу помірні. Для конкурентоспроможності проєкт повинен забезпечувати точний мультимодальний аналіз, інтеграцію з платформами та відповідність стандартам конфіденційності.

На основі аналізу конкуренції, характеристик проєктної ідеї, споживчих вимог до товару та умов маркетингового середовища визначається і обґрунтовується перелік факторів конкурентоспроможності (таблиця 5.10).

Таблиця 5.10 – Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проєктів значущим)
1.	Точний мультимодальний аналіз	Дозволяє значно підвищити точність визначення емоцій, перевершуючи рішення конкурентів, які аналізують лише одну або дві модальності

№ п/ п	Фактор	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проєктів значущим)
2.	Робота в реальному часі	Ключовий фактор для бізнесів, де швидкість аналізу є критичною. У конкурентів часто спостерігаються затримки
3.	Персоналізація результатів	Забезпечує індивідуальний підхід для користувачів, що є запитом у більшості клієнтів. Конкуренти пропонують мінімальну персоналізацію
4.	Відповідність стандартам конфіденційності	Гарантує довіру користувачів і відповідає міжнародним вимогам, як-от GDPR
5.	Масштабованість	Завдяки використанню хмарних сервісів, забезпечує обробку великих обсягів даних, що є перевагою перед конкурентами

Далі відповідно до визначених факторів конкурентоспроможності проводиться аналіз сильних та слабких сторін стартап-проєкту. Його можна побачити у таблиці 5.11.

Таблиця 5.11 – Порівняльний аналіз сильних та слабких сторін проєкту

№ п/ п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з системою комп'ютерного зору для обробки медіафайлів із обличчям людини за допомогою нейронних мереж						
			-3	-2	-1	0	+1	+2	+3
1.	Мультимодальний аналіз	18						+	
2.	Робота в реальному часі	16						+	

№ п/ п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з системою комп'ютерного зору для обробки медіафайлів із обличчям людини за допомогою нейронних мереж						
			-3	-2	-1	0	+1	+2	+3
3.	Персоналізація результатів	14					+		
4.	Стандарти конфіденційності	15				+			
5.	Масштабованість	17							+

Остаточним етапом аналізу ринкових можливостей впровадження проєкту є підготовка SWOT-аналізу (матриці сильних (Strengths) і слабких (Weaknesses) сторін, загроз (Threats) і можливостей (Opportunities)) у таблиці 5.12.

Таблиця 5.12 – SWOT- аналіз стартап-проєкту

Сильні сторони:	Слабкі сторони:
Точний мультимодальний аналіз (міміка, текст). Персоналізація результатів.	Залежність від хмарних сервісів. Витрати на початковий маркетинг та просування. Обмеженість ресурсів для тестування. Висока конкуренція з великими гравцями.
Можливості:	Загрози:
Зростання попиту на проєкт у різних галузях. Розширення ринків через інтеграцію з платформами CRM. Популярність персоналізованих сервісів.	Активність конкурентів на глобальному ринку. Вимоги до сертифікації. Ризик високих витрат на підтримку та обслуговування. Нестабільність економіки через воєнні дії в Україні.

Як можна побачити з таблиці 5.11, стартап-проєкт системи комп'ютерного зору для обробки медіафайлів із обличчям людини за допомогою нейронних мереж

має значні конкурентні переваги завдяки точному мультимодальному аналізу, персоналізації результатів та відповідності стандартам конфіденційності. Основні слабкі сторони включають високу залежність від хмарних сервісів і значні початкові витрати, однак ці фактори можна мінімізувати через ефективну стратегію розвитку та інтеграцію інноваційних рішень.

На основі SWOT-аналізу формуються альтернативні стратегії ринкової поведінки (перелік заходів), спрямовані на успішне виведення стартап-проєкту на ринок. Також визначається орієнтовний оптимальний час їх реалізації з урахуванням можливого запуску конкурентних проєктів. Узагальнені альтернативи представлено у таблиці 5.13.

Таблиця 5.13 – Альтернативи ринкового впровадження стартап-проєкту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1.	Активне просування продукту через цифровий маркетинг	80%	6-12 місяців
2.	Партнерство з корпоративними платформами (CRM, HR-системи)	60%	18 місяців
3.	Забезпечення відповідності міжнародним стандартам (GDPR, ISO)	90%	9 місяців
4.	Участь у міжнародних виставках та презентаціях для інвесторів	70%	12 місяців
5.	Запуск пілотних проєктів із ключовими клієнтами (медицина, бізнес)	75%	12 місяців
6.	Оптимізація продукту для масштабування (AWS, мультимодальний аналіз)	85%	18 місяців

На основі аналізу таблиці 5.13 обраною альтернативою є активне просування продукту через цифровий маркетинг. Ця альтернатива має високу ймовірність

отримання ресурсів і короткі строки реалізації, що дозволяє швидко вивести проєкт на ринок і створити базу для подальшого розвитку.

Додатково, для довгострокової конкурентоспроможності, варто паралельно працювати над відповідністю міжнародним стандартам.

5.4 Розроблення ринкової стратегії проєкту

Розроблення ринкової стратегії є ключовим етапом запуску стартап-проєкту, що забезпечує ефективне охоплення ринку. Для цього визначаються цільові групи споживачів, їхні особливості та вимоги, що подано у таблиці 5.14.

На основі огляду цільових груп основними, на яких слід зосередити увагу, було вибрано медичні установи та сервісні компанії. Ці сегменти демонструють високий попит на технології емоційного аналізу, готовність до впровадження продукту та значний потенціал для масштабування в межах їхньої діяльності.

Таблиця 5.14 – Вибір цільових груп потенційних споживачів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачі в сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1.	Медичні установи, лікарі, психологи	Висока	Помірний	Середня	Складно
2.	HR-менеджери та корпоративні тренери	Висока	Високий	Середня	Помірно складно
3.	Індивідуальні користувачі	Середня	Високий	Низька	Відносно просто
4.	Сервісні компанії та кол-центри	Висока	Високий	Висока	Помірно складно

Оскільки обрано не одну цільову групу, за стратегію охоплення ринку обрано стратегію диференційованого маркетингу.

Для ефективного просування на обраних сегментах ринку необхідно визначити базову стратегію розвитку. У таблиці 5.15 представлено її характеристику. Проєкт орієнтується на стратегію диференційованого маркетингу, що передбачає роботу з кількома цільовими сегментами – медичними установами та сервісними компаніями. Для кожного з цих сегментів розробляються окремі програми ринкового впливу, що дозволяє максимально врахувати їхні специфічні потреби та зміцнити позиції проєкту на ринку.

Після цього йде вибір стратегії конкурентної поведінки. Результат цього вибору можна побачити у таблиці 5.16. За стратегію конкурентної поведінки було взято стратегію наслідування лідера, оскільки головним пріоритетом на першому етапі є зайняття ринку в Україні, де закордонні конкуренти майже не представлені.

Таблиця 5.15 – Визначення базової стратегії розвитку

№ п/п	Обрана альтернатива розвитку проєкту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
1.	Запуск сервісу для медичних установ із базовим функціоналом, з подальшою адаптацією до потреб сегменту	Просування сервісу через медичні асоціації та тематичні конференції	Точний мультимодальний аналіз, відповідність стандартам GDPR, інтеграція з медичними платформами	Адаптація продукту до вимог медичного сектору
2.	Випуск сервісу для сервісних компаній із	Просування сервісу через професійні	Інтеграція з CRM, оптимізація	Розробка окремих програм для

№ п/п	Обрана альтернатива розвитку проєкту	Стратегія охоплення ринку	Ключові конкурентоспроможн і позиції відповідно до обраної альтернативи	Базова стратегія розвитку
	акцентом на інтеграцію з CRM, подальшим розширенням функціоналу	бізнес-спільноти та платформи CRM	обслуговування клієнтів	обслуговування бізнесу

Таблиця 5.16 – Визначення базової стратегії конкурентної поведінки

№ п/п	Чи є проєкт «першопроходцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки
1.	Ні, ринок вже сформований, існують сильні гравці	Переважно шукатиме нових споживачів, з акцентом на незадоволені потреби сегментів	Ні, але будуть впроваджені аналогічні стандарти, такі як відповідність GDPR	Диференціація через унікальні функції (мультимодальний аналіз, персоналізація)

Стратегія позиціонування формується на основі вимог споживачів з обраних сегментів до постачальника (стартап-компанії) та продукту, а також відповідно до базової стратегії розвитку і стратегії конкурентної поведінки. Вона спрямована на створення ринкової позиції (комплексу асоціацій), за якими споживачі зможуть ідентифікувати проєкт або торговельну марку. Результати цієї стратегії відображено у таблиці 5.17.

Таблиця 5.17 – Визначення стратегії позиціонування

№ п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проєкту	Вибір асоціацій, які мають сформувати комплексну позицію власного проєкту (три ключових)
1.	Висока точність аналізу, конфіденційність даних	Стратегія диференціації	Точний мультимодальний аналіз, відповідність GDPR, інтеграція з іншими платформами	Точність Надійність. Персоналізація
2.	Зручність використання, персоналізація результатів		Простий інтерфейс, адаптація до потреб користувачів	

Підсумком виконання підрозділу є розроблена система стратегічних рішень, яка окреслює основні напрями ринкової діяльності та позиціонування стартап-компанії.

5.5 Розроблення маркетингової програми стартап-проєкту

Для успішного продажу продукту необхідно забезпечити наявність попиту на нього. Стимулювання попиту та підвищення впізнаваності бренду досягається завдяки ефективній рекламі, яка є основою маркетингової діяльності. Тому розробка ретельно продуманої маркетингової програми для виходу на ринок є ключовим етапом, що сприяє швидкій окупності інвестицій та максимізації прибутків.

Першим етапом у створенні маркетингової програми є визначення маркетингової концепції продукту, яку отримає споживач. У таблиці 5.18 наведено узагальнені результати аналізу конкурентоспроможності продукту.

Таблиця 5.18 – Визначення ключових переваг концепції потенційного товару

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1.	Точний аналіз емоцій споживачів	Підвищення точності розуміння емоцій через мультимодальний підхід	Висока точність аналізу, мультимодальність, персоналізація
2.	Забезпечення конфіденційності та відповідності стандартам	Захист даних відповідно до GDPR	Відповідність міжнародним стандартам
3.	Зручність використання продукту	Інтуїтивно зрозумілий інтерфейс	Простота інтеграції та використання

Наступним етапом є розробка тривірневої маркетингової моделі товару, яка передбачає деталізацію ідеї послуги, її функціональних характеристик та специфіки процесу надання. Структура моделі представлена у таблиці 5.19.

Таблиця 5.19 – Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Точний мультимодальний аналіз емоцій для медичних установ і сервісних компаній, покращення якості обслуговування, персоналізація під потреби клієнтів.		
II. Товар у реальному виконанні	Властивості/характеристики	М/ Нм	Вр/Тх /Тл/Е/Ор
	1. Мультимодальний аналіз емоцій (міміка, текст)	М	Е
	2. Відповідність GDPR для забезпечення конфіденційності даних	М	Вр
	3. Простота інтеграції з CRM та медичними платформами	Нм	Тх
	4. Локалізація під потреби користувачів різних ринків	М	Тл
	Якість: відповідність міжнародним стандартам, використання сучасних алгоритмів		
	Пакування: інтеграція через онлайн-доступ до платформи (віртуальний сервіс)		
	Марка: «MultimodalEmotionRecognition» – система комп’ютерного зору для обробки медіафайлів із обличчям людини за допомогою нейронних мереж		
III. Товар із підкріпленням	До продажу: доступ до тестового періоду, консультації спеціалістів		
	Після продажу: постійна технічна підтримка, оновлення сервісу, інтеграція нових функцій		
За рахунок чого потенційний товар буде захищено від копіювання: за рахунок використання ноу-хау у створенні унікальних алгоритмів мультимодального			

Рівні товару	Сутність та складові
аналізу емоцій, а також комплексного поєднання функціональних характеристик продукту (відповідність GDPR) і високоякісної сервісної підтримки	

Наступним кроком є визначення цінових меж, якими необхідно керуватись при встановленні ціни на потенційний товар (остаточне визначення ціни відбувається під час фінансово-економічного аналізу проєкту), яке передбачає аналіз ціни на товари-аналоги або товари субституту, а також аналіз рівня доходів цільової групи споживачів (табл. 5.20). Аналіз проводиться експертним методом. Оскільки більшість сервісів працюють за моделлю підписки, ціни вказані за один платний період, який дорівнює одному місяцю. Рівень доходів споживачів визначений як середньостатистичне значення доходу з урахуванням усіх витрат, на основі даних про компанії у сфері медицини та сервісних послуг.

Таблиця 5.20 – Визначення меж встановлення ціни

№ п/п	Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
1.	1000 грн	1500 грн	150000 грн	1200–1400 грн

Наступним етапом є визначення оптимальної системи збуту, яка передбачає прийняття рішень щодо організації збуту власними силами чи залученням сторонніх посередників, визначення оптимальної глибини каналу збуту, вибору та обґрунтування типу посередників (табл. 5.21).

Таблиця 5.21 – Формування системи збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
1.	Медичні установи здійснюють закупівлі	Консультації, технічна підтримка	Прямий канал	Власна система збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
	через тендери та контракти			
2.	Сервісні компанії орієнтуються на рекомендації або онлайн-платформи	Швидке налаштування, навчання	Однорівневи й канал	Залучена система через CRM- платформи

Завершальним етапом маркетингової програми є створення концепції маркетингових комунікацій, яка базується на попередньо визначеному позиціонуванні та специфіці поведінки цільових клієнтів. Узагальнені результати відображені у таблиці 5.22.

Підсумком цього розділу є сформована ринкова (маркетингова) програма, яка охоплює концепції товару, збуту, просування, а також попередній аналіз цінових можливостей. Вона базується на потребах та цінностях потенційних споживачів, конкурентних перевагах проєкту, а також враховує стан та тенденції ринкового середовища, в межах якого планується впровадження стартапу, і обрану стратегію ринкової поведінки.

Таблиця 5.22 – Концепція маркетингових комунікацій

Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонуванн я	Завдання рекламного повідомлення	Концепція рекламного звернення
Потенційні клієнти обирають	Професійні конференції, асоціації,	Надійність. Точність. Персоналізація	Підкреслити відповідність GDPR,	«Інноваційне рішення для аналізу емоцій,

Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
рішення через професійні спільноти, тендери, онлайн-платформи та рекомендації	спеціалізовані сайти, соціальні мережі, CRM-платформи		точність аналізу, простоту та оперативність використання	адаптоване до ваших потреб»

Висновки до розділу 5

У цьому розділі було здійснено розробку стартап-проєкту, яка охоплює опис ідеї, технологічний аудит, аналіз ринкових можливостей, визначення стратегії ринкової поведінки та формування маркетингового плану. Ідея проєкту базується на впровадженні мультимодальної системи аналізу емоцій з високим рівнем точності, персоналізації та відповідності стандартам конфіденційності. Результати аналізу ринку вказують на значний попит даних систем в таких сферах як медичні установи та сервісні компанії. Аналіз конкурентного середовища підкреслив перспективність даного проєкту. Визначено ключові переваги продукту, цінових меж, систем збуту та концепції маркетингових комунікацій. Обрано стратегію просування та позиціонування для ефективного охоплення цільових сегментів

Таким чином, розробка стартап-проєкту «системи комп'ютерного зору для обробки медіафайлів із обличчям людини за допомогою нейронних мереж» закладає міцний фундамент для впровадження інноваційних рішень у сфері охорони здоров'я, професійної діяльності, клієнтського сервісу та індивідуального моніторингу емоцій у повсякденному житті.

ВИСНОВКИ

У результаті цього проєкту було досліджено розроблену систему комп'ютерного зору для мультимодального аналізу емоційного стану людини за допомогою нейронних мереж, тобто система поєднує в собі аналіз обличчя та людської розмови. Це дозволяє її застосовувати в сферах медицини, бізнесу, освіти, а також сфери послуг.

Перший розділ містить детальний аналіз сучасних рішень у сфері емоційного аналізу. Розглянуто існуючі технології, їх можливості, сильні та слабкі сторони. Це дозволило виявити обмеження і переваги, які потребують застосування в інноваційному рішенні. Крім того, було проведено порівняльний аналіз систем, спрямованих на аналіз емоцій, і виділено тенденції розвитку цієї галузі. Було вирішено сфокусуватися на мультимодальному підході та підвищенні конфіденційності даних. Тому в розділі сформулювалась основна проблема дослідження, яка полягає створенні інноваційної системи, здатної виконувати аналіз емоцій з урахуванням вимог сучасного ринку, таких як конфіденційність даних та відповідність Загальному регламенту про захист даних (ЗРЗД).

Другий розділ приділив увагу проектуванню системи. Було продумано загальні архітектурні рішення та описано функціональні складові проєкту. Зокрема на рішення щодо проєкту впливали функції мультимодального аналізу вхідних даних та їх конфіденційності. А також запланована майбутня робота в режимі реального часу, інтеграцію з медичними та корпоративними платформами вплинули на закладання зручної основи для масштабування. Було описано логіку роботи системи, яка передбачає персоналізований підхід до обробки даних.

Третій розділ зосередився на технологічній базі. Було обрано хмарне середовище розгортання системи AWS та детально описано застосування сервісів та засобів побудови надійних, з можливістю масштабування. Сховища даних різними функціями EFS, S3 та DynamoDB були використанні для цілей швидкісного доступу до даних, зберігання великих об'ємів медіаінформації, а також безпечного зберігання даних користувачі та аналізу їх емоційного стану

протягом часу. Lambda як безсерверне рішення допоможе зекономити кошти та змогу розширення системи, при рості навантаження. Також у цьому розділі було розглянуто та відібрано архітектури нейронних мереж (ResNet, XLM-RoBERTa) та датасетами (FER+, GoEmotions) для аналізу вхідних медіаданих.

У четвертому розділі було описано, на основі визначених вимог та архітектури у третьому розділі, розроблення прототипу системи та імплементації функціональності. Поміж функцій точного аналізу емоцій, безпечного зберігання даних, в деталях розібрано типи запитів, типи даних, що використовуються в різних частинах застосунку. Було розібрано використання ResNet та XLM-RoBERTa для аналізу міміки обличчя та сентиментальності тексту транскрибованого з відео для надання системі перевагу поміж конкурентів.

П'ятий розділ показав розроблену маркетингову стратегію. Ця частина записки зосередилась на визначенні цільових аудиторій, а також аналізу цінової політики на основі експертного методу. Було обрано стратегії диференційованого маркетингу, позиціонування проєкту та розроблено систему продажів із застосуванням концепції маркетингових комунікацій.

У результаті детально описаної системи та її розробки було прокладено шлях до потенційного проведення точного аналізу емоцій на постійній основі. Система, підтримуючи персоналізований підхід маючи новизну у використанні мультимодального підходу (обробки даних з різних джерел), цілісно аналізує емоційний стан людини протягом заданого часу. Поєднане використання хмарного рішення AWS та сучасних архітектур нейронних мереж, система здатна видавати високу продуктивність при роботі з великими обсягами даних. А тому розроблювана система відповідає поточним потребам ринку та створює можливості для розширення сфери застосування.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Mental health: [Веб-сайт]. URL <https://www.who.int/data/gho/data/themes/theme-details/GHO/mental-health> (дата звернення: 29.09.2024).
2. COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide: [Веб-сайт]. URL <https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide> (дата звернення: 30.09.2024).
3. Cost Remains Significant Barrier to Therapy Access, 2022 Verywell Mind Survey Finds: [Веб-сайт]. URL <https://www.verywellmind.com/cost-of-therapy-survey-5271327> (дата звернення 2.10.2024).
4. Г. Торнікрофт та ін. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. The Lancet, 2016, т. 387, № 10023. 15–20 с.
5. Investing in treatment for depression and anxiety leads to fourfold return: [Веб-сайт]. URL https://www.who.int/news/item/13-04-2016-investing-in-treatment-for-depression-and-anxiety-leads-to-fourfold-return?utm_source=chatgpt.com (дата звернення 2.10.2024).
6. Deep face recognition: A survey: [Веб-сайт]. URL http://www.whdeng.cn/Wang_Deng_Survey.pdf (дата звернення 4.10.2024).
7. Д. Големан. How the Brain Grew. Emotional Intelligence. Нью-Йорк: Bantam Books, 2006. 30 с.
8. Д. Големан. Emotional Hijacking. Emotional Intelligence. Нью-Йорк: Bantam Books, 2006. 34–35 с.
9. Д. Големан. The Emotional Sentinel. Emotional Intelligence. Нью-Йорк: Bantam Books, 2006. 38–40 с.
10. Dr. Joseph Ledoux: Using Rats to Trace Routes of Fear: [Веб-сайт]. URL <https://www.columbia.edu/~lep1/rry/w3410/LeDoux/NYT.Nov.96.html> (дата звернення: 10.10.2024).

11. Daylio App Review 2024: Pros & Cons, Cost & Who It's Right For: [Веб-сайт]. URL <https://www.choosingtherapy.com/daylio-app-review/> (дата звернення: 10.10.2024).
12. Moodnotes is the new mental health app from the folks that brought you Monument Valley: [Веб-сайт]. URL <https://www.androidauthority.com/moodnotes-developer-interview-777821/> (дата звернення: 10.10.2024).
13. Affectiva: [Веб-сайт]. URL <https://www.affectiva.com/> (дата звернення: 10.10.2024).
14. Introducing Ask Replika: [Веб-сайт]. URL <https://blog.replika.com/posts/ask-replika> (дата звернення: 10.10.2024).
15. Amazon Maintains Cloud Lead as Microsoft Edges Closer: [Веб-сайт]. URL <https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/> (дата звернення: 10.10.2024).
16. Best deep CNN architectures and their principles: from AlexNet to EfficientNet: [Веб-сайт]. URL <https://theaisummer.com/cnn-architectures/> (дата звернення: 11.10.2024).
17. FERPlus: [Веб-сайт]. URL <https://github.com/Microsoft/FERPlus?tab=readme-ov-file> (дата звернення: 12.10.2024).
18. Story of AWS Storage: [Веб-сайт]. URL <https://medium.com/@sumbul.first/story-of-aws-storage-3d1934c1a336> (дата звернення: 13.10.2024).
19. What AWS Cloud Storage Solution Can Best Agree with Your Business: [Веб-сайт]. URL <https://www.infopulse.com/blog/best-aws-cloud-storage-solution> (дата звернення: 15.10.2024).
20. Deploying large packages on AWS Lambda using EFS: [Веб-сайт]. URL <https://towardsdatascience.com/deploying-large-packages-on-aws-lambda-using-efs-3a707f83d918> (дата звернення: 16.10.2024).
21. Deploy multiple machine learning models for inference on AWS Lambda and Amazon EFS: [Веб-сайт]. URL <https://aws.amazon.com/blogs/machine->

learning/deploy-multiple-machine-learning-models-for-inference-on-aws-lambda-and-amazon-efs/ (дата звернення: 20.10.2024).

22. How to Integrate API Gateway with Cognito to Serve Authenticate Users: [Веб-сайт]. URL <https://enlear.academy/how-to-integrate-api-gateway-with-cognito-to-serve-authenticate-users-ed46c2bedb28> (дата звернення: 16.10.2024).

23. The state of serverless: [Веб-сайт]. URL <https://www.datadoghq.com/state-of-serverless/#4> (дата звернення: 18.10.2024).