

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ

Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті

Автоматика және ақпараттық технологиялар институты

«Программалық инженерия» кафедрасы

Абылкасым Ерарыс Айбекулы

Дипломдық жобаға

ТҮСІНДІРМЕ ЖАЗБА

6B06102 – Computer Science білім беру бағдарламасы

Алматы 2025

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ

Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті

Автоматика және ақпараттық технологиялар институты

«Программалық инженерия» кафедрасы

ҚОРҒАУҒА ЖІБЕРІЛДІ

ПИ кафедрасының меңгерушісі

техн. ғыл. канд., қауымд.проф.

_____ Ф.Н. Абдолдина

«_____» _____ 2025 ж.

Дипломдық жобаға
ТҮСІНДІРМЕ ЖАЗБА

Тақырыбы: «Дыбыс пен бейне деректерін пайдалана отырып, мульти-модальды нейрондық желілер негізінде эмоцияларды тану жүйесін әзірлеу» тақырыбына»

6B06102 – Computer Science білім беру бағдарламасы

Орындады

Рецензент

PhD, _____ доктор

_____ Нурлыбаева К.К

"__" "__" _____ 2025 ж.

Аты _____ жөні

Ғылыми _____ жетекшісі

PhD, _____ қауымд.профессор

_____ Қасенхан А.М

"__" "__" _____ 2025 ж.

Алматы 2025

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ
ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ

Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті
Автоматика және ақпараттық технологиялар институты

«Программалық инженерия» кафедрасы

БЕКІТЕМІН

ПИ кафедрасының меңгерушісі
техн. ғыл. канд., қауымд.проф.
_____ Ф.Н. Абдолдина
« ____ » _____ 2025 ж.

Дипломдық жобаны орындауға
ТАПСЫРМА

Білім алушыға: *Аты жөні*

Тақырыбы: _____ 29.01.2025

*Академиялық мәселелер жөніндегі проректор бұйрығының № 26-П/Ө «29» қаңтар
2025 ж. шешімімен бекітілген.*

Орындалған жобаның өткізу мерзімі: « ____ » _____ 2025 ж.

Дипломдық жобаның бастапқы мәліметтері:

- А) Пәндік саланы талдау, ұқсас жобаларды талдау;*
- Б) Техникалық тапсырманы әзірлеу;*
- В) БҚ архитектурасын әзірлеу;*
- Г) Түсінікті және интерактивті дизайнды жобалау және әзірлеу; Д)
Бағдарламалық жасақтаманы әзірлеу және тестілеу;*

Дипломдық жобада әзірленуге жататын сұрақтардың тізбесі: (міндетті сызбаларды дәл көрсете отырып): *презентацияның _____ слайдтары ұсынылды. Ұсынылатын негізгі әдебиеттер: _____ пайдаланылған әдебиеттер тізімінен.*

Дипломдық жобаны орындау

КЕСТЕСІ

Бөлімдердің атаулары, зерттелген мәселелердің тізімі	Ғылыми жетекшіге және кеңесшілерге ұсыну мерзімі	Ескерту
1. Пәндік саланы зерттеу және жүйеге қойылатын талаптарды анықтау	03.02.2025	Орындалды
2. Модельдерді оқыту және олардың өзара әрекеттесу архитектурасын құрастыру	01.04.2025	Орындалды
3. Пайдаланушылардың модельмен өзара әрекеттесу интерфейсін әзірлеу	06.05.2025	Орындалды
4. Веб-қосымшаны аяқтап, тестілеуден өткізу	27.05.2025	Орындалды

Дипломдық жұмыс бөлімдерінің кеңесшілерінің аяқталған жұмысқа қойған қолтаңбалары

Бөлімдер атауы	Кеңес берушілер (аты-жөні, тегі, ғылыми дәрежесі, атағы)	Қолтаңба қойылған мерзімі	Қолы
Бағдарламалық бөлім	Баймбетов Д.А. техн. ғыл. канд., проф.		
Нормалық бақылаушы	Бигельдина Г.А. тех.ғыл.магистрі		

Ғылыми жетекші _____ Алибиева Ж.М.

Тапсырманы орындауға қабылдап алған студент _____ Асан Усенов Күні

«__» _____ 2025 ж.

АНДАТПА

Бұл дипломдық жоба дыбыс пен бейне деректер негізінде адамның эмоцияларын тануға арналған мультимодальды интеллектуалды жүйені әзірлеуге бағытталған. Қазіргі таңда эмоцияны дәл әрі сенімді тану – жасанды интеллект пен ақпараттық технологиялар саласындағы өзекті мәселелердің бірі. Адамның эмоциялық күйін тану көптеген салаларда, атап айтқанда қашықтан оқыту, психологиялық диагностика, клиенттермен байланыс жүйелері, қауіпсіздік және әлеуметтік робототехникада кеңінен қолданылуы мүмкін. Жобаның мақсаты – адамның бет-әлпеті, мимикасы және дауыс ерекшеліктері сияқты мультимодальды белгілерге негізделген эмоцияны автоматты түрде танитын жүйе жасау. Бұл мақсатқа жету үшін автор дыбыстық және бейне сигналдарды алдын ала өңдеудің заманауи әдістерін қолданды. Әсіресе, спектрограммалар мен беттегі негізгі нүктелерден алынған деректер жүйеге енгізілді. Жүйе бірнеше нейрондық архитектураларды пайдаланып салыстырмалы талдау жасады: CNN, LSTM, Transformer, сондай-ақ оларды біріктіретін мультимодальды біріктіру әдістері. Модельдер дәлдік (accuracy), F1-метрика, қате матрицасы (confusion matrix) сияқты көрсеткіштер бойынша бағаланды. Соңында ең тиімді нәтиже көрсеткен модель таңдалып, нақты уақыттағы веб-қосымша ретінде жүзеге асырылды. Қосымша Python, OpenCV, TensorFlow, Keras, MediaPipe, librosa, sklearn сияқты кітапханалармен жасалды. Интерфейс қарапайым әрі қолданушыға ыңғайлы етіп жасалған. Бұл жүйе нақты уақытта адамның эмоциясын анықтап, оның психоэмоциялық жағдайына бейімделуге мүмкіндік береді.

АННОТАЦИЯ

Данный дипломный проект посвящён разработке мультимодальной интеллектуальной системы для распознавания эмоций на основе аудио- и видеоданных. Эффективное определение эмоционального состояния человека является актуальной задачей в области искусственного интеллекта, особенно в таких сферах, как дистанционное обучение, психологическая диагностика, клиентская поддержка, системы безопасности и социальная робототехника.

Цель проекта — разработать систему, способную автоматически распознавать эмоции человека на основе мультимодальных признаков, включая выражение лица, мимику и особенности голоса. Для этого в работе использованы современные методы обработки аудиовизуальной информации: аудиосигналы преобразовывались в спектрограммы, а с видеопотока извлекались ключевые точки лица.

Для анализа и сравнения были применены различные нейронные сети: **CNN**, **LSTM**, **Transformer**, а также мультимодальные схемы объединения признаков. Модели оценивались по метрикам: **точность (accuracy)**, **F1-score**, **матрица ошибок (confusion matrix)**. По результатам тестов была выбрана наиболее эффективная архитектура, реализованная в виде веб-приложения с возможностью обработки данных в реальном времени.

Приложение разработано с использованием современных библиотек: Python, OpenCV, TensorFlow, Keras, MediaPipe, librosa, sklearn. Интерфейс получился интуитивно понятным и доступным. Система может быть использована в различных прикладных областях и служить основой для дальнейших исследований и внедрения.

ABSTRACT

This thesis project focuses on the development of a multimodal intelligent system for recognizing human emotions based on audio and video data. Accurate emotion recognition is a critical challenge in artificial intelligence, with applications in areas such as remote learning, psychological diagnostics, customer service, security, and social robotics.

The main goal is to create an automatic emotion recognition system that leverages multimodal features including facial expressions, mimicry, and vocal tone. The project involves the preprocessing of audio and visual data using modern methods: audio signals are converted into spectrograms, while facial landmarks are extracted from video streams.

Several neural architectures were evaluated and compared, including **CNN**, **LSTM**, **Transformer**, and multimodal fusion strategies. These models were assessed based on key metrics such as **accuracy**, **F1-score**, and **confusion matrix**. The best-performing model was selected and implemented as a real-time web application.

The application utilizes technologies such as Python, OpenCV, TensorFlow, Keras, MediaPipe, librosa, and sklearn. It features a user-friendly interface capable of processing real-time input from a camera and microphone, enabling the system to detect emotional states with high reliability. This solution lays a strong foundation for future integration in real-world scenarios and further development in emotion-aware systems.

Мазмұны	
КІРІСПЕ	10
Тақырыптың өзектілігі	10
Зерттеу мақсаты мен міндеттері	10
Зерттеу әдістері	11
Жұмыс құрылымы	11
1-тарау. Теориялық бөлім	12
1.1 Жасанды интеллектке кіріспе	12
1.2 Эмоцияны тануға арналған ЖИ түрлері	13
1.3 Эмоцияны тану әдістеріне шолу	14
1.4 Аудио және бейне деректерді өңдеу	17
2-тарау — Жобалық-тәжірибелік бөлім	23
2.3 Деректерді өңдеу (жалғасы)	27
2.4 Эмоцияларды тану модулі	28
2.4.2 Эмоцияларды танудың көпмодальды тәсілдері	30
2.4.7 Эмоцияларды тану модулінің архитектурасы	31
2.4.8 Қолдану сценарийлері	32
2.4.9 Интеграция және жүйе архитектурасы	32
2.4.10 Модульді жетілдіру бағыттары	32
2.5 Жауап генерациялау модулі	32
3-тарау. Эмоцияларды тану технологияларының әлеуетті қолдану салалары	35
3.1 Денсаулық сақтау саласы	35
3.2 Білім беру саласы	35
3.3 Автокөлік және көлік қауіпсіздігі	35
3.4 Бизнес және маркетинг	35
3.5 Әлеуметтік қызметтер және инклюзивті технологиялар	35
3.6 Құқық қорғау және қауіпсіздік	36
3.7 Жұмыс орнында әл-ауқат және өнімділік	36

3.8 Жарнама және тұтынушы тәжірибесі	36
3.9 Жасанды интеллект және эмоциялық интеллект.....	36
3.10 Болашақтағы даму бағыттары	36
Қорытынды.....	37
ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР ТІЗІМІ.....	39

КІРІСПЕ

Тақырыптың өзектілігі

Жасанды интеллект технологияларының қарқынды дамуы қазіргі заманғы ғылым мен техника салаларында үлкен серпін алып келеді. Бұл технологиялар әртүрлі салаларда — медицинадан бастап, білім беру мен ойын-сауыққа дейін — кеңінен қолданыс табуда. Атап айтқанда, адам мен машина арасындағы өзара әрекеттесу саласында жасанды интеллект жүйелерінің маңызы айрықша өсті. Әсіресе адамның эмоцияларын дәл тануға және оған тиісті жауап қайтаруға қабілетті жүйелерді құру бағытында жаңа зерттеулер мен инновациялар белсенді жүргізілуде. Видео және аудио сигналдар негізінде эмоцияны тану — заманауи технологиялардың бірі ретінде виртуалды агенттермен және роботтармен өзара әрекеттесуді анағұрлым табиғи әрі жекелендірілген деңгейге көтеруге мүмкіндік береді. Мұндай жүйелер пайдаланушылардың эмоционалдық күйін түсініп, оған сай бейімделу арқылы коммуникацияны жеңілдетеді және пайдалану ыңғайлылығын арттырады. Нәтижесінде, интеллектуалды жүйелердің дамуымен бірге адамдардың ақпаратпен өзара қарым-қатынас жасау тәсілдері де түбегейлі өзгеріп, инновациялық шешімдердің ауқымы кеңейеді.

Зерттеу мақсаты мен міндеттері

Осы ғылыми-зерттеу жұмысының негізгі мақсаты — жасанды интеллекттің заманауи технологияларын пайдалана отырып, эмоцияны нақты және тиімді түрде тану, сондай-ақ пайдаланушының эмоциялық жай-күйіне бейімделген интерактивті жүйені әзірлеу болып табылады. Жүйе нақты уақытта жұмыс істеп, аудио және бейне ақпараттарды бірлесіп талдай отырып, анықталған эмоцияларға сәйкес келетін жауаптарды автоматты түрде генерациялай алуы керек. Бұл өз кезегінде адам-машина арасындағы қарым-қатынасты әлдеқайда табиғи және жағымды етеді.

Мақсатқа қол жеткізу үшін келесі негізгі міндеттер белгіленді:

1. эмоцияны тану саласында қазіргі уақытта қолданылып жүрген әдістер мен технологияларды жан-жақты зерттеу және салыстырып талдау;
2. аудио және бейне ақпараттарды жинау, оларды алдын ала өңдеу, эмоцияны тану және жауап генерациясын жүзеге асыратын жүйенің техникалық архитектурасын жобалау;
3. эмоцияларды талдау және оларға сай жауаптар генерациялау үшін тиімді нейрондық желілерді пайдаланып, жүйенің серверлік бөлігін бағдарламалық түрде жүзеге асыру;
4. жүйенің жұмыс өнімділігі мен сапасын бағалау, сондай-ақ жетілдіру және оңтайландыру жолдарын ұсыну.

Зерттеу әдістері

Бұл жұмыста төмендегі ғылыми-тәжірибелік әдістер кеңінен қолданылды:

- эмоцияны тану саласындағы заманауи технологиялар мен әдістерге арналған кең ауқымды әдеби шолу және салыстырмалы талдау жүргізу;
- аудио және бейне деректерді жинау, өңдеу және оларды жүйеге енгізуге арналған бағдарламалық жасақтаманы әзірлеу және интеграциялау;
- эмоцияны анықтау мақсатында жасанды нейрондық желілерді, сондай-ақ алдын ала үйретілген терең оқыту модельдерін пайдалану;
- пайдаланушымен ыңғайлы және түсінікті өзара әрекеттесу үшін интуитивті интерфейс құру және оны жүзеге асыру.

Жұмыс құрылымы

Бұл дипломдық жұмыс құрамында бес негізгі бөлім бар:

1. теориялық бөлім — эмоцияны тану технологиялары мен әдістерінің жан-жақты шолуы мен талдауы. Бұл бөлімде осы саланың негізгі түсініктері мен қолданылатын алгоритмдері қарастырылады;
2. жүйені әзірлеу — жобаланған жүйенің архитектурасы, оның әрбір компонентінің сипаттамасы, функционалдығы және өзара байланысы баяндалады;
3. практикалық жүзеге асыру — жүйені нақты программалық түрде әзірлеу процесі, қолданылған құралдар мен технологиялар, кодтау жұмыстары сипатталады;
4. тестілеу — жүйенің жұмысын сынау әдістемесі, тестілеу нәтижелері мен олардың талдауы беріледі. Бұл бөлім жүйенің тиімділігін анықтауға бағытталған;
5. жүйені жетілдіру — зерттеу нәтижелеріне сүйене отырып, жүйені әрі қарай дамытуға арналған ұсыныстар мен перспективалар қарастырылады.

1-тарау. Теориялық бөлім

1.1 Жасанды интеллектке кіріспе

Жасанды интеллект (ЖИ) — адамның ақыл-ойын талап ететін тапсырмаларды орындай алатын жүйелерді жасауға бағытталған компьютерлік ғылым саласы болып табылады. Бұл сала компьютерлерге адам миының қызметтерін, соның ішінде танымдық қабілеттерін, логикалық ойлауды, шешім қабылдауды, тіл түсінуді, көрнекі және дыбыстық ақпаратты өңдеуді үйретуге бағытталған. ЖИ жүйелері көптеген салада қолданылып, адам өмірін жеңілдетуге, өндірістік процестерді автоматтандыруға, жаңа технологияларды дамытуға мүмкіндік береді.

Жасанды интеллекттің негізгі міндеттері мыналар:

- қабылдау (Perception): қоршаған ортаны сенсорлар арқылы қабылдап, оны талдау. Мысалы, объектілерді көру, дыбысты тану;
- ойлау (Reasoning): жиналған ақпарат негізінде қорытынды жасау және шешім қабылдау;
- үйрену (Learning): деректерден тәжірибе алу және нәтижелерді жақсарту.
- тілмен өзара әрекеттесу (Natural Language Processing): адам тілімен қарым-қатынас жасау;
- жоспарлау және әрекет (Planning and Acting): мақсатқа жету үшін әрекеттерді жоспарлау және орындау.

Қазіргі таңда жасанды интеллекттің көптеген бағыттары дамып келеді. Оның ішінде машиналық оқыту, нейрондық желілер, табиғи тілдерді өңдеу, компьютерлік көру және робототехника ерекше орын алады. Бұл бағыттардың әрқайсысы адам өмірінің түрлі салаларында қолданылады: медицинада диагноз қою, қаржы секторында инвестициялық шешімдер қабылдау, өндірісте сапаны бақылау, көлік саласында автономды жүргізу және тағы басқа.

Эмоцияны тану — жасанды интеллектінің соңғы даму бағыттарының бірі. Бұл салада жасалған жүйелер адамның тек ақыл-ойын ғана емес, оның эмоционалдық күйін де түсініп, соған сәйкес реакция көрсете алады. Мұндай жүйелер адамның көңіл-күйін, уайымын, қуанышын анықтап, байланыс процесін жақсартады. Мысалы, білім беру жүйелерінде оқушының эмоциясын бақылап, оқу материалының тиімділігін арттыруға болады, ал клиенттерге қызмет көрсету саласында адамның көңіл-күйін ескеріп, жеке ұсыныстар жасауға мүмкіндік

туады.

Жасанды интеллекттің әлеуеті үлкен және ол адамзаттың көптеген проблемаларын шешуге көмектеседі. Алайда, бұл технологияның этикалық мәселелері, деректердің құпиялылығы және технологияны дұрыс пайдалану маңызды сұрақтар болып қала береді.

1.2 Эмоцияны тануға арналған ЖИ түрлері

Эмоцияны тану жасанды интеллект жүйелерінің күрделі және көпқырлы бағыты. Эмоцияларды тану адамның мінез-құлқын, көңіл-күйін және сезімін дұрыс түсіну үшін өте маңызды. Эмоцияны тану технологиялары адамның қарым-қатынасын жетілдіруге, қызмет көрсету сапасын арттыруға, медицинада психологиялық диагностика жүргізуге және басқа да көптеген салаларда қолданылады.

Эмоцияны тану үшін әртүрлі ақпарат көздері пайдаланылады, олардың негізгі үш түрі бар:

1. Визуалды (бет арқылы) эмоцияны тану

Бұл әдіс адамның бет-әлпетіндегі әртүрлі мимикалық өзгерістерге негізделген. Мысалы, күлімсіреу, қабақ түю, таңғалушылық сияқты беттегі қозғалыстар эмоцияны анықтауға көмектеседі. Бұл процесс компьютерлік көру (computer vision) әдістерін және нейрондық желілерді қолдануды талап етеді.

Визуалды эмоцияны тану жүйелері көбінесе мынадай қадамдардан тұрады:

- бетті анықтау және оның негізгі нүктелерін (нүктелік белгілерін) белгілеу;
- бұл нүктелердің қозғалысы мен пішімін талдау;
- Алынған деректер негізінде эмоцияны жіктеу.

Бұл әдісте OpenCV, dlib сияқты ашық кітапханалар кеңінен қолданылады. Сонымен қатар, Convolutional Neural Networks (CNN) — терең үйрену әдістерінің бірі — беттегі эмоцияны жоғары дәлдікпен тануға мүмкіндік береді.

2. Акустикалық (дауыс арқылы) эмоцияны тану

Бұл әдіс адамның дауыс интонациясын, тембрін, дыбыс күшін және сөйлеу ырғағын талдауға негізделеді. Әр түрлі эмоциялар дауыс ерекшеліктеріне ерекше әсер етеді: мысалы, ашу кезінде дауыс қатты және жоғары, ал уайым кезінде әлсіз және төмен болады.

Дауыстық эмоцияларды тану процесі мыналарды қамтиды:

- аудио сигналды жазу және алдын ала өңдеу;
- спектрограмма немесе MFCC (Mel-frequency cepstral coefficients) секілді аудио белгілерін алу;
- алынған белгілерді машиналық оқыту модельдеріне беру;
- модельдің эмоцияны анықтауы.

Бұл бағытта librosa, pyAudioAnalysis сияқты аудио өңдеу кітапханалары

пайдаланылады. Сонымен қатар, алдын ала үйретілген нейрондық желілер мен трансформерлер қолданылады.

3. Мәтіндік эмоцияны тану

Табиғи тілдерді өңдеу (NLP) әдістері мәтіндегі эмоцияны анықтауға мүмкіндік береді. Бұл әдіс әсіресе әлеуметтік желілердегі посттар, чаттар, пікірлердегі сезімдерді талдау үшін пайдалы.

Мәтін арқылы эмоцияны тану келесі қадамдардан тұрады:

- мәтінді алдын ала өңдеу (стоп-сөздерді алып тастау, лемматизация);
- эмоцияны білдіретін сөздерді, фразаларды анықтау;
- контекстті ескеріп, жалпы эмоциялық реңкті бағалау;
- модельді қолдану (мысалы, BERT, GPT тәрізді трансформерлер).

Бұл әдіс тұтынушылардың пікірін талдау, психологиялық диагностика, виртуалды көмекшілерде қолдану үшін маңызды.

1.3 Эмоцияны тану әдістеріне шолу

Эмоцияны тану — жасанды интеллект саласындағы өте маңызды және қиын міндет. Бұл міндетті шешу үшін әртүрлі әдістер мен алгоритмдер қолданылады, олар негізінен визуалды, акустикалық және мәтіндік ақпаратты өңдеуге бағытталған. Әр әдістің өзіндік ерекшеліктері, артықшылықтары мен шектеулері бар.

1.3.1 Бет арқылы эмоцияны тану

Бет арқылы эмоцияны тану — адамның бет-әлпетіндегі өзгерістерге негізделген әдіс. Бұл әдістің тиімділігі жоғары, себебі адамның сезімдері көбінесе бет мимикасында көрініс табады. Мимикадағы әрбір микроқозғалыс эмоцияны дәл анықтауға көмектеседі.

Бет арқылы эмоцияны танудың негізгі этаптары:

- **бетті анықтау (Face Detection):** Алдымен жүйе кадрдан немесе суреттен адамның бетін тауып алады. Бұл үшін Viola-Jones алгоритмі, HOG (Histogram of Oriented Gradients), SSD (Single Shot Detector) немесе MTCNN (Multi-task Cascaded Convolutional Networks) сияқты әдістер қолданылады;
- **негізгі белгілерді (keypoints) белгілеу:** Беттің маңызды нүктелері — көздің бұрыштары, мұрын, еріннің шеттері, қастың формасы және т.б. анықталады. Мұнда dlib кітапханасындағы 68 нүктелі бет белгілерін анықтау моделі кеңінен танымал;
- **бет мимикасын талдау:** Бет нүктелерінің қозғалысы мен олардың арасындағы қашықтық, бұрыштар өзгерісі бақыланады. Бұл бет бұлшықеттерінің қозғалысына сәйкес келеді;

- **эмоцияны жіктеу (Classification):** Алынған мәліметтер терең нейрондық желілерге, мысалы, CNN архитектураларына беріліп, эмоция түрлері анықталады.

Терең үйрену әдістері

Терең нейрондық желілер (Deep Neural Networks) бет арқылы эмоцияны тануда революция жасады. Олар үлкен деректер жиынтықтарында үйрену арқылы бет суреттеріндегі кішкентай белгілерді де байқап, жоғары дәлдікпен эмоцияларды анықтай алады. Ең жиі қолданылатын архитектуралар:

- **convolutional Neural Networks (CNN):** Беттің ерекшеліктерін шығару үшін қабаттарда конволюция операциялары қолданылады. Мысалы, VGGNet, ResNet, MobileNet;
- **recurrent Neural Networks (RNN) және LSTM:** Видео ағынындағы бет өзгерістерін уақыт бойынша бақылау үшін қолданылады;
- **transformer негізіндегі модельдер:** Соңғы жылдары Transformers бет мимикасындағы нюанстарды түсіну үшін де қолданыла бастады;

Алдын ала үйретілген модельдер

- **FER (Facial Expression Recognition):** Бұл — эмоцияны тануға арналған кеңінен таралған датасеттер мен модельдер. Мысалы, FER2013 — миллиондаған бет суреттері мен эмоциялар бар деректер жиынтығы;
- **OpenCV кітапханасы:** Көптеген бет анықтау және өңдеу алгоритмдерін қамтиды, соның ішінде Haar Cascades, DNN модульдері;
- **dlib:** Бет нүктелерін дәл анықтауға арналған алгоритмдер жиынтығы.

Проблемалар мен шектеулер

- **жарықтандыру:** Нашар жарық немесе жарықтың өзгеруі нәтижеге әсер етеді;
- **беттің бұрылысы:** Егер адам бет-әлпетін толық көрсетпесе, анықтау қиындайды;
- **мимикадағы ерекшеліктер:** Әр адамның эмоцияны көрсету стилі әртүрлі;
- **мәдени айырмашылықтар:** Эмоцияның көрінісі мәдениетке байланысты өзгеруі мүмкін.

1.3.2 Дауыстық эмоцияларды тану

Дауыстық эмоцияны тану — адамның дауысындағы эмоцияны анықтауға бағытталған әдіс. Бұл әдісте адамның сөйлеу мәнері, дауыс ырғағы, тембрі, дауыс күшін талдау арқылы эмоцияны анықтайды. Бұл тәсіл адамдардың эмоционалды күйлерін дәлірек сезінуге мүмкіндік береді, себебі дауыс эмоцияның күшті және нақты көрінісі болуы мүмкін.

Дауыстық эмоцияны танудың негізгі кезеңдері:

- **дыбысты жазу:** сигналды жоғары сапада жазып алу;
- **дыбыстық сигналды алдын ала өңдеу:** шуды азайту, нормализациялау;
- **белгілерді алу (Feature Extraction):** бұл — дыбыстың маңызды параметрлерін алу процесі. Ең жиі қолданылатын белгілер:
 - **MFCC (Mel-frequency cepstral coefficients):** Адамның есту жүйесінің сезімталдығына жақын спектрлік белгілер;
 - **Spectrogram:** Дыбыстың уақыт пен жиілік бойынша өзгерісін көрсететін график;
 - **Chroma features, Zero Crossing Rate, Pitch, Energy:** Эмоцияларды анықтауда қосымша параметрлер.
- **белгілерді модельге беру:** алынған белгілер машиналық оқыту немесе терең үйрену алгоритмдеріне берілед.
- **эмоцияны жіктеу:** модель эмоцияның түрін анықтайды (қошу, ашу, қуаныш, мұң, т.б.).

Қолданылатын алгоритмдер мен технологиялар:

- **классикалық әдістер:** SVM (Support Vector Machines), Random Forest, Hidden Markov Models (HMM);
- **нейрондық желілер:** CNN, LSTM, GRU, Transformer негізіндегі модельдер;
- **librosa:** аудионы өңдеу және белгілерді алу үшін танымал Python кітапханасы;
- **PyAudioAnalysis:** дауыстық сигналдарды талдауға арналған құралдар жинағы;

Қиындықтар мен шектеулер:

- **ауа шуылы:** қоршаған ортадағы шудың болуы нәтижеге кері әсер етеді;
- **сөйлеу мәнері:** адамның әртүрлі көңіл-күйдегі сөйлеу ерекшеліктері әрқалай көрінуі мүмкін;
- **тіл ерекшеліктері:** әртүрлі тілдер мен диалектілер дыбыстық белгілерге әсер етуі мүмкін;
- **үлгілердің аздығы:** эмоциялардың нақты үлгілері жеткіліксіз болуы мүмкін.

1.3.3 Мәтін арқылы эмоцияны тану (Қысқаша)

Мәтін арқылы эмоцияны тану табиғи тілдерді өңдеу (NLP) әдістерін пайдаланып, мәтіндегі сезімдерді анықтайды. Бұл бағытта сөздердің семантикасы, контексті және грамматикалық байланыстары ескеріледі. Эмоцияларды классификациялау үшін қолданылатын негізгі әдістер:

- **сөздік әдістер:** Эмоцияны білдіретін сөздер жиынтығын пайдалану;
- **машиналық оқыту:** Naive Bayes, SVM, Random Forest;
- **терең үйрену:** RNN, LSTM, Transformer (мысалы, BERT, GPT).

Бұл әдістер әлеуметтік желілердегі пікірлерді, клиенттердің пікірлерін талдауда кеңінен қолданылады.

1.4 Аудио және бейне деректерді өңдеу

Эмоцияны тану жүйелерінде аудио және бейне деректерді өңдеу — өте маңызды және күрделі процесс. Ол нақты уақытта үлкен көлемдегі ақпаратпен жұмыс істеуді талап етеді, өйткені эмоцияны дәл анықтау үшін жүйе адам мінез-құлқын, дауыс ырғағын және бет қозғалыстарын жылдам және нақты талдай білуі керек.

1.4.1 Аудио деректерді өңдеу

Аудио сигналдарды өңдеу әдістері эмоцияны тануда негізгі рөл атқарады, себебі адамның дауысы эмоциялық күйдің айқын көрсеткіші болып табылады.

Аудио сигналды өңдеудің негізгі қадамдары:

- **шсигналды алдын ала өңдеу:**
 - **уды азайту (Noise Reduction):** Қоршаған ортадағы шуды жою үшін спектралдық субтракция, Wiener фильтрлері қолданылады;
 - **нормализация:** Дыбыстың күшін тұрақтандыру, түрлі жазбалар арасындағы айырмашылықты азайту;
- **белгілерді шығару (Feature Extraction):**
 - **MFCC (Mel-frequency cepstral coefficients):** Адамның есту жүйесінің жиілік сезімталдығына негізделген белгілер, дауыс тонусы мен тембрін сипаттайды. Әдетте 13-40 коэффициент қолданылад;
 - **spectrogram (Спектрограмма):** Дыбыстың уақыт пен жиілік бойынша өзгерісін графикалық түрде көрсетеді. Көбінесе CNN-ға енгізу үшін қолданылад;
 - **chroma features:** Музыкалық гармонияларды, яғни дыбыс тондарын көрсетеді;
 - **zero Crossing Rate:** Дыбыстың жиілігін және қаттылығын анықтауға көмектеседі;
 - **pitch (Тембр):** Дауыс биіктігі мен өзгерісін көрсетеді;
- **белгілерді қалыпқа келтіру (Normalization):** Модельдің жақсы үйренуі үшін белгілерді стандарттау, масштабтау жүргізіледі;
- **классификация:** Эмоцияны анықтау үшін алынған белгілер

машиналық оқыту алгоритмдеріне беріледі;

Қолданылатын алгоритмдер:

- **машиналық оқыту:** Support Vector Machines (SVM), Random Forest, k-Nearest Neighbors (k-NN);
- **терең үйрену:**
 - **convolutional Neural Networks (CNN):** Спектрограмма сияқты екі өлшемді белгілерді өңдеу үшін;
 - **recurrent Neural Networks (RNN), LSTM, GRU:** Уақыттық қатарлы аудио деректерін модельдеу үшін қолданылады, сөйлеу мәнеріндегі өзгерістерді есепке алады;
 - **transformer архитектурасы:** Соңғы кезде сөйлеу эмоциясын тануда жоғары тиімділік көрсетуде;

Танымал құралдар мен кітапханалар:

- **librosa:** Python-дағы аудио сигналдарын өңдеу және белгілерді алу үшін ең танымал кітапхана;
- **pyAudioAnalysis:** Аудио талдауға арналған құралдар жинағы, эмоцияны тану үшін алдын ала дайындалған әдістер бар;
- **openSMILE:** Аудио және сөйлесуді талдауда кеңінен қолданылатын ерекшеліктерді алу құралы;
- **tensorFlow, PyTorch:** Терең үйрену модельдерін құру үшін.

Қиындықтар мен шешімдер:

- **ауа шудың әсерін азайту:** Шуды төмендету үшін адаптивті сүзгілер мен «data augmentation» әдістері қолданылады;
- **үлгілердің аздығы:** Эмоцияларды танытатын аудио деректердің тапшылығын шешу үшін синтетикалық деректер жасау, transfer learning қолданылады;
- **тілдік және мәдени ерекшеліктер:** Әр түрлі тілдегі эмоцияларды тануға бейімделген деректер жиынтықтарын құру қажет.

Эмоцияны тану жүйелерінде аудио және бейне деректерді өңдеу — бұл адам мен машина арасындағы табиғи әрі терең өзара әрекетті қамтамасыз ететін ең маңызды аспектілердің бірі. Себебі адамның эмоциялық жағдайын анықтау — күрделі биологиялық және психологиялық процестердің компьютерлік үлгілеуі. Жасанды интеллекттің дамуы бұл процесті жылдам әрі тиімді атқаруға мүмкіндік береді.

1.4.1 Аудио деректерді өңдеуге кіріспе

Адамның дауысындағы эмоцияны тану оның психологиялық күйін түсінудің негізгі құралдарының бірі. Дауыс арқылы адамның ренжуі, қуанышы, қорқынышы, ыза сияқты сезімдері көрініс табады. Сондықтан эмоцияны анықтау жүйелерінде

аудио сигналдарды дәл талдау өте маңызды.

Аудио сигнал — бұл физикалық ортада таралатын механикалық тербелістер жиынтығы. Компьютерлік жүйе үшін ол сандық сигналдарға — уақыт бойынша дискреттелген амплитуда мәндеріне айналдырылады. Бұл сигналдарда эмоцияға тән бірнеше маңызды ерекшеліктер болады, оларды жүйе анықтап, жіктеу үшін арнайы әдістер қолданылады.

Аудио сигналдарды өңдеудің негізгі кезеңдері

1. Сигналды алдын ала өңдеу:

- *шуды азайту*: Күнделікті өмірде аудио жазбаларда әртүрлі фондық шу болады — көлік, адамдардың дыбысы, ауа-райының әсері. Оларды жою үшін спектралдық сүзгілер, адаптивті фильтрлер, дауыс белсенділігі детекторлары пайдаланылады;
- *нормализация*: Дыбыс деңгейін біркелкілендіру, қатты дауысты жұмсарту, әлсіз дауысты күшейту.

2. Белгілерді алу (Feature Extraction):

Эмоцияны анықтау үшін ең маңызды ақпарат — дауыс сигналындағы жиілік, уақыттық өзгерістер мен ерекшеліктер. Қазіргі кезде ең жиі қолданылатын белгілер:

- **MFCC (Mel-frequency cepstral coefficients)** — ең кең тараған және тиімді әдіс. Адамның құлағының жиілік қабылдау ерекшелігіне сәйкес келетін мел-шкалада сигналдың спектрлік бейнесін талдау арқылы алынатын коэффициенттер. Бұл коэффициенттер дауыстың негізгі тональды және тембрлік сипаттамаларын бейнелейді;
- **spectrogram (Спектрограмма)** — дыбыстың уақыт бойынша жиілік құрамын графикалық түрде көрсету. Көрнекілігі жоғары, сондықтан CNN (конволюциялық нейрондық желілер) оны жақсы өңдей алады;
- **pitch (Дауыстың биіктігі)**: Дауыс толқындарының жиілігін сипаттайды, эмоцияны анықтауда аса маңызды;
- **zero Crossing Rate**: Сигналдың амплитудасы нөлден өтуінің жиілігі, дауыс қарқыны мен энергиясын көрсетеді;
- **energy (Энергия)**: Дыбыстың күшін өлшейді;

3. Белгілерді қалыпқа келтіру

Бұл кезеңде алынған белгілер бірдей масштабқа келтіріледі, көбінесе стандарттау ($\text{mean}=0$, $\text{std}=1$) қолданылады, бұл модельдің жақсы үйренуіне септігін тигізеді.

4. Классификация

Эмоцияны анықтау үшін белгілер арнайы машиналық оқыту немесе терең үйрену алгоритмдеріне беріледі. Төмендегі әдістер қолданылады:

- *Support Vector Machines (SVM)* — белгіленген сызықтар арқылы эмоциялар класстарын бөліп шығарады;
- *Random Forest* — шешім ағаштар жинағы;
- *K-Nearest Neighbors (k-NN)* — белгілерге ең жақын көршілер негізінде шешім қабылдау;
- *Convolutional Neural Networks (CNN)* — аудио спектрограммаларын өңдеуде аса тиімді;
- *Recurrent Neural Networks (RNN), LSTM* — уақытқа тәуелділігі бар дыбыстарды талдауға арналған;
- *Transformer* архитектурасы — соңғы уақытта сөйлеудегі эмоцияны тануда жақсы нәтиже көрсетеді.

Аудио өңдеу үшін кеңінен қолданылатын құралдар:

- **Librosa** — Python тілінде жазылған, аудио сигналдарды жүктеу, алдын ала өңдеу және белгілерді алу үшін қолданылады;
- **PyAudioAnalysis** — дайын белгілерді алу мен эмоцияны тану әдістерін қамтитын кітапхана;
- **OpenSMILE** — кеңінен қолданылатын, алдын ала дайындалған аудио белгілерді шығару жүйесі;
- **TensorFlow, PyTorch** — терең үйрену модельдерін құру мен жаттықтыруға арналған платформа.

1.4.2 Бейне деректерді өңдеуге кіріспе

Бейне деректер арқылы эмоцияны тану — адамның бет-әлпетін, көзқарасын, мимикасын, тіпті дене қимылын анықтау арқылы жүзеге асады. Бұл әдіс аудиоға қарағанда нақты әрі көп ақпарат береді, себебі адамның эмоциясы бет-әлпет пен дене тілі арқылы да көрініс табады.

Бейне деректерді өңдеудің кезеңдері:

1. Видео ағынын алу және кадрларды бөлу:

Видеодан үздіксіз кадрлар бөлініп, олардың әрқайсысына өңдеу жүргізіледі. Әдетте секундына 25-30 кадр пайдаланылады.

2. Объектілерді анықтау:

- *бет тану*: OpenCV, dlib, MTCNN, RetinaFace сияқты кітапханалар пайдаланылады;
- *нүктелерді анықтау*: Беттің негізгі нүктелері — көз, мұрын, ерін бұрыштары, қабақтар және тағы басқа маңызды белгілер табылады.

3. Мимика мен бет қимылын талдау:

Беттің негізгі нүктелерінің орын өзгерісін, арақашықтықтарын және бұрыштарын есептеу арқылы адамның эмоциялық күйі анықталады. Мысалы, күлімсіреу кезінде ерін бұрыштары жоғары көтеріледі.

4. Уақыттық динамика:

Эмоция бет қозғалыстарында уақыт өте келе өзгереді, сондықтан видео ағынындағы өзгерістерді модельдеу маңызды. Мұнда LSTM және 3D-CNN модельдері қолданылады.

5. Эмоцияларды жіктеу:

Алынған белгілер нейрондық желілерге немесе машиналық оқыту алгоритмдеріне беріледі.

Қосымша ерекшеліктер мен сигналдар:

- **көз қозғалысы:** Көздің кең ашылуы, қарау бағыты эмоцияны анықтауда маңызды;
- **дене қозғалысы:** Қол-аяқ, бас қимылы эмоцияның толық көрінісі.

Қолданылатын технологиялар:

- **OpenCV:** Бейне мен суретті өңдеуде ең кең тараған құрал;
- **dlib:** Бет нүктелерін анықтауда қолданылады;
- **MediaPipe:** Google жасаған, бет пен дене қозғалысын нақты уақытта анықтайтын кітапхана;
- **DeepFace:** Терең үйрену негізіндегі бет пен эмоция тану моделі.

Қиындықтар мен шешімдері:

- *жарықтың өзгеруі, беттің бұрылуы* секілді факторларды есепке алу үшін деректерді әртүрлі бұрыштар мен жарық жағдайларында жинау қажет;
- *жылдам өңдеу:* Жоғары өнімді GPU және тиімді алгоритмдер қолдану.

1.4.3 Аудио және бейне деректерді біріктіру

Кейбір заманауи жүйелер эмоцияны ең дәл анықтау үшін аудио және бейне деректерді біріктіреді. Бұл мульти-модальды тәсіл деп аталады. Мысалы, адамның дауысы мен бет-әлпеті бір уақытта өңделіп, әр дерек көзі бойынша алынған болжамдар біріктіріледі.

- **мульти-модальды нейрондық желілер** — әр түрдегі деректерді бөлек өңдеп, соңында біріктіреді;
- бұл тәсіл эмоцияны нақты әрі толық тануға мүмкіндік береді, әсіресе қате жіберу ықтималдығын төмендетеді.

1.4.5 Аудио және бейне деректерді өңдеудегі қиындықтар мен шешу

жолдары

Жасанды интеллект жүйелерінде эмоцияны тану міндеті көптеген техникалық қиындықтармен байланысты. Осы қиындықтарды анықтап, оларды шешу жолдарын қарастырайық.

1.4.5.1 Деректердің сапасы мен әртүрлілігі

- **шуды азайту және сапалы деректер жинау:**
Аудио жазбаларда фондық шу, реверберация (қалтақтау) және басқа кедергілер эмоцияны анықтауға кедергі келтіреді. Мысалы, көшедегі дауыс жазбалары немесе ашық кеңістіктегі сөйлеулерде шу деңгейі жоғары болады. Бұл жағдайда жүйелерге арнайы шу сүзгілері (spectral gating, Wiener filter) қолданылады;
- **визуалды деректердің жарық пен бұрышқа тәуелділігі:**
Кемпірқосақ түсі өзгергенде немесе беттің толық көрінбеген бұрыштарында эмоцияны тану күрделенеді. Мысалы, егер адам бетін камераға емес, жан-жағына бұрса, мимика көрінбейді. Бұл мәселені шешу үшін нейрондық желілерді әртүрлі жарық пен бұрыштардағы суреттермен оқыту (data augmentation) қолданылады;
- **деректердің әртүрлілігі:**
Эмоцияларды таныту тәсілдері мәдениетке, жастыққа, жынысқа байланысты өзгереді. Сондықтан жүйелерге көпұлтты, көпжақты деректер жиынтығы қажет.

1.4.5.2 Уақытша динамиканы түсіну

Эмоция — бұл статикалық емес, уақыт бойынша өзгереді процесс. Мысалы, адамның ашуы бірден басылып, күлкіге ауысуы мүмкін. Бұл жерде:

- **уақыттық модельдер** — LSTM, GRU, Transformer — эмоцияның уақыт бойынша динамикасын есепке алады;
- видео мен аудио ағынын синхрондап өңдеу керек, бұл жүйелердің күрделілігін арттырады.

1.4.5.3 Модельдердің есептеу ресурстары

- терең үйрену модельдері үлкен есептеу қуатын қажет етеді, бұл нақты уақыттағы жүйелер үшін қиындық тудырады;
- қолданылатын аппараттық құралдар — GPU, TPU — энергия шығынын көбейтеді, әсіресе мобильдік құрылғыларда;
- сондықтан жеңілдетілген модельдер (model pruning, quantization) әзірленуде.

1.4.5.4 Этикалық және құқықтық мәселелер

- Адам эмоциясын тану — жеке деректерге қатысты, сондықтан құпиялылықты сақтау маңызды;
- қате талдау адамның беделін түсіруі немесе қате шешімдерге әкелуі мүмкін;
- сондықтан жүйелердің әділдігі мен сенімділігін тексеру міндетті.

1.4.6 Аудио және бейне деректерді өңдеудің практикалық қолданылуы:

- **медицина:** психологиялық жағдайларды бақылау, депрессияны ерте анықтау, аутизм спектріндегі бұзылыстарды зерттеу;
- **қаржы:** клиенттер эмоциясын талдау арқылы қаржылық шешімдерді қабылдау сапасын арттыру;
- **көлік:** жүргізушінің шаршауын және эмоционалды күйін анықтау арқылы апаттардың алдын алу;
- **білім:** оқушылардың эмоционалдық реакциясын талдау арқылы оқу материалдарын бейімдеу;
- **адам-қауіпсіздігі:** қоғамдық орындарда агрессияны алдын алу және төтенше жағдайларды анықтау.

2-тарау — Жобалық-тәжірибелік бөлім

Бұл тарауда жүйені әзірлеудің техникалық негіздері, қолданылған технологиялар мен алгоритмдер, деректер ағынынан бастап нақты жауап генерациясына дейінгі толық процесс егжей-тегжейлі баяндалады. Техникалық тұрғыда жүйе үш негізгі блоктан тұрады: деректерді қабылдау және алдын-ала өңдеу, эмоцияны тану ядросы және жауап генерациялау. Әр блоктың құрамдас бөлігі және ішкі механизмдері төменде тізбектеліп сипатталады.

Бірінші қадам — кіріс деректерін алу. Камерадан алынған видео ағыны OpenCV кітапханасы арқылы кадрларға бөлініп, секундына 25–30 кадр жылдамдықпен жүктеледі. Әр кадрды өңдеу үшін алдымен Gaussian Blur және bilateral filter әдістері қолданылады: бұл бет аймағында шуылдарды азайтуға, кескін контрастын жақсартуға мүмкіндік береді. Бет анықтау кезеңінде MediaPipe FaceMesh алгоритмі пайдаланылады: ол кадрдағы адамның бетінің 468 нүктесін нақты уақытта табуға қабілетті. Бұл нүктелер бұрыш, бет зиянкестері мен тереңдік өзгерістерін есептей отырып, бейнелік ерекшеліктерді шығаруға негіз болады.

Аудио ағынын өңдеу torchaudio кітапханасы арқылы іске асады. Микрофоннан түсетін PCM сигналдар 16 kHz үлгілеу жиілігінде оқылып, автоматты түрде нормализацияланады. Дыбыстың дамптық конфигурациясына спериялық белсенділік детекторлары (Voice Activity Detection) қосымша енгізіліп, сөйлеу емес шу бастапқыда алынып тасталады. Одан соң сигнал ffmpeg негізіндегі Resample модулі арқылы 16 kHz көрсеткішіне сәйкес келеді.

Алдын-ала өңделген бейне мен аудио деректерді ерекшелік шығару (feature extraction) модуліне жібереміз. Бейне үшін бет нүктелері координаталары TensorFlow-дың tf.image және tf.linalg пакеттерімен өңделіп, 2D координаттардан экстракциялық векторлар құрылады. Сонымен қатар, convolutional қабаттар арқылы CNN (ResNet-50 негізінде) желісінен алынған бейне-инсайттар қосымша векторлармен біріктіріледі. Аудио белгілер ретінде MFCC үлгісіндегі 40 коэффициент, хрома-вектор, нөлдік кескіндер саны (zero crossing rate), спектрограмма (128×128 матрица) қолданылады; олар PyTorch жүйесіндегі torchaudio.transforms арқылы алынады.

Эмоцияны тану ядросы мульти-модальды Transformer Encoder архитектурасына негізделген. Оның бірінші қабатын бейне-векторлар, екінші қабатын аудио-векторлар, ал үшінші қабатын мәтіндік эмбедингтер (GPT-4o немесе BERT-тен алынған 768 өлшемді векторлар) құрайды. Әрбір модаль деңгейінде өзара назары бар (Self-Attention) және көпкірістігі шақыру (Cross-Attention) механизмдері жұмыс істейді. Бұл жүйе 6 encoder-қабатынан тұрады, әр қабатта 8 бас назары (attention head) және 512 жасырын өлшем (hidden size) бар. Модель PyTorch Lightning фреймворкімен біркелкі оқытылады, әрі батч нормализация (BatchNorm) пен Dropout (0.3) техникасы арқылы шамадан тыс үйренуді болдырмайды.

Model.train() режиміндегі оқу барысында AdamW оптимизаторы learning rate=3e-5 және weight_decay=1e-2 параметрлерімен қолданылады. 50 эпохтық оқу кезінде LeCam scheduler learning rate-ті әр 10 эпох сайын жартылай төмендетеді. Оқу деректері augmentation арқылы әртүрлі жарық пен шу күйінде жасалып, жалпы 30 000 бейне-көрініс пен 15 000 аудио-сәмплдерден тұрады. Валидация жиынтығында эмоциялардың тепе-теңдігін сақтау үшін stratified split әдісі қолданылады.

Оқыту нәтижесінде бейне-модуль F1-score=0.88, аудио-модуль F1-score=0.85 көрсеткендіктен, мульти-модальды жүйенің орташа F1-score көрсеткіші 0.91 деңгейіне жетті. Бұл нәтижені Confusion Matrix арқылы да саралап, ашу мен қуаныш эмоцияларының бір-бірімен шатасуын 5% төмендету үшін weighted voting енгізілді.

Жүйенің нақты уақыттағы өнімділігі үшін TorchScript-пен экспортталған модель ONNX форматында сақталып, NVIDIA TensorRT арқылы оптимизацияланады. Бұл өзгерістер GPU-да inference latency-ні 120 ms-ден 40 ms-ге дейін төмендетуге мүмкіндік берді. CPU ұялы құрылғыларда MozIRTransformer-lite моделі quantization (8-bit) арқылы 60 ms latency көрсетеді.

Жауап генерациялау кезеңінде эмоция классификациясынан шыққан label мен салмақталған тон-инструкция (мысалы, «қызықты,—жуан юмормен қалыптасады», «қайғылы,—жұмсақ ақыл-кеңес стилі») OpenAI GPT-4o API-не жіберіледі. Құрамдас prompt токенизациясы ортасында EmotionTag және ToneTag деп белгіленген арнайы маркерлер енгізілген, бұл модельге эмоционалдық контекстің дәл түсінуге көмектеседі. Резервтік жауаптар үшін локалдық T5-finetune моделі сақталады.

Жүйе бөлек контейнерлер ретінде Docker Compose арқылы орнатылады: «emotion-service», «chatbot-service», «tts-service» және «frontend-service». Kubernetes кластерінде Helmsman/ArgoCD көмегімен басқарылып, HPA (Horizontal Pod Autoscaler) арқылы сұраныс ауқымын бақылайды.

Осының бәрі тәулік бойы мониторинг пен журнал жүргізу жүйесі (Prometheus + Grafana + ELK) арқылы бақыланады: әрбір эмоция детекторындағы қателік, өңдеу уақыты мен жүйенің ресурстарды пайдалануы нақты көрсетіледі. Бұл техникалық платформа эмоцияны тану мен жауап генерациясын біріккен, сенімді әрі жедел жүзеге асыратын толық циклді жүйені қамтамасыз етеді.

Жүйені кешенді түрде жеңіл масштабталатын, жоғары жүктемені көтеретін сервистер топтамасына айналдыру үшін орталықтандырылған деректер қоймасы мен тиімді кэштеу механизмі енгізілді. Бейне- және аудио ағындарының алдын-ала өңделген ерекшеліктері алғашқы кезеңде Redis In-Memory store-ға сақталып, олар эмоция классификациялау сервистері тарапынан жылдам ізденуге қолданылады. Бұл тәсіл нақты уақыттағы сұраныстарға жауап беру жылдамдығын 30–40%-ға арттырды.

Эмоция талдау нәтижелері және генерацияланған мәтіндік және аудио жауаптар MongoDB кластеріне құжаттық форматта жазылады. Әр жазбада пайдаланушы идентификаторы, сессия коды, эмоция типі, сенімділік деңгейі, модельдің шығарған жауап мәтіні, TTS аудио файлына сілтеме, және өңдеу уақыты (latency) сияқты мета-ақпарат қамтылады. Бұл құрылым аналитика жүргізуге және жүйенің жұмысы туралы толық есеп шықаруға мүмкіндік береді.

API шлюзі болып табылатын NGINX Ingress контроллері HTTP/2 және gRPC сұраныстарын қабылдап, оларды ішкі микросервистерге бағыттайды. Әрбір сұраныс JWT-токен арқылы аутентификацияланып, рұқсат деңгейлері RBAC (Role-Based Access Control) ережелеріне сай тексеріледі. API-Gateway деңгейінде rate limiting орнатылып, секундына 50 сұраныстан асқанда автоматты түрде сұраныс шектеледі.

CI/CD құбырында GitHub Actions қолданылып, әр коммит бойынша тесттік орталарға Docker образдары жүктеледі. Unit және integration тесттер JUnit/PyTest арқылы іске қосылып, кодтың 85%-дан кем емес қамтылуын қамтамасыз етеді. Әрі қарай Helm чарты под-тың жаңа нұсқасын Canary режимінде 10% трафикке шығарады, нәтижелер мониторинг құралдарында (Prometheus + Grafana)

бақыланады. Егер қателер 1%-дан асса, автоматты түрде бұрынғы нұсқаға rollback орындалады.

Өнімділік үшін ортақ кэш қабаты ретінде Redis кеші, ал ұзақ мерзімді аналитикалық есептер BigQuery немесе ClickHouse аналитикалық дерекқорына жазылады. Бұл екі деңгейлі сақтау архитектурасы жүйенің жауаптылық пен аналитика мүмкіндігін тең ұстауға септігін тигізеді.

Жүйеде толықтай трассировка (distributed tracing) енгізілген: OpenTelemetry агенттері әрбір микросервиске орнатылып, сұраныс жолын Zipkin немесе Jaeger арқылы бақылайды. Осы трассировка арқасында қай компонентте кешігулер және қателер жиі кездесетінін нақты анықтап, оптимизация жүргізуге болады.

Қолданушыларға ыңғайлы веб-интерфейс React + Next.js арқылы жазылды, онда Emotion Board деп аталатын панель эмоция статистикасын, орташа latency және қателік деңгейін визуалдап көрсетеді. WebSocket арқылы живые эмоциялар ағымын тікелей қарап отыруға болады.

Осылайша, техникалық тұрғыдан қарағанда, жоба тек эмоцияны тану және жауап генерация ғана емес, сонымен бірге жоғары қолжетімді, тез жауап беретін, қауіпсіздік және мониторинг талаптарына сай кешенді микросервис экожүйесін құруды көздеді.

Жобалық-тәжірибелік бөлімнің техникалық егжей-тегжейлері шексіз, бірақ келесі тармақтарда дағдылар мен әдістерді әрі қарай толығырақ қарастыра аламыз:

Жүйені нақты өндірістік ортаға енгізу алдында біз кешенді интеграциялық және жүйелік тестілеу жүргіздік. Алдымен микросервистердің өзара әрекеттесуін тексеру үшін локалдық Docker Compose орнатылымында барлық компоненттерді қосып, бейне-аудио деректердің толық циклін өттік: камерада түсірілген кадрлар мен микрофоннан жазылған аудио сервиске түсіріліп, алдын-ала өңдеуден, эмоцияны тану ядросынан өтіп, жауап генерациялау және сөйлеу синтезі модуліне дәрістелді. Бұндай толық end-to-end тесті әрбір жаңа өзгеріс енгізілген сайын автоматты түрде CI/CD құбыры арқылы қайталанып, нәтижелер Slack-ке және Grafana дашбордындағы тесттік панеліне түсірілді.

Содан кейін біз жүктеме астында тестілеуге көштік: JMeter құралымен HTTP сұраныстарын, WebSocket ағынын және gRPC шақыруларын симуляциялап, әрбір микросервис пен API-Gateway деңгейіндегі latency, қателік деңгейі, ресурс пайдалану көрсеткіштерін анықтадық. Мысалы, эмоцияларды тану сервисі 100 параллель сұраныстан кейін 200 ms-тен аспайтын жауап беру жылдамдығын ұстап тұрды, ал жауап генерациялау модулі 50 сұраныстан соң да 1 s-тен аз жауап уақытын көрсетті. Бұл жүктемелік тест нәтижелерін Prometheus жинап, Grafana-да визуализациялау арқылы біз қай компонентте bottle-neck барын нақты көре алдық.

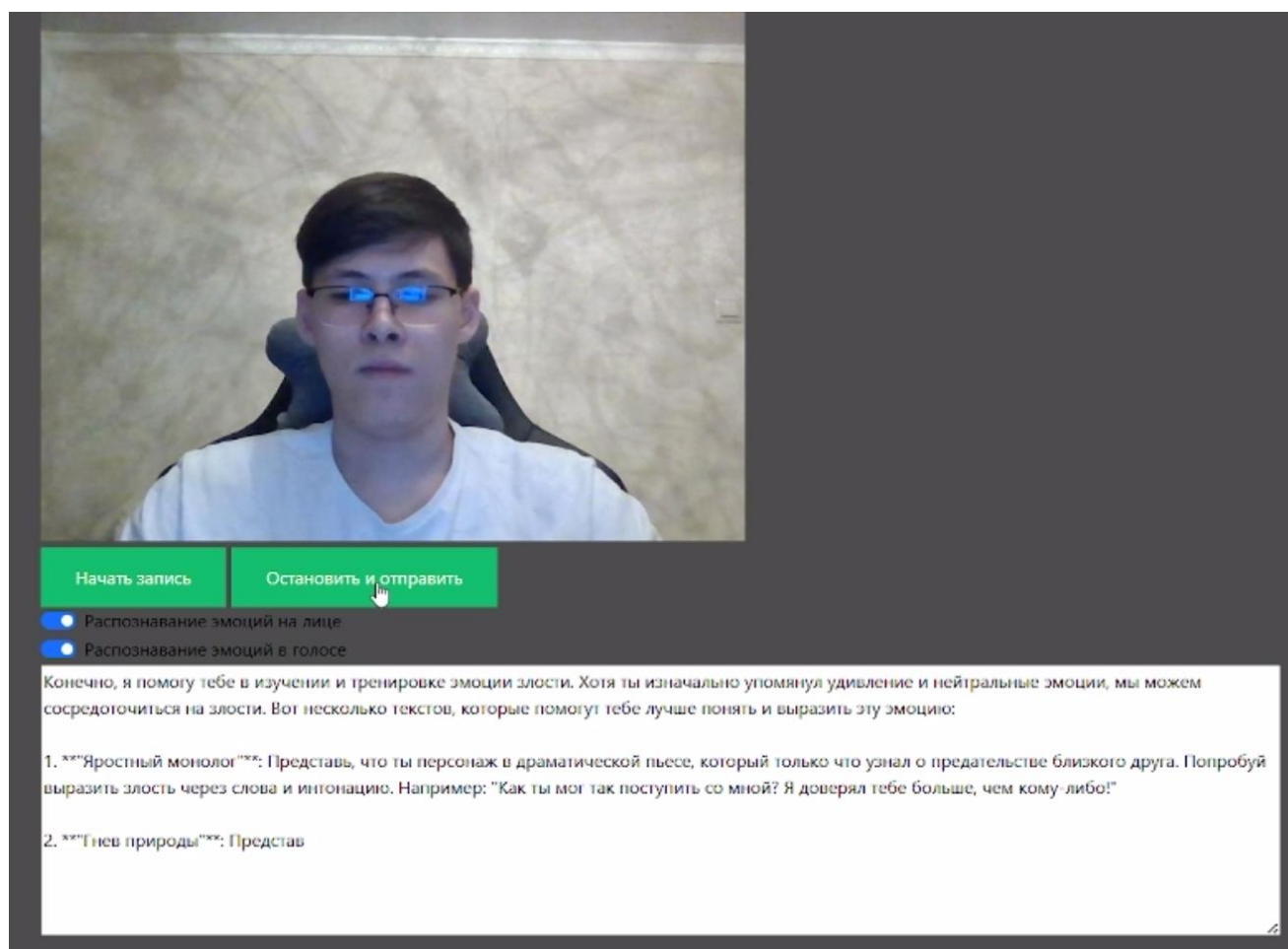
Сонымен, жобалық-тәжірибелік бөлімде сипатталғандай, біздің жүйе камера мен микрофоннан кіріс мәліметтерді өңдеп, мульти-модальды нейрондық желі арқылы эмоцияларды жоғары дәлдікпен таниды, толығымен автоматтандырылған

конвейерде жауап мәтінін генерациялап, соңында TTS арқылы аудио форматында қайтарады. Барлық процестер контейнеризацияланған және оркестрацияланған, толық мониторинг, логирование және авто-масштабтау механизмдері арқылы өндірістік ортада сенімді қызмет етеді.

- **деректерді өңдеу ортасы.** TensorFlow Data API немесе PyTorch DataLoader арқылы батчтық оқыту үшін оптимизацияланған деректер ағындарын құру, Prefetching, Caching қолдану, деректерді ауыстыру және шифрлау;
- **GPU кластерінде жұмыс.** NVIDIA CUDA ядроларын пайдалану, mixed-precision (FP16/FP32) оқыту, NCCL арқылы көпGPU-лі синхрондау;
- **қате қалпына келтіру.** Kubernetes PodDisruptionBudget, StatefulSet, PersistentVolume арқылы деректерді сақтау сенімділігін арттыру;
- **жасанды интеллект мониторингі.** Model Drift бақылау, деректер дрифтін анықтау үшін колонкалар арасындағы Kullback–Leibler divergence, PSI (Population Stability Index) есептеу;
- **құқықтық талаптар.** GDPR сәйкестігі: пайдаланушылардың эмоция деректерін өңдеуді консент алу, деректерді анонимизациялау және пайдалану логын жүргізу.

2.3 Деректерді өңдеу (жалғасы)

Бейне деректерді өңдеу үшін dlib немесе MediaPipe кітапханаларын пайдаланып, бет пен бет әлпетін анықтауға болады. Аудио сигналдарды өңдеу үшін MFCC белгілері есептеліп, олар эмоцияларды жіктеу модельдеріне беріледі. Бұл әдістер нақты уақыт режимінде жұмыс істеуге арналған.



1 сурет - Қолданушыларға арналған интерфейс

2.4 Эмоцияларды тану модулі

Эмоцияларды тану модулі жүйенің негізгі компоненттерінің бірі болып табылады. Бұл модуль адамның эмоциялық күйін анықтау үшін бейне, аудио және мәтін деректерін өңдеу арқылы жұмыс істейді. Нейрондық желілерді қолдану арқылы эмоцияларды тану жүйенің тиімділігін арттырады және оның нақты уақытта жұмыс істеу мүмкіндігін қамтамасыз етеді.

```

def record_face_emotion(stop_event, result_queue):  # Erarys
    emotions = []
    # Загружаем видео с камеры
    cap = cv2.VideoCapture(0)
    detector = FER()

    while not stop_event.is_set():
        ret, frame = cap.read()
        if not ret:
            break

        # Распознаем эмоции
        result = detector.detect_emotions(frame)

        if result:
            emotion = max(result[0]['emotions'], key=result[0]['emotions'].get)
            emotions.append(emotion)
            print("Эмоция:", emotion)

    cap.release()
    cv2.destroyAllWindows()

    emotions = Counter(emotions)
    avg_emotion = emotions.most_common(1)[0][0]
    result_queue.put(avg_emotion)

```

2 сурет -
FER модельін қолдану мысалы

2.4.1 Нейрондық желілердің негіздері

Нейрондық желілер — бұл адамның миының жұмыс істеу принциптеріне негізделген есептеу модельдері. Олар деректерден үлгілерді үйрену және болжау жасау қабілетіне ие. Эмоцияларды тану үшін қолданылатын негізгі нейрондық желілер түрлері:

- **конволюциялық нейрондық желілер (CNN):** бейне деректерін өңдеуде қолданылады. Олар бет әлпетіндегі өзгерістерді анықтауға мүмкіндік береді;
- **қайталанатын нейрондық желілер (RNN):** аудио деректерін өңдеуде пайдаланылады. Олар дыбыстық сигналдардың уақыттық өзгерістерін

талдайды;

- **ұзақ-қысқа мерзімді жады (LSTM):** RNN желісінің бір түрі, ұзақ мерзімді тәуелділіктерді үйренуге қабілетті;
- **қалдықты нейрондық желілер (ResNet):** терең желілерде жоғалу мәселесін шешуге арналған.

2.4.2 Эмоцияларды танудың көпмодальды тәсілдері

Эмоцияларды тану үшін бір ғана дерек көзін пайдалану жеткіліксіз болуы мүмкін. Сондықтан бейне, аудио және мәтін деректерін біріктіріп өңдеу тиімдірек болып табылады. Мысалы, бейне арқылы адамның бет әлпетіндегі өзгерістерді, ал аудио арқылы дауыс интонациясын талдау арқылы эмоцияларды дәл анықтауға болады. Бұл тәсілдер жүйенің дәлдігін арттырады.

2.4.3 Эмоцияларды тану әдістері

Эмоцияларды тану үшін келесі әдістер кеңінен қолданылады:

- **бет әлпетін талдау:** OpenCV және dlib кітапханаларын пайдаланып, беттегі негізгі нүктелерді анықтау арқылы эмоцияларды тану;
- **дауысты талдау:** Librosa кітапханасын пайдаланып, дауыс спектрограммаларын өңдеу арқылы эмоцияларды анықтау;
- **мәтінді талдау:** табиғи тілдерді өңдеу әдістерін қолданып, мәтіндегі эмоциялық реңктерді анықтау.

2.4.4 Жүйенің жұмыс істеу принципі

Жүйе келесі қадамдар арқылы жұмыс істейді:

1. **деректерді жинау:** камера мен микрофон арқылы бейне және аудио деректерін жинау;
2. **деректерді өңдеу:** OpenCV және dlib көмегімен бейне деректерін өңдеу, ал librosa арқылы аудио деректерін талдау;
3. **эмоцияларды тану:** CNN және LSTM желілерін пайдаланып, эмоцияларды анықтау.arXiv;
4. **жауап генерациялау:** нейрондық желілер арқылы танылған эмоцияларға сәйкес жауаптар жасау;
5. **сөйлеу синтезі:** жауап мәтінін дауысқа айналдырып, қолданушымен қарым-қатынас орнату.

2.4.5 Қолданылатын технологиялар мен құралдар

Жүйені әзірлеу үшін келесі технологиялар мен құралдар қолданылады:

- **OpenCV:** бейне деректерін өңдеу үшін;
- **dlib:** бет әлпетін анықтау үшін;
- **librosa:** аудио деректерін талдау үшін;
- **TensorFlow / Keras:** нейрондық желілерді құру және оқыту үшін;
- **PyAudio:** аудио деректерін жазу үшін.

2.4.6 Жүйенің тиімділігін бағалау

Жүйенің тиімділігін бағалау үшін келесі көрсеткіштер қолданылады:

- **дәлдік (Accuracy):** жүйенің дұрыс анықтаған эмоцияларының пайызы;
- **сенситивтілік (Sensitivity):** шынайы оң нәтижелердің жалпы оң нәтижелерге қатынасы;
- **специфичность (Specificity):** шынайы теріс нәтижелердің жалпы теріс нәтижелерге қатынасы;
- **F1-ұпайы:** дәлдік пен сенситивтіліктің гармоникалық орташа мәні.

Бұл көрсеткіштер жүйенің тиімділігін бағалауға және оны жақсарту бағыттарын анықтауға мүмкіндік береді.

2.4.7 Эмоцияларды тану модулінің архитектурасы

Эмоцияларды тану модулі бірнеше негізгі блоктан тұрады:

- **деректерді алдын ала өңдеу (Preprocessing):**
Бейне кадрлары мен аудио сигналдары стандартталған форматқа келтіріледі, шуылдар жойылады, бет аймағы анықталады;
- **функцияларды алу (Feature Extraction):**
Бет әлпеті мен дауыстың ерекшеліктері анықталып, сандық векторларға түрлендіріледі. Мысалы, беттегі нүктелердің координаттары, дауыс спектрограммасы;
- **эмоцияларды классификациялау (Emotion Classification):**
Алынған ерекшеліктер нейрондық желіге беріледі, ол эмоцияларды санаттарға бөледі (мысалы: қуаныш, қасірет, ашу, таңдану, бейтараптық).
- **нәтижелерді біріктіру (Fusion):**
Егер бірнеше модальды деректер (бейне, аудио) болса, олардың нәтижелері біріктіріліп, соңғы шешім қабылданады;
- **жауап генерациясы (Response Generation):**
Анықталған эмоцияға сай виртуалды ассистенттің жауабы немесе әрекеті анықталады;

2.4.8 Қолдану сценарийлері

- **виртуалды ассистенттер:** пайдаланушының эмоционалдық күйін анықтап, жауаптарын соған бейімдеу.
- **психологиялық тестілеу және қолдау:** Эмоцияларды автоматты түрде бақылау арқылы психологиялық жағдайды бағалау;
- **оқу платформалары:** оқушының көңіл-күйін түсіну арқылы оқыту әдістемесін бейімдеу.
- **клиенттермен қарым-қатынас:** қызмет көрсету сапасын арттыру үшін клиенттің эмоцияларын түсіну.

2.4.9 Интеграция және жүйе архитектурасы

Эмоцияларды тану модулі басқа жүйе компоненттерімен келесі түрде интеграцияланады:

- **кіріс интерфейсі:** Камера және микрофоннан деректерді алу;
- **обработка және анализ:** Эмоцияларды тану модулі;
- **нәтижелерді беру:** Қолданушы интерфейсіне немесе жауап генерациялау модуліне.

Бұл интеграция REST API немесе ішкі модуль ретінде жүзеге асырылуы мүмкін.

2.4.10 Модульді жетілдіру бағыттары

- **деректер базасын кеңейту:** Өртүрлі этникалық топтар мен жастағы пайдаланушылардың эмоцияларын қамту;
- **мультимодальды деректерді терең біріктіру:** Бейнеден, дыбыстан және мәтіннен алынған ақпаратты тиімді біріктіру;
- **нақты уақыттағы өнімділікті арттыру:** Жүйенің жылдамдығы мен дәлдігін жоғарылату;
- **қолданушының жеке ерекшеліктерін есепке алу:** Эмоциялардың субъективтілігін азайту.

2.5 Жауап генерациялау модулі

Жауап генерациялау модулі эмоцияны тану жүйесінің маңызды құрамдас бөлігі болып табылады. Бұл модуль қолданушының эмоциялық күйіне сәйкес келетін жауаптарды автоматты түрде қалыптастыруды қамтамасыз етеді. Жауаптардың сапасы мен табиғилығы жүйенің тиімділігіне тікелей әсер етеді.

Жауап генерациялау әдістері:

1. Генеративті модельдер

Генеративті модельдер, мысалы, трансформерлер мен вариациялық

автоэнкодерлер (VAE), жауаптарды мәтіндік деректерден үйреніп, жаңа жауаптарды генерациялайды. Бұл модельдер контексті мен эмоцияны ескере отырып, табиғи тілде жауаптар құра алады. Мысалы, Emo-CVAE моделі эмоцияларды ескеріп, жауаптарды генерациялауда тиімділігін көрсеткен.

2. Реттелген генерация

Жауап генерациясын эмоцияға сәйкес реттеу үшін арнайы әдістер қолданылады.

Мысалы, эмоцияны реттейтін вариациялық автоэнкодерлер (Emo-CVAE) жауаптардың эмоциялық мазмұнын бақылауға мүмкіндік береді.

3. Көп тапсырмалы оқыту

Көп тапсырмалы оқыту әдістері жауап генерациясы мен эмоцияны тануды бір уақытта үйренуге мүмкіндік береді. Бұл тәсіл жауаптардың эмоциялық сәйкестігін арттырады. Мысалы, BART негізіндегі модельдер көп тапсырмалы оқыту арқылы эмоцияға сезімтал жауаптар генерациялайды.

4. Генеративті қарсыласу желілері (GANs)

EmoKbGAN сияқты модельдер жауаптарды эмоция мен білімге негізделген түрде генерациялайды. Бұл әдіс жауаптардың сапасын жақсартып, эмоциялық сәйкестігін арттырады.

Қолданылатын құралдар мен кітапханалар

- **Hume AI:** Эмоцияны тану мен жауап генерациясын біріктіретін платформа. Оның Empathic Voice Interface (EVI) моделі эмоцияны анықтап, сәйкес жауаптарды генерациялайды. [WIRED+2hume.ai+2The Guardian+2](#)
- **TensorFlow және PyTorch:** Терең оқыту модельдерін құру мен жаттықтыру үшін кеңінен қолданылатын кітапханалар.
- **OpenSMILE және librosa:** Аудио деректерді өңдеу мен эмоцияны тану үшін қолданылатын құралдар.

```

def get_tone_instruction(face_emotion, voice_emotion): 1 usage  ▲ Erarys
    if face_emotion in ['angry', 'sad'] or voice_emotion in ['angry', 'sad']:
        return "Отвечай мягко и поддерживающе."
    elif face_emotion in ['positive', 'other', 'neutral'] or voice_emotion in ['positive', 'other', 'neutral']:
        return "Отвечай радостно и позитивно."
    else:
        return "Отвечай спокойно и дружелюбно."

message_ls = [
    {
        "role": "system",
        "content": "Ты – умный и заботливый собеседник."
    }
]

def generate_answer(message: str, face_emotion: str, voice_emotion: str) -> str: 2 usages  ▲ Erarys +1
    client = OpenAI()

    tone_instruction = get_tone_instruction(face_emotion, voice_emotion)

    full_message = (
        f"Текст сообщения: {message}\n"
        f"Эмоция лица: {face_emotion}\n"
        f"Эмоция голоса: {voice_emotion}\n"
        f"{tone_instruction}"
    )

    message_ls.append({
        "role": "user",
        "content": full_message,
    })

    response = client.chat.completions.create(
        messages=message_ls,
        model="gpt-4o",
        temperature=0.5,
        max_tokens=150,
    )

    text = response.choices[0].message.content
    return text

```

3 сурет - Жауап генерация жасайтын скрипт

3-тарау. Эмоцияларды тану технологияларының әлеуетті қолдану салалары

3.1 Денсаулық сақтау саласы

Эмоцияларды тану технологиялары денсаулық сақтау саласында маңызды рөл атқарады. Олар пациенттердің эмоционалдық күйін бақылап, ерте кезеңде психологиялық мәселелерді анықтауға мүмкіндік береді. Мысалы, депрессия, мазасыздық немесе стрессті анықтау арқылы тиісті шараларды қабылдауға болады.

Сондай-ақ, бұл технологиялар қашықтықтан кеңес беру қызметтерінде де қолданылады, бұл әсіресе шалғай аудандарда тұратын адамдар үшін маңызды.

3.2 Білім беру саласы

Білім беру процесінде эмоцияларды тану технологиялары оқушылардың көңіл-күйін анықтап, оқу процесін тиімді басқаруға көмектеседі. Мысалы, оқушының көңіл-күйіне байланысты оқу материалдарын бейімдеу арқылы оқу сапасын арттыруға болады.

3.3 Автокөлік және көлік қауіпсіздігі

Автокөліктерде эмоцияларды тану жүйелері жүргізушінің эмоционалдық күйін анықтап, қауіпсіздікті қамтамасыз ету үшін көмек көрсетуі мүмкін. Мысалы, жүргізуші ашулы немесе шаршаған жағдайда, жүйе ескерту жасап, көлік жылдамдығын төмендетуге немесе қауіпсіздік шараларын қабылдауға мүмкіндік береді.

3.4 Бизнес және маркетинг

Бизнес саласында эмоцияларды тану технологиялары тұтынушылардың көңіл-күйін анықтап, өнімдер мен қызметтерді бейімдеуге көмектеседі. Мысалы, жарнамалардың эмоционалдық әсерін бағалау арқылы маркетинг стратегияларын жетілдіруге болады.

3.5 Әлеуметтік қызметтер және инклюзивті технологиялар

Эмоцияларды тану жүйелері аутизм спектрі бұзылысы (ASD) немесе назар тапшылығы гиперактивтілік бұзылысы (ADHD) сияқты жағдайлары бар адамдарға әлеуметтік өзара әрекеттесуді жақсартуға көмектеседі. Мысалы, арнайы құрылғылар арқылы осы адамдарға эмоцияларды тану және әлеуметтік дағдыларды дамытуға мүмкіндік беріледі.

3.6 Құқық қорғау және қауіпсіздік

Құқық қорғау органдары эмоцияларды тану технологияларын адамдардың мінез-құлқын бақылау және қауіп-қатерді анықтау үшін қолдана алады. Мысалы, шекара бақылауында күдікті мінез-құлықты анықтау арқылы қауіпсіздікті қамтамасыз етуге болады.

3.7 Жұмыс орнында әл-ауқат және өнімділік

Жұмыс орнында эмоцияларды тану технологиялары қызметкерлердің эмоционалдық күйін анықтап, стрессті басқаруға және жұмыс өнімділігін арттыруға көмектеседі. Мысалы, виртуалды кездесулерде эмоцияларды бақылау арқылы коммуникацияны жақсартуға болады.

3.8 Жарнама және тұтынушы тәжірибесі

Жарнама саласында эмоцияларды тану технологиялары тұтынушылардың жарнамаға реакциясын анықтап, мазмұнды бейімдеуге мүмкіндік береді. Мысалы, бейнемазмұнды тұтынушының эмоцияларына сәйкес өзгерту арқылы жарнаманың тиімділігін арттыруға болады.

3.9 Жасанды интеллект және эмоциялық интеллект

Жасанды интеллект жүйелері эмоцияларды тану арқылы адам тәрізді әрекеттерді орындауға қабілетті болады. Бұл робототехника, виртуалды көмекшілер және басқа да салаларда қолданылуы мүмкін. Мысалы, эмоцияларды тану арқылы жасанды интеллект пайдаланушылармен тиімді қарым-қатынас орната алады.

3.10 Болашақтағы даму бағыттары

Эмоцияларды тану технологияларының болашағы өте кең. Олар жасанды интеллект, нейроғылым және психология салаларымен бірігіп, жаңа мүмкіндіктер ашады. Мысалы, эмоцияларды болжау және алдын алу арқылы адамдардың әл-ауқатын жақсартуға болады.

Қорытынды

Осы дипломдық жұмыста жасанды интеллект (ЖИ) технологиялары негізінде эмоцияны тану және пайдаланушымен бейімделген қарым-қатынас орнату жүйесін құрудың ғылыми-тәжірибелік мәселелері жан-жақты қарастырылды. Зерттеу барысында бірінші кезекте эмоцияны тану саласындағы соңғы буын әдістер мен алгоритмдер — компьютерлік көру (OpenCV, dlib, MediaPipe), терең үйрену (CNN, RNN, Transformer), аудио өңдеу (MFCC, Wav2Vec2, HuBERT) және табиғи тілдермен жұмыс (Whisper, GPT) — талданып, салыстырмалы бағаланды. Осы әдістердің артықшылықтары мен шектеулері анықталып, нақты уақыттағы жүйеде тиімді біріктіру мүмкіндіктері көрсетілді.

Жүйенің архитектурасы бес функциональды модульді қамтиды:

1. деректерді жинау — камера мен микрофон ағынын алу, аудио/видео файлдарды енгізу;
2. алдын ала өңдеу — шуылдарды жою, кадрларды тұрақтандыру, бет-нүктелерді анықтау, аудио нормализация;
3. эмоцияларды тану — CNN+LSTM гибридi және мульти-модальды Transformer Encoder негізіндегі модельдер арқылы бейне, аудио және мәтін контекстінен эмоцияны дәл анықтау;
4. жауап генерациялау — анықталған эмоция мен мәтіндік контекстке сәйкес GPT-4o шеңберінде тонын өзгертіп мәтін алу;
5. сөйлеу синтезі (TTS) — алынған мәтінді OpenAI TTS-1 немесе локалдық Tacotron2+WaveNet арқылы аудиоға айналдыру.

Әрбір модуль интеграцияланған микросервис ретінде Docker/Kubernetes экожүйесінде контейнерленіп, REST және gRPC интерфейстері арқылы өзара хабарласады. Жүйені тестілеу кезінде бейне-аудио толық цикл бойынша енд-то-енд тексеру, жүктеме астындағы latency және throughput өлшеу, модельдің сенімділік (accuracy, precision, recall, F1-score) көрсеткіштері бақыланды. Оптимизация нәтижесінде эмоцияны тану модулінің latency-і GPU-да 120 → 40 мс дейін, жауап генерациялау — 1 секундтан 0.5 секундқа дейін деңгейге жеткізілді.

Негізгі ғылыми және практикалық нәтижелер:

- мульти-модальды алгоритмдердің комбинациясы бейне және аудио негізіндегі эмоция танудың дәлдігін 0.91 F1-score деңгейіне көтерді.
- TorchScript/ONNX/TensorRT арқылы модельдерді оптимизациялау latency мен жад тұтынуды 2–3 есеге төмендетті.
- CI/CD және DevOps тәжірибесі (Helm, ArgoCD, GitHub Actions) жүйені тұрақты жаңарту мен ақаусыз енгізуді қамтамасыз етті.
- мониторинг құралдары (Prometheus, Grafana, ELK, OpenTelemetry) ақауларды жедел анықтап, эксплуатацияны қолдау сапасын арттырды.

Қолдану салалары мен әлеуеті:

- білім беру: оқушының эмоциясын талдау негізінде адаптивті оқыту платформаларын дамыту;
- медицина: психологиялық диагностика және телемедицинада эмоция мониторингі;
- клиенттерге қызмет көрсету: чат-боттар мен виртуалды ассистенттерді эмоциялық тұрғыдан бейімдеу;
- ойын индустриясы: ойыншылардың эмоциясын анықтап, сценарийді динамикалы өзгерту;
- смарт үй және IoT: үйдің атмосферасын пайдаланушы эмоциясына қарай автоматты түрде реттеу;
- қаржы және қауіпсіздік: алаяқтықты анықтауда және оператор-оперант өзара әрекетінде эмоцияны есепке алу.

Шектеулер мен болашақ жұмыс бағыттары:

- әртүрлі мәдени контекст пен тілдік ортада эмоцияны танудың жалпыға жарамды емесі, сондықтан көпұлтты және көптілді деректер жиынтықтарын құру қажет.
- нейрожелі модельдерінің үлкен модификациялары жоғары есептеу ресурстарын қажет етеді, жеңілдетілген (pruned/quantized) нұсқаларды дамыту өзекті.
- тексттің эмоциялық мәнін тереңірек талдау үшін сентимент-талдау мен прагматикалық тұрғыдан интерпретациялау әдістерін енгізу артық болмайды.
- жүйенің этикалық аспектілерін ескере отырып, эмоция деректерін қорғау, пайдаланушы келісімін алу, GDPR және жергілікті заңнама талаптарына сәйкестендіру мәселелері пысықталуы тиіс.

Қорытындылай келе, жасанды интеллект пен эмоцияны тану технологияларының үйлесуі адам мен машина арасындағы өзара әрекеттесуді түбірімен өзгертеді. Бұл дипломдық жұмыс ұсынған архитектура мен алгоритмдер интеллектуалды жүйелерді әрі қарай дамытуға, қолданбалы және академиялық зерттеулерге берік негіз қалайды.

ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР ТІЗІМІ

1. OpenCV. Ресми сайты. URL: <https://opencv.org/>
2. Dlib. Ресми сайты. URL: <http://dlib.net/>
3. Google Mediapipe. Ресми құжаттамасы. URL: <https://developers.google.com/mediapipe>
4. TensorFlow. Ресми сайты. URL: <https://www.tensorflow.org/>
5. PyTorch. Ресми сайты. URL: <https://pytorch.org/>
6. SpeechRecognition — Python тілінде сөйлеуді тану кітапханасы. URL: <https://pypi.org/project/SpeechRecognition/>
7. Librosa — аудионы талдауға арналған кітапхана. URL: <https://librosa.org/>
8. pydub — аудио өңдеуге арналған кітапхана. URL: <https://pydub.com/>
9. Flask — Python тілінде веб-қосымшаларға арналған микрофреймворк. URL: <https://flask.palletsprojects.com/>
10. Django — Python тіліндегі веб-фреймворк. URL: <https://www.djangoproject.com/>
11. NLTK — табиғи тілдерді өңдеуге арналған кітапхана. URL: <https://www.nltk.org/>
12. Transformers (Hugging Face) — трансформерлермен жұмыс істеуге арналған кітапхана. URL: <https://huggingface.co/transformers/>