

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА КОМП'ЮТЕРНОЇ ІНЖЕНЕРІЇ

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «Оптимізація процесу обробки даних в державних
базах даних за рахунок інтеграції штучного інтелекту»

на здобуття освітнього ступеня бакалавра
зі спеціальності 123 Комп'ютерна інженерія
(код, найменування спеціальності)
освітньо-професійної програми Комп'ютерна інженерія
(назва)

Кваліфікаційна робота містить результати власних досліджень.
Використання ідей, результатів і текстів інших авторів мають посилання на
відповідне джерело

(підпис) Євгеній ПРИЙМАК
Ім'я ПРИЗВИЩЕ здобувача

Виконав: здобувач вищої освіти гр. КІД-42
Євгеній ПРИЙМАК
Ім'я, ПРИЗВИЩЕ

Керівник: Світлана КУФТЕРІНА
науковий ступінь, Ім'я, ПРИЗВИЩЕ
вчене звання

Рецензент: _____
науковий ступінь, Ім'я, ПРИЗВИЩЕ
вчене звання

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Навчально-науковий інститут Інформаційних технологій

Кафедра Комп'ютерної інженерії

Ступінь вищої освіти бакалавр

Спеціальність 123 Комп'ютерна інженерія

Освітньо-професійна програма комп'ютерна інженерія

ЗАТВЕРДЖУЮ

Завідувач кафедри Наталія
ЛАЩЕВСЬКА Ім'я, ПРІЗВИЩЕ
« » 2024 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Приймак Євгеній Олександрович

(прізвище, ім'я, по батькові здобувача)

1. Тема кваліфікаційної роботи: Оптимізація процесу обробки даних в державних базах даних за рахунок інтеграції штучного інтелекту

керівник кваліфікаційної роботи Світлана КУФТЕРІНА
(Ім'я, ПРІЗВИЩЕ, науковий ступінь, вчене звання)

затверджені наказом Державного університету інформаційно-комунікаційних технологій
від «27» лютого 2024 р. №36

2. Строк подання кваліфікаційної роботи «03» червня 2024 р.

3. Вихідні дані до кваліфікаційної роботи:

1) теза за дослідженням на тему: «Потенціал використання штучного інтелекту для збільшення швидкості та точності обробки даних в державних інформаційних системах»;

2) теза за дослідженням на тему: «Використання штучного інтелекту для вирішення помилок в державних базах даних».

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

1. Питання аналізу існуючих державних баз даних та принципів обробки даних в них.

2. Питання оптимізації процесу обробки інформації за допомогою штучного інтелекту.

3. Питання створення прикладу алгоритму оптимізації завдань обробки інформації.

5. Перелік ілюстративного матеріалу: презентація

6. Дата видачі завдання «27» лютого 2024 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Збір даних	28.02-22.03.2024	Виконано
2	Обробка матеріалу	23.03-05.04.2024	Виконано
3	Розробка теоретичної частини	06.04-10.04.2024	Виконано
4	Аналіз державних баз даних	11.04-30.04.2024	Виконано
5	Розробка алгоритму	01.05-03.05.2024	Виконано
6	Висновки за результатами аналізу	04.05-08.05.2024	Виконано
7	Розробка презентації	09.05-11.05.2024	Виконано
8	Передзахист	14.05-17.05.2024	Виконано
9	Здача в деканат	25.05-03.06.2024	Виконано

Здобувач вищої освіти

(підпис)

Євгеній ПРИЙМАК

(ім'я, ПРИЗВИЩЕ)

Керівник
кваліфікаційної роботи

(підпис)

Світлана КУФТЕРІНА

(Ім'я, ПРИЗВИЩЕ)

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня бакалавра: 37 стор., 11 рис., 1 табл., 30 джерел.

Мета роботи – Аналіз сучасних підходів до обробки інформації в державних базах даних та вивчення можливостей інтеграції алгоритмів штучного інтелекту для покращення цих процесів.

Об'єкт дослідження – Процеси обробки інформації з використанням алгоритмів ШІ та традиційний метод обробки даних в державних реєстрах.

Предмет дослідження – Державні бази даних, алгоритми штучного інтелекту.

Короткий зміст роботи: У цій роботі досліджується потенціал штучного інтелекту для покращення роботи державних системах. Аналізується сучасний стан обробки інформації в реєстрах, висвітлюються проблеми та виклики, пов'язані з обробкою великих об'ємів даних.

Досліджуються можливості інтеграції штучного інтелекту в державні бази даних. Розглядаються перспективи розвитку методів обробки інформації, аналізується світовий досвід використання алгоритмів машинного навчання в цій сфері, наводяться приклади їх роботи у відкритих сервісах України.

Розроблений та протестований алгоритм обробки інформації за допомогою мови Python. Обґрунтовано ефективність інтеграції досліджуваних методів в державні бази даних.

КЛЮЧОВІ СЛОВА: ШІ, МАШИННЕ НАВЧАННЯ, ДЕРЖАВНІ БАЗИ ДАНИХ, ОБРОБКА ІНФОРМАЦІЇ, АНАЛІЗ ДАНИХ, ОПТИМІЗАЦІЯ, ЕФЕКТИВНІСТЬ, РЕЄСТРИ.

ABSTRACT

Text section of the qualification work for obtaining a bachelor's degree: 37 pages, 11 figures, 1 table, 30 sources.

The aim of the work is to analyze modern approaches to information processing in government databases and to explore the possibilities of integrating artificial intelligence algorithms to improve these processes.

Object of the research: Information processing processes using AI algorithms and traditional data processing methods in state registers.

Subject of the research: Government databases, artificial intelligence algorithms.

Summary of the work: This paper explores the potential of artificial intelligence to enhance the functioning of government systems. It analyzes the current state of information processing in registers, highlights the problems and challenges associated with processing large volumes of data.

The possibilities of integrating artificial intelligence into government databases are investigated. The prospects for the development of information processing methods are considered, the global experience of using machine learning algorithms in this field is analyzed, and examples of their use in Ukraine's open services are provided.

An information processing algorithm using Python has been developed and tested. The effectiveness of integrating the studied methods into government databases is substantiated.

KEYWORDS: AI, MACHINE LEARNING, GOVERNMENT DATABASES, INFORMATION PROCESSING, DATA ANALYSIS, OPTIMIZATION, EFFICIENCY, REGISTERS.

ЗМІСТ

ВСТУП.....	9
РОЗДІЛ 1. СУЧАСНИЙ СТАН ОБРОБКИ ІНФОРМАЦІЇ В ДЕРЖАВНИХ БАЗАХ ДАНИХ	11
1.1 Огляд сучасного підходу до обробки, збереження та обміну інформацією в державних базах даних	11
1.2 Проблематика обробки великих об'ємів даних різних типів.....	14
1.3 Огляд існуючих баз даних, класифікація даних в них.....	16
1.4 Огляд сучасних технологій збереження інформації	18
РОЗДІЛ 2. ПОТЕНЦІАЛ ІНТЕГРАЦІЇ АЛГОРИТМІВ ШТУЧНОГО ІНТЕЛЕКТУ В ДЕРЖАВНІ БАЗИ ДАНИХ.....	21
2.1 Перспективи розвитку методів обробки інформації в державних системах з використанням штучного інтелекту	21
2.2 Аналіз використання штучного інтелекту для обробки даних у світі.....	24
2.3 Приклади роботи штучного інтелекту в відкритих сервісах України	30
2.4 Стандарти та протоколи використання ШІ	32
2.5 Економічний та соціальний вплив інтеграції алгоритмів машинного навчання в державні бази даних, можливі недоліки	35
РОЗДІЛ 3. РЕАЛІЗАЦІЯ ОБРОБКИ ІНФОРМАЦІЇ ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ НА ПРИКЛАДІ МОВИ PYTHON	38
3.1 Розробка та тестування архітектури алгоритму обробки даних	38
3.2 Обґрунтування ефективності інтеграції досліджуваних алгоритмів в державні бази даних.....	43
ВИСНОВКИ	45
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ	46
ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ (Презентація)	50

ВСТУП

У сучасному світі державні бази даних стають дуже важливими для забезпечення ефективного управління інформацією про громадян. Проте, з поширенням цифрових технологій і зростанням обсягів даних, стає очевидною необхідність удосконалення процесів обробки цих даних.

Актуальність теми. Актуальність обраної теми проявляється у необхідності підвищення ефективності управління в умовах постійного зростання обсягів інформації та швидких змін у суспільстві. Нові технології, такі як штучний інтелект, можуть стати ключовими інструментами для оптимізації цих процесів.

Мета дослідження. Метою цієї роботи є дослідження сучасних підходів до обробки інформації в державних базах даних та вивчення можливостей інтеграції алгоритмів штучного інтелекту для покращення цих процесів. З цією метою передбачається проведення наступних завдань:

- провести аналіз наявних державних баз даних;
- провести аналіз процесу збору та обробки інформації в державні бази даних;
- визначити недоліки традиційної моделі обробки інформації;
- провести аналіз прикладів використання штучного інтелекту;
- визначити потенціал застосування штучного інтелекту в державних базах даних;
- створити приклад алгоритму штучного інтелекту для оптимізації обробки даних.

Об'єкт дослідження. Об'єктом дослідження є процес обробки інформації в державних базах даних.

Предмет дослідження. Аналіз сучасних підходів та методів оптимізації процесу обробки інформації з використанням інтеграції алгоритмів штучного інтелекту.

Методи дослідження. У цій роботі використовувалися методи аналізу літературних джерел, порівняльного аналізу, моделювання та імплементації програмного забезпечення.

Наукова новизна та практична значущість. Результати цього дослідження можуть стати важливим внеском у розвиток сучасних методів управління державними базами даних та забезпечити практичне застосування нових технологій для оптимізації управління державними системами.

Апробація результатів та публікації, які були опубліковані в ході проведення дослідження:

1. Куфтеріна С. Р., Приймак Є. О. ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИРІШЕННЯ ПОМИЛОК В ДЕРЖАВНИХ БАЗАХ ДАНИХ. «Розвиток наук в умовах нової реальності: проблеми та перспективи» : II Міжнародна наукова конференція, м. Київ, 3 травня 2024 р. С. 139–140.

2. Куфтеріна С. Р., Приймак Є. О. ПОТЕНЦІАЛ ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ЗБІЛЬШЕННЯ ШВИДКОСТІ ТА ТОЧНОСТІ ОБРОБКИ ДАНИХ В ДЕРЖАВНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ. «Діджиталізація науки як виклик сьогодення» : VI Міжнародна студентська конференція, м. Суми, 3 травня 2024 р. С. 135–136.

РОЗДІЛ 1 СУЧАСНИЙ СТАН ОБРОБКИ ІНФОРМАЦІЇ В ДЕРЖАВНИХ БАЗАХ ДАНИХ

1.1 Огляд сучасного підходу до обробки, збереження та обміну інформацією в державних базах даних

В Україні обробка інформації в державних базах даних передбачає багатогранний процес, який включає в себе збір, введення, перевірку, обробку та зберігання даних. Розуміння тонкощів цього механізму дозволить зрозуміти всю структуру роботи органів влади. Тут пропонується огляд процесу обробки даних.

Збір і введення даних. Процес починається зі збору інформації із різних джерел, зокрема державних установ, громадян, підприємств та інших зацікавлених сторін. Дані можуть збиратися через онлайн-форми, паперові документи, електронні заяви або автоматизовані системи.

В Україні державні органи часто покладаються на громадян, які надають інформацію через онлайн-портали або фізичні документи, далі дані подаються до відповідних установ. Після збору вона вноситься в онлайн реєстри призначеними працівниками або посадовими особами, відповідальними за її введення. Цей етап передбачає введення інформації в державні системи, але якість його виконання залежить від людини, із-за людського фактору співробітники можуть допускати помилки при внесенні даних, задля вирішення цього питання був передбачений наступний крок.

Перевірка даних. Після надходження до баз даних зібрана інформація проходить перевірку для забезпечення точності, узгодженості та відповідності попередньо визначеним стандартам.

В Україні державні органи проводять валідацію та паралельну перевірку, щоб виявити помилки та невідповідності в даних.

Валідація - це процес перевірки введених даних на відповідність визначеним правилам або форматам. Вона зазвичай використовується для перевірки коректності та достовірності введених даних перед їх подальшим використанням у системі. Валідація може включати перевірку на наявність обов'язкових полів, правильний формат введених даних (наприклад, електронної пошти або номера телефону), а також логічну перевірку відповідності даних заданим умовам (наприклад, перевірка дати на коректність).

Паралельна перевірка - це процес порівняння даних, збережених в одній базі даних, з відповідними даними з інших джерел для виявлення розбіжностей і помилок.

Якщо валідацію на сьогоднішній день виконують завдяки програмному коду, то паралельну перевірку проводить людина, це знову повертає нас до проблеми людського фактору та можливих супутніх помилок.

Зберігання. Оброблені дані заносяться в державні реєстри, які служать сховищами для зберігання, упорядкування та доступу до величезних обсягів інформації.

В Україні державні органи використовують сучасні системи управління базами даних (СУБД) для ефективного та безпечного зберігання та керування даними. Ці системи забезпечують доступність, надійність і цілісність інформації за допомогою надійних заходів безпеки, процедур резервного копіювання та механізмів аварійного відновлення. Крім того, інфраструктура управління даними регулює, як вони зберігаються, доступ і керування ними, забезпечуючи відповідність нормативним вимогам і захищаючи конфіденційність та права громадян.

Доступ і пошук. Авторизовані користувачі, включаючи урядовців, службовців і громадян, можуть мати доступ і отримувати потрібну інформацію з державних баз даних через призначені форми та установи.

В Україні державні реєстри забезпечують безпечний доступ до даних через онлайн-портали та інформаційні системи. Для обмеження несанкціонованого доступу та захисту конфіденційної інформації реалізовано засоби контролю та

верифікації. Крім того, зручні інтерфейси та функції пошуку дозволяють користувачам швидко й ефективно отримувати певні набори даних.

Не менш важливим є доступ державних баз даних до документів одна одної, саме тому була створена система «Трембіта».

«Трембіта» - це національна система електронної взаємодії між державними інформаційними ресурсами, електронними реєстрами, інформаційними системами тощо. Вона покликана надати органам державної влади та місцевого самоврядування оперативний, безперешкодний і надійний доступ до інформації, необхідний їм для здійснення їх повноважень[21].

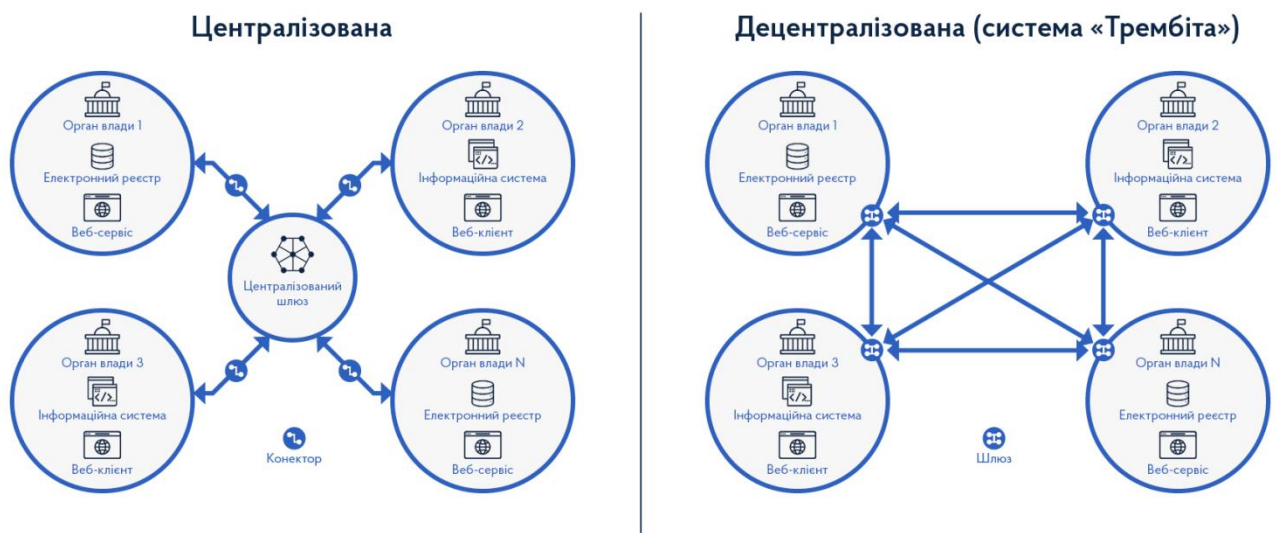


Рисунок 1.1 – Принцип роботи системи «Трембіта». Порівняння зі стандартним рішенням

Завдяки своїй структурі «Трембіта» дозволяє проводити паралельну перевірку в рази швидше.

Моніторинг і технічне обслуговування. Постійний моніторинг і технічне обслуговування є важливими для забезпечення постійної продуктивності, надійності та безпеки державних баз даних.

Державні установи України використовують інструменти та процеси моніторингу для відслідковування продуктивності бази даних, виявлення поломок

та завчасного вирішення проблем. Щоб запобігти втраті даних, системним збоям і порушенням безпеки, виконуються регулярні дії з технічного обслуговування, такі як резервне копіювання бази даних, оновлення програмного забезпечення та виправлення безпеки.

Сучасні підходи до обробки інформації в державних базах даних в Україні передбачають систематичний і структурований процес, який охоплює всі кроки від збору до зберігання, але сильно залежить від людини та ефективності її роботи[19].

1.2 Проблематика обробки великих об'ємів даних різних типів

Державні органи, як і будь-які інші організації, накопичують величезні обсяги даних про своїх громадян, ресурси, діяльність та інші аспекти функціонування. Ці дані можуть використовуватися для прийняття кращих управлінських рішень, підвищення ефективності роботи, надання кращих послуг населенню та забезпечення прозорості та підзвітності.

Однак обробка великих обсягів даних стає все більш складною задачею. Традиційні методи обробки даних, засновані на людській праці, часто є неефективними, дорогими та схильними до помилок. Зростання обсягів інформації та їх складності робить ці методи все більш нездатними впоратися з викликами сучасності.

Існує ряд проблем, з якими стикаються державні органи при обробці великих обсягів даних:

- **Обсяг даних.** Державні органи накопичують величезні обсяги даних з різних джерел, таких як переписи населення, реєстри майна, податкові декларації, медичні записи та кримінальні справи. Ці обсяги даних постійно зростають, що ускладнює їх зберігання, обробку та аналіз;

– **Різноманітність даних.** Дані, які збирають державні органи, часто є різноманітними за форматом, структурою та змістом. Це ускладнює їх інтеграцію та аналіз;

– **Швидкість обробки даних.** Державним органам часто потрібно швидко обробляти великі обсяги даних, щоб приймати своєчасні та обґрунтовані рішення. Традиційні методи обробки даних часто не можуть впоратися з цією швидкістю;

– **Якість даних.** Дані, які збирають державні органи, можуть бути неточними, неповноцінними або дублюватися. Це призводить до помилкових рішень та неефективного використання ресурсів;

– **Безпека.** Державні органи зобов'язані захищати конфіденційність даних своїх громадян. Це складна задача при обробці та збереженні великих обсягів даних.

Проблеми обробки великих обсягів інформації створюють серйозні наслідки для державних органів. Установи не зможуть вчасно приймати управлінські рішення, підвищувати ефективність роботи та надавати кращі послуги населенню, якщо вони не зможуть обробляти дані. Неточні або неповноцінні дані можуть призвести до помилкових рішень, що матиме негативні наслідки для громадян та держави. А нездатність ефективно обробляти інформацію може призвести до непрозорості та підзвітності державних органів. Громадяни можуть втратити довіру до державних органів, якщо вони вважають, що їхні дані не захищені або не використовуються належним чином.

Обробка великих обсягів даних стає все більш складною задачею для державних органів. Традиційні методи обробки даних можуть уникнути цієї проблеми лише за допомогою збільшення штату працівників, але це понесе за собою великі витрати з бюджету задля малоефективного тимчасового рішення.

1.3 Огляд існуючих баз даних, класифікація даних в них

Україна має розгалужену систему державних реєстрів, які зберігають в собі різноманітну інформацію, відомості про громадян, юридичні особи, ресурси, діяльність та інші ключові аспекти життя країни. Ці бази даних є невід'ємною частиною інфраструктури державного управління, оскільки вони сприяють ефективності владних структур, оптимізації надання послуг населенню та прийняттю відповідальних управлінських рішень.

Один з найбільш значущих реєстрів — **Єдиний державний реєстр фізичних та юридичних осіб**, що включає дані про всіх зареєстрованих в Україні фізичних та юридичних осіб. Ця база містить інформацію про ім'я, прізвище, дату народження, адресу, реєстраційний номер, а також відомості про місце роботи кожної зареєстрованої особи, що дозволяє уряду забезпечити контроль за діяльністю громадян на території держави.

Державний реєстр речових прав на нерухоме майно є іншим критично важливим інструментом, який містить інформацію про всі об'єкти нерухомості на території України, включаючи житлові будинки, земельні ділянки та комерційні приміщення. Він включає кадастрові номери, адреси, відомості про власників та право власності, що є основою для правового регулювання власності та забезпечує прозорість угод.

Єдиний державний реєстр судових рішень об'єднує дані про всі судові рішення, прийняті в Україні. Цей реєстр містить інформацію про номер справи, дату прийняття рішення, суть справи, сторони справи та повний текст рішення, що сприяє прозорості судових процесів та дозволяє громадянам слідкувати за легітимністю та обґрунтованістю судових вироків.

Єдиний державний реєстр кримінальних проваджень забезпечує облік всіх кримінальних справ, що зареєстровані в Україні. Він включає номери справ, дати порушення проваджень, статті Кримінального кодексу, інформацію про

підозрюваних, потерпілих, що є важливим для забезпечення правопорядку та ефективного розслідування злочинів.

Державний реєстр бенефіціарів містить інформацію про фактичних власників юридичних осіб, включаючи їх імена, прізвища, країни громадянства та частки участі в капіталі. Цей реєстр сприяє забезпеченню прозорості ведення бізнесу та боротьбі з відмиванням грошей.

Державний реєстр ліцензій та сертифікатів об'єднує інформацію про всі видані в Україні ліцензії та сертифікати, що забезпечує регулювання різних видів діяльності, відомості про суб'єкти господарювання, їхні види діяльності, що є основою для контролю та нагляду держави за дотриманням законодавства.

Державний реєстр платників податків включає інформацію про всіх платників податків, зареєстрованих в країні, включаючи їхні ідентифікаційні номери, адреси, види діяльності та податкові зобов'язання. Цей реєстр є важливим для адміністрування податкових надходжень та забезпечення фінансової стабільності держави.

Єдиний державний реєстр територіальних громад відіграє ключову роль у забезпеченні інформації про адміністративно-територіальний устрій країни, включаючи дані про назви громад, їхні адміністративні центри, склад та керівництво. Він містить важливі дані про населення, бюджети та інші ключові показники діяльності громад.

Зазначені реєстри є невід'ємною частиною управлінської системи України, сприяючи підвищенню прозорості, ефективності та відповідальності у державному управлінні, що є важливим для забезпечення стійкого розвитку країни в умовах глобальних викликів і змін.

Кожен реєстр створений для зберігання значних об'ємів даних різних типів, від простих текстових даних (імена, адреси, дати) до повноцінних документів та їх змісту.

Якщо класифікувати дані з цих реєстрів по типам, то їх можна представити у вигляді таблиці 1.1.

Таблиця 1.1 – Класифікація даних в державних реєстрах

Назва бази даних	Тип даних	Формат даних	Опис
Єдиний державний реєстр фізичних та юридичних осіб	Персональні, адміністративні	Структуровані	Текстові дані про громадян, зображення.
Державний реєстр речових прав на нерухоме майно	Адміністративні	Структуровані	Документація, текстові дані про громадян.
Єдиний державний реєстр судових рішень	Адміністративні	Структуровані, неструктуровані	Документація.
Єдиний державний реєстр кримінальних проваджень	Адміністративні	Структуровані	Документація, текстові дані про громадян, зображення.
Єдиний державний реєстр бенефіціарів	Персональні, адміністративні	Структуровані	Текстові дані про громадян.
Державний реєстр ліцензій та сертифікатів	Адміністративні	Структуровані	Документація, зображення.
Державний реєстр платників податків	Персональні, адміністративні	Структуровані	Текстові дані про громадян.
Єдиний державний реєстр територіальних громад	Адміністративні	Структуровані	Текстові дані.

1.4 Огляд сучасних технологій збереження інформації

Сучасні технології збереження даних можна розділити на наступні типи:

Основні технології обробки інформації

В державних бази даних (БД) використовуються різні технології для обробки інформації, кожна з яких має свої сильні та слабкі сторони.

1. SQL (Structured Query Language) – мова програмування, що використовується для доступу, маніпулювання та обробки даних у реляційних БД. SQL є стандартом для реляційних БД і широко використовується в державному секторі;

2. NoSQL (Not Only SQL) – категорія БД, що не ґрунтується на реляційній моделі даних. NoSQL БД часто використовуються для зберігання та обробки неструктурованих або напівструктурованих даних, таких як текстові документи, зображення та відео;

3. hadoop – платформа для розподіленої обробки даних, що використовується для роботи з великими обсягами даних. Hadoop може бути використаний для аналізу даних, машинного навчання та інших задач;

4. spark – платформа для розподіленої обробки даних з відкритим кодом, що використовується для роботи з великими обсягами даних. Spark може бути використаний для аналізу даних, машинного навчання, потокової обробки даних та інших задач.

Архітектурні рішення

Архітектура БД визначає, як дані зберігаються, організовуються та доступні. Існує кілька типів архітектурних рішень для державних БД:

1. централізовані БД: Дані зберігаються на одному сервері. Цей тип архітектури простий у реалізації та керуванні, але може мати обмеження щодо масштабованості та продуктивності;

2. розподілені БД: Дані розподілені на кілька серверів. Цей тип архітектури може забезпечити кращу масштабованість та продуктивність, але може бути складнішим у реалізації та керуванні;

3. хмарні БД: Дані зберігаються в хмарі, що може забезпечити кращу масштабованість, гнучкість та доступність. Хмарні БД також можуть бути більш економічно вигідними, ніж традиційні БД.

Стандарти та протоколи обміну даними

Для обміну даними між державними органами та іншими зацікавленими сторонами використовуються різні стандарти та протоколи.

1. XML (Extensible Markup Language) – мова розмітки даних, що використовується для структурування та обміну даними. XML є популярним форматом для обміну даними в державному секторі;

2. JSON (JavaScript Object Notation) – легкий формат обміну даними, що використовується для структурування та обміну даними. JSON є більш компактним та простим у використанні, ніж XML;

3. EDI (Electronic Data Interchange) – стандарт для обміну електронними документами між підприємствами. EDI часто використовується для обміну даними між державними органами та постачальниками.

РОЗДІЛ 2 ПОТЕНЦІАЛ ІНТЕГРАЦІЇ АЛГОРИТМІВ ШТУЧНОГО ІНТЕЛЕКТУ В ДЕРЖАВНІ БАЗИ ДАНИХ

2.1 Перспективи розвитку методів обробки інформації в державних системах з використанням ШІ

Інтеграція штучного інтелекту може допомогти державним органам подолати проблеми, з якими вони стикаються при обробці даних, та покращити прийняття рішень, ефективність роботи та надання послуг населенню.

Існує ряд перспективних методів обробки інформації, які можуть бути використані в державних системах.

Машинне навчання. Машинне навчання може революціонізувати роботу урядових органів, завдяки можливості аналізувати великі обсяги даних з державних баз даних для покращення прийняття рішень, надання послуг та ефективності[16].

Використання ШІ для аналізу складних даних з різних джерел, таких як перепис населення, дані про злочинність та податкові декларації, дозволить урядам виявляти закономірності та тенденції, які важко побачити неозброєним оком. Це веде до кращого розуміння проблем та прийняття більш обґрунтованих рішень щодо політики та ресурсів.

Машинне навчання може автоматизувати завдання, такі як обробка заявок, виявлення шахрайства та надання персоналізованих послуг. Це звільняє час та ресурси для співробітників державних органів, зосереджуючи їх на більш складних завданнях[16].

Аналіз даних про громадян дозволяє персоналізувати послуги, які вони отримують. Наприклад, Машинне навчання може рекомендувати освітні програми, які відповідають потребам учня, або надавати медичні послуги, які враховують індивідуальні фактори ризику.

Саме на основі машинного навчання базуються наступні галузі обробки даних.

Обробка природної мови (ОНМ). ОНМ - це динамічно розвиваюча галузь штучного інтелекту, яка прагне подолати бар'єр між людською мовою та машинним розумінням. Її мета – наділити комп'ютери здатністю розуміти, аналізувати та генерувати мову так само, як це роблять люди.

В основі ОНМ лежать складні алгоритми, які здатні обробляти великі обсяги текстових даних, виявляючи закономірності, структуру та значення мови. Ці алгоритми навчаються на масивних наборах даних, що включають книги, статті, веб-сайти та розмови в соціальних мережах, щоб опанувати нюанси людської мови.

Хоч в публічних комерційних сервісах ОНМ набуло великого значення у якості голосових помічників, але в державних установах поки не використовується.

Не зважаючи на це, ОНМ має великий потенціал для інтеграції в державні установи для оптимізації збору даних, не маючи потреби в переривах, голосовий помічник на базі такого ШІ зможе цілодобово обслуговувати громадян задля збереження та занесення інформації в реєстри.

Візуальна обробка. Візуальна обробка - це галузь ШІ, яка дозволяє комп'ютерам розуміти та інтерпретувати візуальну інформацію. Вона може бути використана для автоматичного розпізнавання об'єктів, відстеження руху та аналізу зображень[7].

В основі візуальної обробки лежать алгоритми машинного навчання, які здатні виявляти об'єкти, категоризувати їх, відстежувати їхні рухи та розуміти просторові відносини між ними. Ці алгоритми навчаються на величезних наборах даних зображень та відео, щоб опанувати візуальні характеристики світу.

Візуальна обробка вже застосовується в деяких реєстрах для аналізу рис обличчя громадян і порівняння з зображенням на документах, але не використовується для обробки самих документів, що є потенційним напрямком.

Використання нових методів обробки інформації може мати ряд переваг для державних систем:

1. підвищення ефективності: Нові методи обробки інформації можуть допомогти державним органам автоматизувати рутинні завдання, що може звільнити час співробітників для більш важливої роботи. ІІІ не потребує переривів, відпусток, а саме тому працює постійно та виключно керуючись заданими інструкціями;

2. покращення прийняття рішень: Нові методи обробки інформації можуть допомогти державним органам аналізувати великі обсяги даних та виявляти закономірності, які неможливо побачити без ІІІ. Це може допомогти їм приймати кращі та більш обґрунтовані рішення;

3. підвищення якості послуг: Нові методи обробки інформації можуть допомогти державним органам надавати кращі послуги населенню. Наприклад, ОНМ може бути використано для автоматичного обслуговування клієнтів, а візуальна обробка може бути використано для виявлення підроблених документів;

4. підвищення прозорості: Нові методи обробки інформації можуть допомогти державним органам зробити свою роботу більш прозорою. ІІІ неможливо підкупити, він буде діяти виключно за заданими інструкціями, що може забезпечити систематичну звітність для .

Україна має значний потенціал для використання нових методів обробки інформації в державних системах. Верховна Рада вже розробила стратегію щодо розвитку цифрової економіки та електронного урядування, вона визначає важливість ІІІ для покращення державних послуг[1].

Однак, Україні необхідно подолати ряд викликів для того, щоб повною мірою використовувати переваги нових методів обробки інформації. Державі необхідно інвестувати в інфраструктуру, навчання персоналу та розробку подальших правових норм для використання ІІІ. Також важливо забезпечити довіру громадян до того, що їхні дані будуть використовуватися штучним інтелектом.

2.2 Аналіз використання штучного інтелекту для обробки даних у світі

Щоб проаналізувати ефективність використання штучного інтелекту для обробки даних потрібно дослідити наявні випадки його інтеграції:

1. Автоматизація рутинних текстових завдань:

Відомо багато випадків використання ШІ для автоматизації рутинних процесів, наведемо деякі приклади.

Наприклад, компанія Amazon використовує чат-бот «Q» на базі ШІ, щоб відповідати на запитання клієнтів, надавати підтримку та створювати графіки.

Бот успішно використовується самою компанією для обробки та аналізу інформації, формування висновків з них та формувати звіти. Для особистого використання Amazon надає готову внутрішню базу даних, до якої ШІ звертається задля формування потрібних форматів.

Для публічного використання чат боту «Q» надаються загальні дані з мережі інтернет.

Таким чином ШІ компанії Amazon значно оптимізує процеси обробки інформації, шляхом тривалого машинного навчання його «навчили» потрібним навичкам і тепер він здатний на аналіз великих об'ємів інформації, яку надає саме підприємство, після чого сам формує висновки, звіти, графіки, рекомендації[25]. Принцип роботи «Q» представлена на рисунку 2.1

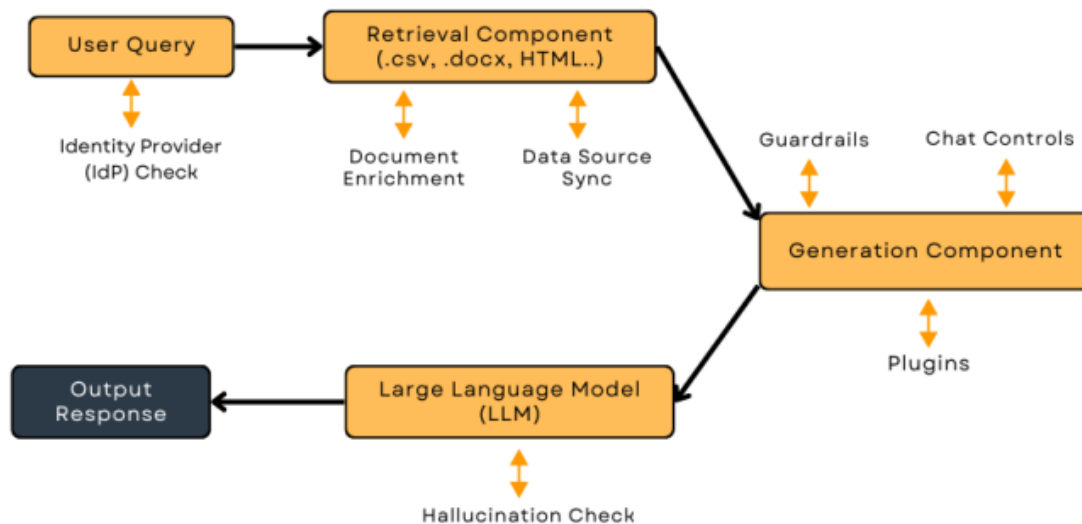


Рисунок 2.1 – Принцип роботи ШІ компанії Amazon

Далі проаналізуємо міжнародну платіжну систему PayPal, в середині компанії ШІ вона використовується для ведення бухгалтерського обліку, автоматичної обробки даних з рахунків, квитанцій та інших фінансових документів.

Для стандартного користувача платіжної системи ШІ оптимізує процеси платежів, за допомогою алгоритмів машинного навчання штучний інтелект може надавати рекомендації на основі попередніх покупок, також автоматично вибирає адресу доставки, сам вводить дані і формує квитанцію.

За допомогою алгоритмів ШІ компанія змогла автоматизувати обробку платежів, отримання даних з рахунків і створення платіжних квитанцій.

2. Комп'ютерний зір:

Візуальна обробка за допомогою ШІ набула великого попиту в підприємствах, загалом вона використовується для систем захисту, які за допомогою штучного інтелекту можуть визначати аналізувати рухи людини, її обличчя, але цей метод активно використовують компанії-виробники для перевірки якості товару та навіть для налагодження автоматизованого виробництва. Ось деякі із таких прикладів.

Компанія Zebra Medical Vision в 2014 створила однойменний ШІ для аналізу рентгенівських знімків та інших медичних зображень на предмет ознак захворювань[26].

За словами засновника компанії, Еяла Гури, алгоритм був розроблений на 350 000 мамограмах з підтвердженим діагнозом раку грудей і може виявляти рак грудей з ймовірністю 92%, у порівнянні з 87% для радіологів[26].

Ефективність ШІ підтверджується оскільки вже 50 медичних закладів в США, Австралії та Європі почали активно використовувати Zebra Medical Vision, подальше інтеграція штучного інтелекту обмежується лише кількістю серверів у компанії. Приклад роботи алгоритму можна побачити на рисунку 2.2.

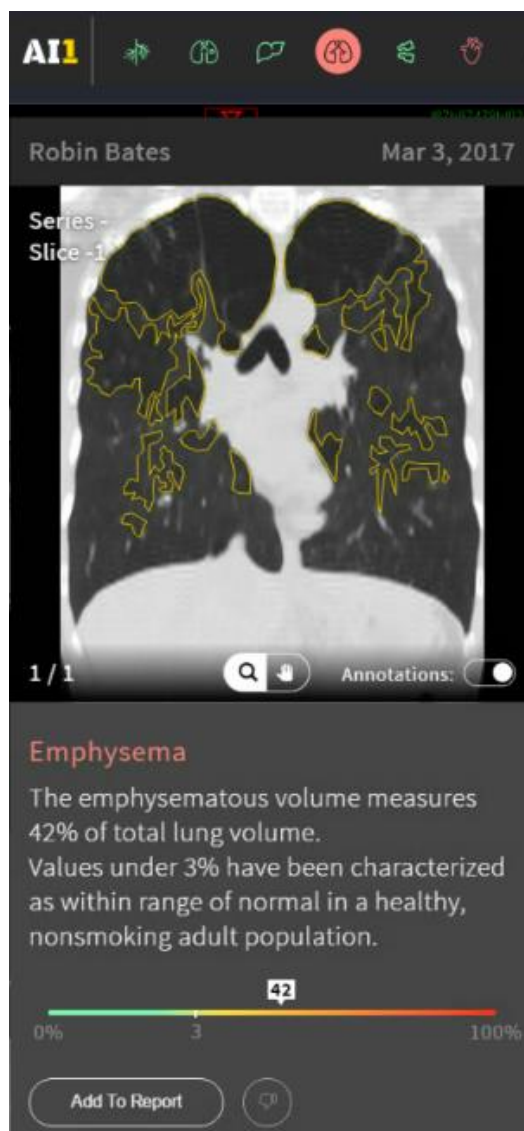


Рисунок 2.2 – Приклад роботи Zebra Medical Vision

Наступна компанія – Hikvision, вона змогла створити та інтегрувати алгоритми ШІ для розпізнання рухів та обличчя в свої системи захисту[24].

За допомогою налаштувань ШІ може розпізнавати людей, які вторгаються в приватну територію та зберігати зображення їх лиць. Також системи Hikvision використовують в Китаї для знаходження розшукуваних злочинців. Приклад роботи ШІ показаний на рисунку 2.3

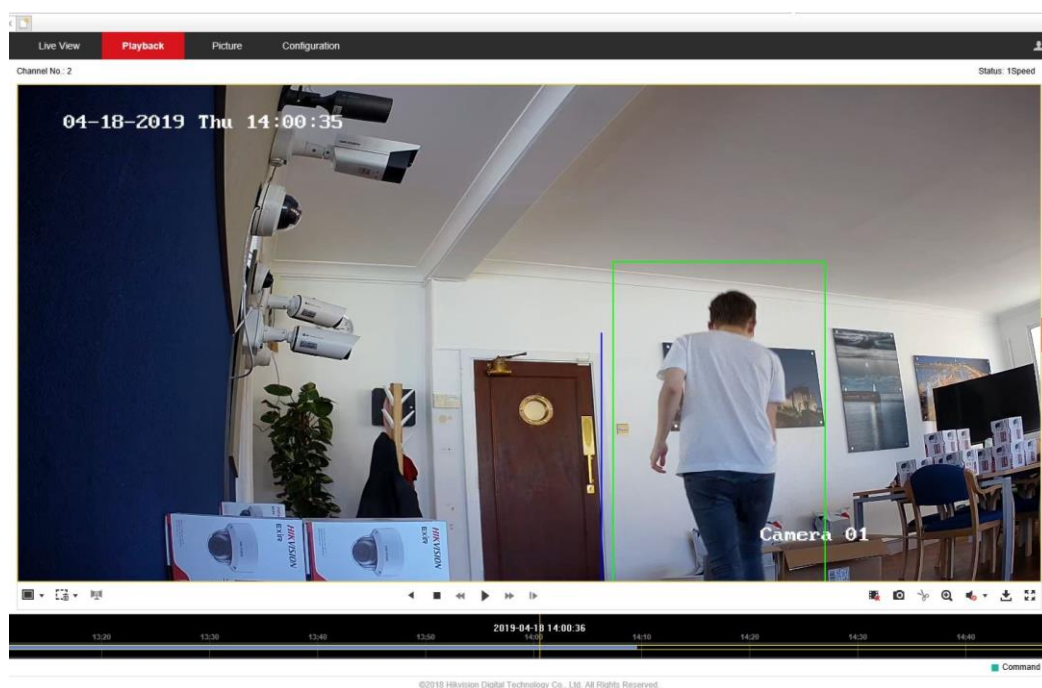


Рисунок 2.3 – Робота системи захисту з інтегрованим ШІ

3.Обробка природної мови (ОНМ):

ОНМ насамперед використовується для розуміння штучним інтелектом людської усної мови, такі алгоритми використовуються насамперед в голосових помічниках.

Alexa від компанії Amazon, популярний в США голосовий помічник, який взаємодіє з технологією розумного будинку та дуже точно сприймає людську мову[23]. Для отримання таких результатів був проведений тривалий процес

машинного навчання завдяки якому ШІ досягнув таких результатів у розумінні людей.

За допомогою алгоритмів штучного інтелекту та програмної взаємодії з системами розумного будинку, голосовий помічник може виконувати самі різноманітні задачі, від автоматичного здійснення онлайн замовлень до повного контролю над елементами дому, все це можливо контролювати використовуючи лише усну мову.

Алгоритм роботи Alexa представлений на рисунку 2.4

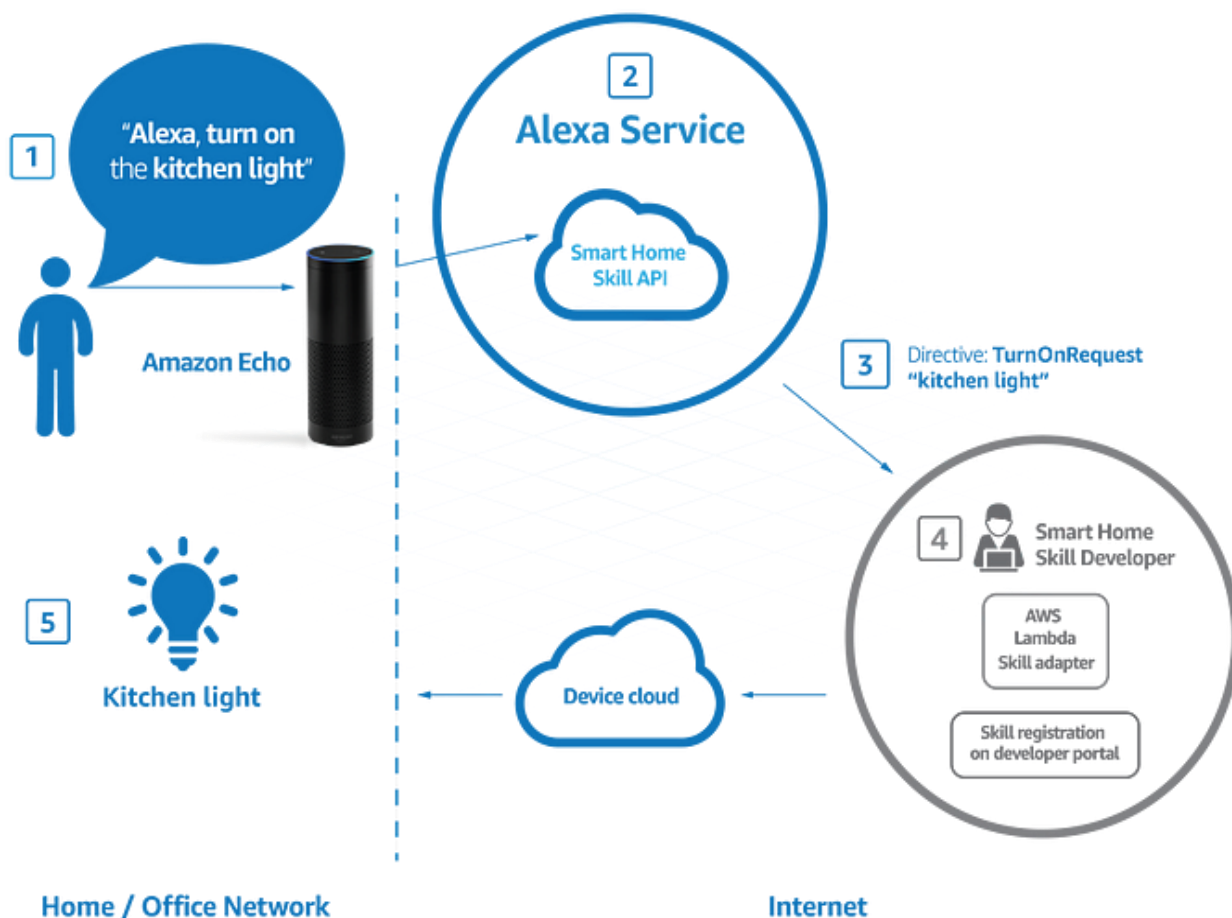


Рисунок 2.4 – Принцип роботи голосового помічника Alexa

Один із найвідоміших ШІ, це Google Translate від компанії Google, за допомогою машинного навчання штучний інтелект змогли «навчити» понад 100 мовам[22].

Хоч цей ШІ і використовується лише для перекладу, але він теж містить в собі складні алгоритми, які дозволяють йому працювати з усним мовленням багатьох мов Землі.

Алгоритм роботи ШІ від Google представлений на рисунку 2.5

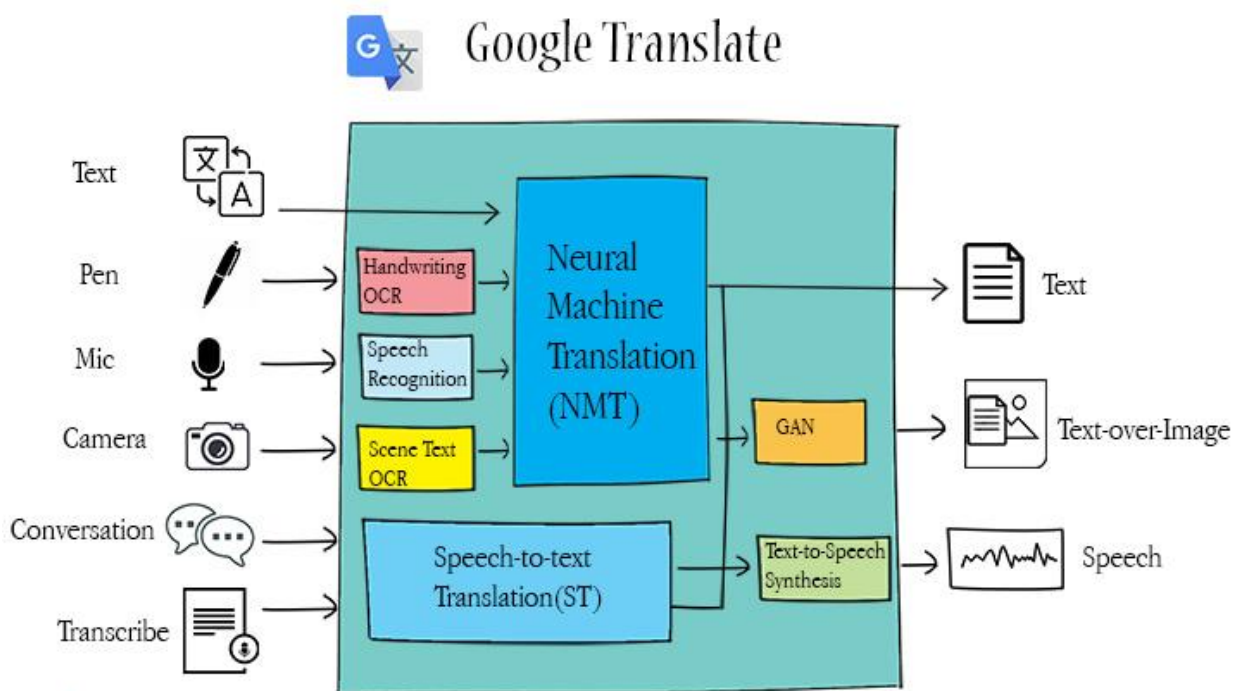


Рисунок 2.5 – Принцип роботи Google Translate

За результатами аналізу світової практики використання ШІ для різного роду завдань від роботи з документами, статистикою до розуміння людської мови та рис обличчя можна стверджувати, що штучний інтелект має великий потенціал в питаннях його інтеграції в державні бази даних.

На прикладі Amazon Q можна буде створити ШІ для роботи з документацією, який буде формувати потрібні державній установі звіти без участі співробітників. Алгоритми подібні тим, що використовує PayPal дадуть змогу спростити роботу реєстру платників податків шляхом глибокого аналізу платіжних даних. Аналоги комп'ютерного зору Hikvision та Zebra Medical Vision можна використати для перевірки документів на факт підробки та підтвердження особистості. Створення голосових помічників подібних Alexa дозволить

налагодити функціонування автономних робочих місць, на яких буде здійснюватись обробка запитів громадян шляхом обробки усної мови.

Розвиток сфери ШІ в державному секторі України потребує тісної співпраці між державними органами, науковими установами та ІТ-компаніями. За успішного впровадження ШІ Україна може стати лідером серед країн регіону щодо оптимізації обробки даних та використання цих даних для покращення добробуту громадян.

2.3 Приклади роботи штучного інтелекту в відкритих сервісах України

Хоч штучний інтелект і не набув широкого використання в державних реєстрах, але активно використовується різними сервісами та компаніями України.

Штучний інтелект (ШІ) застосовується у різних сферах економіки України, що відкриває нові можливості для підвищення ефективності, прозорості та інноваційності в кількох ключових галузях. Огляд практичного впровадження ШІ в Україні дозволяє виявити важливі тенденції, виклики та перспективи, пов'язані з інтеграцією цих технологій. Розглянемо найбільш значущі сектори, де ШІ вже знайшов застосування або показує потенціал для подальшого розвитку.

Для оптимізації судової системи України був створений ШІ у складі проекту "Суд на долоні" та сервісу WINCOURT[20]. Ці платформи використовують алгоритми машинного навчання для аналізу судових документів і прогнозування рішень. Вони забезпечують юристам можливість швидшого доступу до релевантних справ і допомагають аналізувати тенденції та прогнозувати результати розглядів, значно знижуючи час на підготовку та вивчення матеріалів. Це сприяє ефективності правосуддя і відкриває нові горизонти для автоматизації судових процедур.

Для аналізу державних закупівель також був створений ІІІ. Платформа DOZORRO використовує машинне навчання для моніторингу тендерів і виявлення корупційних ризиків[20]. Це дозволяє виявляти та вирішувати проблеми прозорості і корупції на ранніх стадіях, покращуючи ефективність використання державних ресурсів.

В медичній галузі, розробки Інституту кібернетики ім. В.М. Глушкова НАН України внесли значний вклад у покращення діагностики захворювань. Автоматичний аналіз медичних зображень за допомогою ІІІ дозволяє виявляти патології на ранніх стадіях, значно підвищуючи шанси на успішне лікування.

В сільському господарстві ІІІ забезпечує зменшення втрат урожаю та оптимізацію використання ресурсів. Компанія "Агрокедр" використовує алгоритми для аналізу зображень посівів, допомагаючи виявляти шкідників та хвороби. Це дозволяє фермерам своєчасно реагувати на загрози та застосовувати захисні заходи, збільшуючи продуктивність та знижуючи витрати.

У транспортній галузі, "Укрзалізниця" інтегрувала ІІІ для оптимізації розкладів та маршрутів поїздів. Ці заходи сприяють зменшенню часу поїздок та підвищенню комфорту пасажирів, роблячи залізничний транспорт більш привабливим для користувачів.

В освітній сфері ІІІ реалізован через платформу Prometheus, яка використовує персоналізацію для адаптації навчальних програм до індивідуальних потреб учнів. Це підвищує ефективність навчання та мотивацію студентів.

У фінансовій галузі, Ощадбанк впровадив чат-боти на базі ІІІ для спрощення банківських операцій та підвищення рівня обслуговування клієнтів. Це забезпечує швидкий доступ до послуг та зменшує навантаження на персонал.

Розвиток ІІІ в Україні супроводжується рядом викликів, включаючи питання захисту даних, етичні дилеми пов'язані з автоматизацією важливих рішень, та необхідність регулятивних змін для підтримки інновацій. Проте, потенціал для позитивного впливу на економіку і суспільство робить перспективу використання ІІІ в Україні особливо значущою. В майбутньому можна очікувати

подальших досліджень, інвестицій і поліпшення умов для ширшого впровадження ІІІ в усіх сферах діяльності.

2.4 Стандарти та протоколи використання ІІІ

Зростаюча інтеграція штучного інтелекту в державні бази даних несе в собі значні можливості для оптимізації роботи державних органів, покращення якості державних послуг та стимулювання економічного та соціального розвитку. Однак, як і будь-яка нова технологія, ІІІ має використовуватися відповідально та етично, з урахуванням ризиків та потенційних проблем.

Для того, щоб гарантувати належне використання ІІІ в державних базах даних, необхідні чітко визначені стандарти. Ці протоколи мають слугувати орієнтирами для розробників, впровадників та користувачів ІІІ систем, забезпечуючи:

1. прозорість: Громадськість повинна мати чітке уявлення про те, як використовуються ІІІ системи в державних базах даних, які дані використовуються, як приймаються рішення та на яких принципах ґрунтуються алгоритми;

2. підзвітність: Державні органи та розробники ІІІ систем повинні нести відповідальність за свої дії та бути готовими до того, щоб звітувати про те, як використовуються ІІІ системи;

3. відповідальність: ІІІ системи повинні розроблятися таким чином, щоб не завдавати шкоди людям, не дискримінувати їх та не порушувати їхні права та свободи;

4. ефективність: ІІІ системи повинні бути ефективними та дієвими, відповідати поставленим цілям та не створювати зайвих витрат або бюрократії;

5. сумісність: ІІІ системи повинні бути сумісними між собою та з іншими системами, що використовуються в державному секторі.

Наразі існує декілька міжнародних та національних стандартів та протоколів, які стосуються використання ШІ. Наведемо деякі з найважливіші.

Принципи ООН щодо штучного інтелекту: Ці принципи, розроблені Організацією Об'єднаних Націй, описують ключові етичні принципи, які повинні бути враховані при розробці та використанні ШІ-систем.

Керівні принципи ЄС щодо надійного ШІ: Ці керівні принципи, розроблені Європейською Комісією, описують ключові характеристики, яким повинні відповідати ШІ-системи, щоб вважатися надійними.

Національна стратегія розвитку штучного інтелекту в Україні: Ця стратегія, розроблена Кабінетом Міністрів України, описує цілі та пріоритети розвитку ШІ в Україні, а також містить рекомендації щодо використання ШІ в державному секторі.

Зростає кількість досліджень, присвячених питанням стандартів та протоколів використання ШІ в державних базах даних. Деякі з цих досліджень фокусуються на розробці загальних рамок для стандартів, тоді як інші досліджують більш конкретні аспекти, такі як прозорість, підзвітність, або безпека.

Одне з ключових досліджень, проведених Організацією економічного співробітництва та розвитку (ОЕСР), описує п'ять основних принципів, які повинні бути враховані при розробці стандартів для ШІ в державному секторі[30]. Ці принципи включають:

1. відповідність цінностям: ШІ-системи повинні розроблятися та використовуватися таким чином, щоб відповідати основним цінностям, таким як повага до людської гідності, недискримінація та верховенство права;

2. відповідальність: Державні органи та розробники ШІ-систем повинні нести відповідальність за свої дії та бути готовими до того, щоб звітувати про розробку та використання ШІ-систем;

3. прозорість: Громадськість повинна мати чітке уявлення про те, як використовуються ІІ-системи, які дані збираються та використовуються, та як приймаються рішення;

4. безпека, надійність та стійкість: ІІ-системи повинні бути розроблені та експлуатовані таким чином, щоб мінімізувати ризики кібератак, несанкціонованого доступу до даних та помилок у роботі системи;

5. підзвітність людини: Людина повинна завжди залишатися підзвітною за рішення, що приймаються за допомогою ІІ-систем.

Інше дослідження, проведене Всесвітнім форумом економіки, пропонує більш детальну рамку для стандартів ІІ, що охоплює такі аспекти, як управління даними, розробка та тестування алгоритмів, а також моніторинг та аудит ІІ-систем[29].

Дослідження також підкреслюють важливість міжнародного співробітства у розробці стандартів ІІ[29]. Оскільки технології штучного інтелекту швидко розвиваються, а державні бази даних часто містять інформацію, що стосується громадян різних країн, узгоджені на міжнародному рівні стандарти є необхідними для забезпечення послідовного та ефективного використання алгоритмів машинного навчання.

Для подолання цих викликів та забезпечення ефективного впровадження стандартів ІІ в державному секторі пропонуються такі рекомендації:

Інвестиції в дослідження та розробку: Необхідні інвестиції в дослідження та розробку стандартів ІІ, щоб враховувати складність технологій та швидкість їхнього розвитку.

Багатосторонній підхід: Розробку стандартів ІІ слід здійснювати за участю державних органів, приватного сектору, громадянського суспільства та наукової спільноти.

Гнучкість та адаптованість: Стандарти повинні бути гнучкими та адаптованими, щоб їх можна було оновлювати відповідно до розвитку технологій.

Міжнародне співробітництво: Необхідне міжнародне співробітництво для розробки узгоднених стандартів, які сприятимуть глобальному впровадженню відповідального ШІ.

Підвищення обізнаності: Необхідно підвищувати обізнаність щодо стандартів ШІ серед державних органів, розробників та громадськості.

Важливо, щоб ШІ використовувався відповідально та етично. Розробка та впровадження ефективних стандартів та протоколів є критично важливим кроком для досягнення цієї мети. За допомогою багатостороннього підходу, інвестицій у дослідження та розробку, а також міжнародного співробітництва держави можуть створити основу для відповідального використання ШІ в державному секторі.

2.5 Економічний та соціальний вплив інтеграції алгоритмів машинного навчання в державні бази даних, можливі недоліки

Інтеграція алгоритмів машинного може не лише покращити ефективність та якість державних послуг, але й мати значний економічний та соціальний вплив.

Економічний вплив

Одним із ключових економічних аспектів інтеграції алгоритмів машинного навчання в державні бази даних є зменшення витрат. Автоматизація рутинних завдань, таких як обробка заявок, введення даних, верифікація інформації, може призвести до значної економії коштів для країни. Це звільнить час персоналу для більш складних завдань, що може підвищити загальну продуктивність та якісну роботу.

Крім того, алгоритми машинного навчання можуть допомогти у оптимізації розподілу ресурсів, таких як бюджетні кошти. Це може призвести до більш ефективного використання цих ресурсів, що може зменшити витрати та покращити загальну ефективність розподілення бюджету для державних органів.

Інтеграція алгоритмів машинного навчання також може стимулювати інновації в державному секторі. Використання нових технологій може призвести до розробки нових продуктів та послуг, які можуть покращити життя громадян та збільшити економічне зростання.

Соціальний вплив

Інтеграція алгоритмів машинного навчання може мати значний позитивний соціальний вплив. Покращення якості та доступності державних послуг може призвести до збільшення задоволеності громадян та лояльності до влади. Це може зміцнити довіру до влади та згуртувати суспільство.

Крім того, алгоритми машинного навчання можуть допомогти у виявленні та вирішенні соціальних проблем, таких як шахрайство, нерівність та дискримінація.

Важливість етичного та відповідального використання

Важливо зазначити, що економічний та соціальний вплив інтеграції алгоритмів машинного навчання буде залежати від того, як ці алгоритми будуть розроблені, впроваджені та використані. Необхідно вжити заходів для мінімізації ризиків, таких як упередженість алгоритмів, втрата контролю над даними та зловживання технологіями[1].

Прозорість, етичність та відповідальність є ключовими принципами, які слід дотримуватися при використанні алгоритмів машинного навчання в державному секторі. Громадськість повинна бути проінформована про те, як використовуються ці алгоритми, і повинна мати можливість впливати на їх розробку та використання.

Додаткові переваги

Окрім економічних та соціальних переваг, інтеграція алгоритмів машинного навчання в державні бази даних може також мати й інші позитивні наслідки. Наприклад:

1. підвищення рівня безпеки: Алгоритми можуть допомогти виявити шахрайську діяльність, кіберзлочини та інші загрози безпеці державних систем;

2. покращення прийняття рішень: Алгоритми можуть аналізувати великі обсяги даних та допомагати державним органам приймати більш обґрунтовані рішення;

3.сприяння відкритому уряду: Алгоритми можуть допомогти зробити державні дані більш відкритими та доступними для громадськості, що може підвищити прозорість та підзвітність уряду.

Виклики та бар'єри

Незважаючи на значний потенціал, інтеграція алгоритмів машинного навчання в державні бази даних також пов'язана з певними викликами та бар'єрами. Деякі з них включають:

1. технічні складнощі: Впровадження та підтримка складних алгоритмів машинного навчання може потребувати значних інвестицій в інфраструктуру для підтримки функціонування ІІІ;

2. питання довіри та безпеки: Громадськість може мати побоювання щодо використання алгоритмів машинного навчання урядом, особливо стосовно питань конфіденційності даних та упередженості алгоритмів;

3.правові та етичні питання: Необхідно розробити чітке законодавство та етичні норми щодо використання алгоритмів машинного навчання в державному секторі.

Інтеграція алгоритмів машинного навчання в державні бази даних має потенціал значно покращити роботу державного сектору, стимулювати економічне зростання та сприяти розвитку більш справедливого та соціально орієнтованого суспільства. Однак для реалізації цього потенціалу державні органи повинні вжити заходів для подолання технічних, правових та етичних бар'єрів, а також повинні забезпечити прозорість, підзвітність та відповідальність при використанні цих потужних технологій.

РОЗДІЛ 3 РЕАЛІЗАЦІЯ ОБРОБКИ ІНФОРМАЦІЇ В БАЗІ ДАНИХ НА ПРИКЛАДІ БАЗИ ДАНИХ INFLUXDB І МОВИ PYTHON

3.1 Розробка архітектури алгоритму обробки даних

Для досягнення мети дослідження був розроблений алгоритм скану даних з паспорту громадянина України, алгоритм передбачає обробку зворотньої частини ID картки, де знаходиться MRZ код. Структура представлена на рисунках 3.1 та 3.2

```
import cv2
from PIL import Image
import pytesseract
import re
import numpy as np

pytesseract.pytesseract.tesseract_cmd = 'C:\\Program Files\\Tesseract-OCR\\tesseract.exe'

def preprocess_image(image):
    gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
    gray = cv2.equalizeHist(gray)
    sharpen_kernel = np.array([[ -1, -1, -1], [ -1, 9, -1], [ -1, -1, -1]])
    sharpen = cv2.filter2D(gray, -1, sharpen_kernel)
    thresh = cv2.adaptiveThreshold(sharpen, 255, cv2.ADAPTIVE_THRESH_GAUSSIAN_C, cv2.THRESH_BINARY, 11, 2)
    return thresh

def find_mrz_region(image):
    height, width = image.shape[:2]
    roi_height = int(height * 0.40)
    mrz_region = image[height - roi_height:height, :]
    return mrz_region

def parse_mrz(mrz_text):
    lines = mrz_text.strip().splitlines()
    if len(lines) < 3:
        return "Error"

    doc_type, country_code, doc_number = lines[0][:2], lines[0][2:5], lines[0][5:14]
    birth_date, gender, expiry_date = lines[1][:6], lines[1][7], lines[1][8:14]
    nationality, surname, given_names = lines[2][:3], lines[2][5:17].replace('<', ' '), lines[2][18:].replace('<', ' ').strip()

    return {
        'Document Type': doc_type,
        'Country Code': country_code,
        'Document Number': doc_number,
        'Birth Date': format_date(birth_date),
        'Gender': gender,
        'Expiry Date': format_date(expiry_date),
        'Nationality': nationality,
        'Surname': surname,
        'Given Names': given_names
    }
```

Рисунок 3.1 – Алгоритм обробки MRZ (частина 1)

```
def format_date(date_str):
    year, month, day = int(date_str[:2]), int(date_str[2:4]), int(date_str[4:6])
    return f"{day:02}/{month:02}/{'20' + str(year) if year < 30 else '19' + str(year)}"

def extract_mrz_data(image_path):
    image = cv2.imread(image_path)
    mrz_region = find_mrz_region(image)
    processed_image = preprocess_image(mrz_region)
    pil_image = Image.fromarray(processed_image)
    custom_config = r'--oem 3 --psm 6 -c tesseract_char_whitelist=ABCDEFGHIJKLMNOPQRSTUVWXYZ0123456789<'
    extracted_text = pytesseract.image_to_string(pil_image, config=custom_config)
    print(f"MRZ: {extracted_text}")
    mrz_data = parse_mrz(extracted_text)
    return mrz_data

image_path = 'C:\\passport_app\\uploaded_images\\UAID_Be.jpg'
extracted_mrz = extract_mrz_data(image_path)
print(extracted_mrz)
```

Рисунок 3.2 - Алгоритм обробки MRZ (частина 2)

Залежності та бібліотеки

Цей код покладається на кілька зовнішніх бібліотек Python для виконання своїх функцій:

OpenCV (cv2): OpenCV - це популярна бібліотека Python для обробки зображень та комп'ютерного зору. Він використовується в цьому коді для читання зображення паспорта, перетворення його в градації сірого, застосування розмиття та виконання адаптивної порогової обробки для підготовки області MRZ до розпізнавання тексту.

Pillow (PIL Fork) (Image): Pillow - це бібліотека Python для обробки зображень. У цьому коді вона використовується для перетворення обробленої області MRZ із масиву OpenCV NumPy в формат зображення PIL, необхідний для роботи pytesseract.

pytesseract: PyTesseract - це бібліотека Python із оберненням для оптичної програми розпізнавання символів Tesseract. Вона використовується для перетворення обробленого зображення MRZ на текст. Функція `pytesseract.image\_to\_string` застосовується для цієї мети, і додаткова конфігурація може бути налаштована для покращення точності розпізнавання документів MRZ (наприклад, `--oem 3 --psm 6` для використання моделі LSTM ocr та режиму PSM 6 для паспортів).

re: Регулярні вирази (re) - це стандартний модуль Python для роботи з рядками, що дозволяє виконувати пошук та заміну на основі шаблонів. Хоча в наданому коді вони явно не використовуються, регулярні вирази можуть бути корисними для більш надійного вилучення конкретних полів даних із рядків MRZ.

Основні функції

Цей код Python призначений для автоматичного вилучення інформації про власника з машиночитабельної зони (MRZ) зображення паспорта. MRZ - це стандартизований оптично зчитуваний блок, який містить ключові дані про власника паспорта у стислому вигляді.

1. «`preprocess_image`»: Функція `preprocess_image` виконує попередню обробку зображення паспорта, щоб покращити точність розпізнавання тексту. Вона перетворює зображення в градації сірого, розмиває його та застосовує адаптивну порогову обробку для виділення тексту на тлі;

2. «`find_mrz_region`»: Функція `find_mrz_region` знаходить та вирізає область MRZ зображення. Вона використовує розміри вхідного зображення для визначення приблизного розташування MRZ, зазвичай займаючи нижню частину паспорта;

3. «`parse_mrz`»: Функція `parse_mrz` аналізує розпізнаний текст MRZ, який зазвичай складається з трьох рядків. Кожен рядок містить закодовану інформацію про власника паспорта. Функція використовує регулярні вирази або інші методи для вилучення даних, таких як тип документа, код країни, номер документа, дата народження, стать, дата закінчення терміну дії, національність, прізвище та ім'я;

5. «`format_date`»: Функція `format_date` перетворює дати народження та дати закінчення терміну дії, отримані з MRZ, з формату YYMMDD на більш читабельний формат, наприклад DD/MM/YYYY;

6. «`extract_mrz_data`»: Функція `extract_mrz_data` об'єднує інші функції для виконання повного процесу розпізнавання даних MRZ. Вона читає зображення паспорта, обробляє його за допомогою `preprocess_image`, розпізнає текст MRZ за допомогою `pytesseract`, аналізує розпізнаний текст за допомогою `parse_mrz`, формує словник з даними та повертає його.

3.2 Обґрунтування ефективності інтеграції досліджуваних алгоритмів в державні бази даних

Алгоритм розпізнавання даних MRZ з зображення паспорта володіє значним потенціалом для використання в державних базах даних, адже він може суттєво покращити ефективність та безпеку різноманітних державних послуг. Його можливості можуть знайти застосування в таких сферах:

Автоматизація перевірки особистості: Алгоритм може автоматизувати процес верифікації особи громадян при оформленні внутрішніх паспортів, посвідок водія, інших документів, а також при проходженні контролю на кордоні. Це значно скоротить час очікування та зменшить навантаження на персонал, роблячи процес більш оперативним та зручним. Окрім того, його можна використовувати для підтвердження особистості при отриманні соціальних послуг, подачі заяв на візи, голосуванні, реєстрації в онлайн-сервісах державних органів.

Попередження шахрайства: Розпізнавання MRZ може допомогти виявити підроблені паспорти та інші документи, що посвідчують особу, шляхом порівняння даних з зображення з даними, що зберігаються в державних базах даних. Це допоможе запобігти шахрайству з документами, крадіжкам особистості, незаконній міграції та іншим злочинам, роблячи державні системи більш стійкими до шахрайських дій.

Покращення пошуку людей: Алгоритм може використовуватися для пошуку людей у державних базах даних за зображеннями їхніх паспортів. Це може бути корисно для правоохоронних органів при розслідуванні злочинів, для пошуку зниклих безвісти осіб, а також для надання допомоги людям, які втратили документи. Швидкий та точний пошук людей може допомогти врятувати життя, розкрити злочини та возз'єднати людей з близькими.

Створення єдиної системи ідентифікації: Розпізнавання MRZ може стати основою для створення єдиної системи ідентифікації громадян, яка поєднуватиме

дані з різних державних баз даних. Це може спростити доступ до державних послуг, зменшити адміністративний тягар та підвищити рівень безпеки. Завдяки єдиній системі ідентифікації громадяни зможуть отримувати послуги швидше та зручніше, не витрачаючи час на повторне надання своїх даних.

Забезпечення доступу до державних послуг онлайн: Алгоритм може використовуватися для верифікації особистості при доступі до державних послуг онлайн, таких як подання податкової декларації, запис на прийом до лікаря, реєстрація на транспорт. Це може зробити доступ до послуг більш зручним та доступним для громадян, адже їм не потрібно буде особисто відвідувати державні установи.

Важливо зазначити, що використання алгоритму розпізнавання даних MRZ в державних базах даних має супроводжуватися строгими заходами щодо захисту персональних даних та дотримання етичних норм. Держава повинна гарантувати, що дані громадян будуть використовуватися лише за законним призначенням, з повагою до їхньої приватності та свободи.

Загалом, алгоритм розпізнавання даних MRZ може стати цінним інструментом для покращення ефективності, безпеки та зручності державних послуг. Його впровадження може суттєво спростити життя громадян, зробити державні системи більш стійкими до шахрайства та покращити загальний рівень надання державних послуг.

ВИСНОВКИ

Для досягнення результату та підведенню підсумків в роботі були виконані наступні завдання:

- розглянута класифікація державних реєстрів та даних в них;
- досліджені процеси збору, обробки, збереження та передачі інформації в державних базах даних;
- визначені основні недоліки традиційного методу обробки інформації;
- розглянуті приклади використання штучного інтелекту для обробки різних типів даних;
- визначні позитивні сторони інтеграції штучного інтелекту в державні реєстри для використання в різних сферах діяльності;
- розроблений алгоритм зчитування машинного коду з ID-карток для швидкого отримання основної інформації про громадянина.

За результатами роботи можна стверджувати, що інтеграція штучного інтелекту в державні системи збереження інформації є оптимальним рішенням, яке призведе до скорочення витрат, автоматизації виконання рутинних завдань та в цілому підвищить ефективність функціонування баз даних і установ України.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Про схвалення Концепції розвитку штучного інтелекту в Україні : Розпорядж. Каб. Міністрів України від 02.12.2020 р. № 1556-р : станом на 29 груд. 2021 р. Режим доступу: <https://zakon.rada.gov.ua/laws/show/1556-2020-p#Text> (дата звернення: 01.03.2024).

2. Гугля В. Ю. Практика використання штучного інтелекту у сучасних умовах : thesis. 2017. Режим доступу: https://er.knutd.edu.ua/bitstream/123456789/10418/1/APPRMU2017_P185.pdf (дата звернення: 01.03.2024).

3. Совгира Т. Проблеми впровадження цифрових технологій («штучного інтелекту») у процесі створення роботизованого сценічного образу. Вісник Київського національного університету культури і мистецтв. Серія: сценічне мистецтво. 2021. Т. 4, № 1. С. 55–64. Режим доступу: <https://doi.org/10.31866/2616-759x.4.1.2021.234237> (дата звернення: 01.03.2024).

4. Технології штучного інтелекту в медичній практиці / О. П. Мінцер та ін. Medical informatics and engineering. 2020. № 2. С. 17–27. Режим доступу: <https://doi.org/10.11603/mie.1996-1960.2020.2.11171> (дата звернення: 01.03.2024).

5. Onyshchuk I. I. Правове регулювання технологій штучного інтелекту: теоретико-прикладні та етичні засади. Scientific papers of the legislation institute of the verkhovna rada of ukraine. 2020. № 3. С. 50–57. Режим доступу: <https://journals.indexcopernicus.com/api/file/viewByFileId/959219> (дата звернення: 01.03.2024).

6. Sabov I. Політичний аспект використання штучного інтелекту. Міжнародні відносини, суспільні комунікації та регіональні студії. 2018. № 1 (3). С. 77–80. Режим доступу: <https://doi.org/10.29038/2524-2679-2018-01-77-80> (дата звернення: 01.03.2024).

7. Tulashvili Y., Lukianchuk Y. Використання програмного забезпечення на основі штучного інтелекту для обробки зображень. Computer-integrated

technologies: education, science, production. 2021. № 43. С. 218–222. Режим доступу: <https://doi.org/10.36910/6775-2524-0560-2021-43-36> (дата звернення: 01.03.2024).

8. Андрушак В. С. Моделі управління потоками інфокомунікаційних мереж з використанням методів штучного інтелекту і машинного навчання : dissertation. 2021. Режим доступу: <https://ena.lpnu.ua/handle/ntb/56769> (дата звернення: 01.03.2024).

9. Бурдаєв В. П. Системи навчання з елементами штучного інтелекту : монографія. Харків : ХНЕУ, 2009. 392 с.

10. Козак Я. М., Kozak Y. Дослідження методів штучного інтелекту для оптимального вибору архітектурних шаблонів програмних систем : thesis. 2017. Режим доступу: <http://elartu.tntu.edu.ua/handle/123456789/19021> (дата звернення: 01.03.2024).

11. Пилипенко А. В. Застосування статистичних методів та методів штучного інтелекту для оптимізації процесу електронного навчання : магістерська робота. 2020. Режим доступу: <https://dspace.znu.edu.ua/jspui/handle/12345/5000> (дата звернення: 01.03.2024).

12. Пчелянський Д. П., Воїнова С. А. Штучний інтелект: перспективи та тенденції розвитку. Automation of technological and business processes. 2019. Т. 11, № 3. С. 59–64. Режим доступу: <https://doi.org/10.15673/atbp.v11i3.1500> (дата звернення: 01.03.2024).

13. Федоров Ф. Ф. Системи штучного інтелекту в економіці : thesis. 2016. Режим доступу: <http://dspace.zsea.edu.ua/jspui/handle/12345/8> (дата звернення: 01.03.2024).

14. Пацула В. І. Штучний інтелект : thesis. 2006. Режим доступу: <http://essuir.sumdu.edu.ua/handle/123456789/60273> (дата звернення: 01.03.2024).

15. Мороз О. Я. Штучний інтелект VERSUS людський інтелект: проблема комп'ютерного розуміння : thesis. 2009. Режим доступу: http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?I21DBN=LINK&P21DBN=UJRN&Z21ID=&S21REF=

10&S21CNR=20&S21STN=1&S21FMT=ASP_meta&C21COM=S&2_S21P03=FILE
=&2_S21STR=Pv_2014_72_4 (дата звернення: 01.03.2024).

16. Колесніков А. П., Карапетян О. М. Штучний інтелект: переваги та загрози використання. *Efektivna ekonomika*. 2023. № 8. Режим доступу: <https://doi.org/10.32702/2307-2105.2023.8.9> (дата звернення: 15.03.2024).

17. Квітка С., Новіченко Н., Бардах О. Штучний інтелект у муніципальному управлінні: вектори розвитку. *Public administration aspects*. 2021. Т. 9, № 4. С. 85–94. Режим доступу: <https://doi.org/10.15421/152140> (дата звернення: 16.03.2024).

18. Богом'я В., Гудзь А. Штучний інтелект: сучасний стан і перспективи застосування. Сучасні інформаційні технології у сфері безпеки та оборони. 2023. Т. 46, № 1. С. 13–17. Режим доступу: <https://doi.org/10.33099/2311-7249/2023-46-1-13-17> (дата звернення: 17.03.2024).

19. Дриньов Д. М., Войтех К. Р., Тимошенко Р. Р. Штучний інтелект в процесі прийняття та реалізації управлінських рішень. *Таврійський науковий вісник. серія: економіка*. 2023. № 18. С. 74–79. Режим доступу: <https://doi.org/10.32782/2708-0366/2023.18.7> (дата звернення: 17.03.2024).

20. Як штучний інтелект використовується у сфері відкритих даних. *Дія.Відкриті дані*. Режим доступу: <https://diia.data.gov.ua/info-center/aiod> (дата звернення: 17.03.2024).

21. Про систему Трембіта. *Трембіта*. Режим доступу: <https://trembita.gov.ua/about> (дата звернення: 17.03.2024).

22. Google translate architecture. *La Vivien Post*. Режим доступу: <https://www.lavivienpost.com/google-translate-and-transformer-model/> (дата звернення: 18.03.2024).

23. Gonfalonieri A. How amazon alexa works? Your guide to natural language processing (AI). *Medium*. Режим доступу: <https://towardsdatascience.com/how-amazon-alexa-works-your-guide-to-natural-language-processing-ai-7506004709d3> (дата звернення: 19.03.2024).

24. Hikvision company profile. *Hikvision*. Режим доступу: <https://www.hikvision.com/europe/about-us/company->

profile/#:~:text=Hikvision%20provides%20a%20broad%20range,safety%20management%20and%20business%20intelligence. (дата звернення: 19.03.2024).

25. Wiggers K. Amazon unveils Q, an AI-powered chatbot for businesses at AWS re:invent. TechCrunch. Режим доступу: <https://techcrunch.com/2023/11/28/amazon-unveils-q-an-ai-powered-chatbot-for-businesses/> (дата звернення: 19.03.2024).

26. Zebra medical vision unveils ai-based algorithm to detect brain bleeds. Healthcare IT News. Режим доступу: <https://hitconsultant.net/2018/03/05/zebra-medical-vision-unveils-ai-based-algorithm-detect-brain-bleeds/> (дата звернення: 19.03.2024).

27. Горбенко С. С., Мелешкевич Л. М. Штучний інтелект як технологія створення автоматизованих інтелектуальних систем : thesis. 2016. Режим доступу: <https://er.knutd.edu.ua/handle/123456789/5044> (дата звернення: 20.03.2024).

28. Дікарєв О. І., Барановська В. М. Концепт «штучний інтелект» у міжнародному та правовому дискурсах. Право.ua. 2024. № 1. С. 162–171. Режим доступу: <https://doi.org/10.32782/law.ua.2024.1.25> (дата звернення: 20.03.2024).

29. Artificial Intelligence: Opportunity, risk, and regulation in financial services - European Business Association. European Business Association. Режим доступу: <https://eba.com.ua/en/shtuchnyj-intelekt-mozhlyvosti-ryzyky-ta-regulyuvannya-finansovyh-poslug/> (дата звернення: 20.03.2024).

30. OECD AI Principles overview. OECD AI Policy Observatory Portal. Режим доступу: <https://oecd.ai/en/ai-principles> (дата звернення: 20.03.2024).

ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ
Кафедра комп'ютерної інженерії

ПРЕЗЕНТАЦІЯ ДО КВАЛІФІКАЦІЙНОЇ РОБОТИ

на тему

**«ОПТИМІЗАЦІЯ ПРОЦЕСУ ОБРОБКИ ДАНИХ В ДЕРЖАВНИХ БАЗАХ
ДАНИХ ЗА РАХУНОК ІНТЕГРАЦІЇ ШТУЧНОГО ІНТЕЛЕКТУ»**

Виконав:

Студент 4 курсу групи КІД-42

Приймак Євгеній Олександрович

Науковий керівник:

Куфтеріна Світлана Ростиславівна

Київ 2024

АКТУАЛЬНІСТЬ

2

Актуальність обраної теми проявляється у необхідності підвищення ефективності управління в умовах постійного зростання обсягів інформації та швидких змін у суспільстві. Технології, такі як штучний інтелект, можуть стати ключовими інструментами для оптимізації цих процесів та забезпечення ефективного функціонування державних реєстрів в питаннях збору та обробки інформації.

МЕТА, ОБ'ЄКТ, ПРЕДМЕТ

3

Мета дослідження: Аналіз сучасних підходів до обробки інформації в державних базах даних та вивчення можливостей інтеграції алгоритмів штучного інтелекту для покращення цих процесів.

Об'єкт дослідження: Процеси обробки інформації з використанням алгоритмів ШІ та традиційний метод обробки даних в державних реєстрах.

Предмет дослідження: Державні бази даних, алгоритми штучного інтелекту.

ЗАВДАННЯ ДОСЛІДЖЕННЯ

4

- провести аналіз наявних державних баз даних;
- провести аналіз процесу збору та обробки інформації в державні бази даних;
- визначити недоліки традиційної моделі обробки інформації;
- провести аналіз прикладів використання штучного інтелекту;
- визначити потенціал застосування штучного інтелекту в державних базах даних;
- створити приклад алгоритму штучного інтелекту для оптимізації обробки даних.

ДЛЯ ВИРІШЕННЯ ПОСТАВЛЕНИХ ЗАВДАНЬ В ТЕОРЕТИЧНІЙ ЧАСТИНІ РОБОТИ БУЛО:

5

- визначено структуру процесу обробки інформації в державних базах даних;
- описані приклади державних баз даних та класифікація інформації в них;
- описано успішні приклади використання штучного інтелекту;
- описано потенціал інтеграції штучного інтелекту;
- створено приклад алгоритму штучного інтелекту для обробки даних.

ПРИКЛАДИ ДЕРЖАВНИХ РЕЄСТРІВ ТА КЛАСИФІКАЦІЯ ДАНИХ В НИХ:

6

Назва бази даних	Тип даних	Формат даних	Опис
Єдиний державний реєстр фізичних та юридичних осіб	Персональні, адміністративні	Структуровані	Текстові дані про громадян, зображення.
Державний реєстр речових прав на нерухоме майно	Адміністративні	Структуровані	Документація, текстові дані про громадян.
Єдиний державний реєстр судових рішень	Адміністративні	Структуровані, неструктуровані	Документація.
Єдиний державний реєстр кримінальних проваджень	Адміністративні	Структуровані	Документація, текстові дані про громадян, зображення.
Єдиний державний реєстр бенефіціарів	Персональні, адміністративні	Структуровані	Текстові дані про громадян.
Державний реєстр ліцензій та сертифікатів	Адміністративні	Структуровані	Документація, зображення.
Державний реєстр платників податків	Персональні, адміністративні	Структуровані	Текстові дані про громадян.
Єдиний державний реєстр територіальних громад	Адміністративні	Структуровані	Текстові дані про громадян, зображення.

АНАЛІЗ ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ОБРОБКИ ДАНИХ У СВІТІ

7

Для визначення потенціалу інтеграції штучного інтелекту в державні бази даних, спочатку потрібно проаналізувати приклади використання ШІ для роботи з трьома стандартними типами даних.

Основні приклади використання наведені в таблиці 2:

Обробка текстових даних

- Amazon Q – аналіз даних компанії, формування звітів.
- PayPal – аналіз платіжних даних, формування квитанцій.

Обробка зображень

- Zebra Medical Vision – аналіз рентгенівських знімків.
- Hikvision – аналіз рухів об'єктів в реальному часі.

Обробка аудіо

- Google Translate – можливість аналізу усної мови.
- Alexa – можливість аналізу усної мови, звуків.

РЕАЛІЗАЦІЯ ПРИКЛАДУ АЛГОРИТМУ ОБРОБКИ ДАНИХ

8

Для реалізації алгоритму була обрана мова програмування Python, цей скрипт буде налаштований для обробки даних з ID-картки.

Принцип дії: за допомогою бібліотек візуальної обробки алгоритм здатен працювати із зображенням, далі йому вказується область роботи – це нижня частина ID-картки, в цій частині скрипт аналізує нанесений машинний код з якого і отримує основні дані про людину.

РЕАЛІЗАЦІЯ ПРИКЛАДУ АЛГОРИТМУ ОБРОБКИ ДАНИХ

Структура коду наведена на рисунка 1 та 2:

```
import cv2
from PIL import Image
import pytesseract
import re
import numpy as np

pytesseract.pytesseract_tesseract_cmd = 'C:\\Program Files\\Tesseract-OCR\\tesseract.exe'

def preprocess_image(image):
    gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
    gray = cv2.equalizeHist(gray)
    sharpen_kernel = np.array([[ -1, -1, -1], [-1, 9, -1], [-1, -1, -1]])
    sharpen = cv2.filter2D(gray, -1, sharpen_kernel)
    thresh = cv2.adaptiveThreshold(sharpen, 255, cv2.ADAPTIVE_THRESH_GAUSSIAN_C, cv2.THRESH_BINARY, 11, 2)
    return thresh

def find_mrz_region(image):
    height, width = image.shape[:2]
    roi_height = int(height * 0.40)
    mrz_region = image[height - roi_height:height, :]
    return mrz_region

def parse_mrz(mrz_text):
    lines = mrz_text.strip().splitlines()
    if len(lines) < 3:
        return "Error"

    doc_type, country_code, doc_number = lines[0][:2], lines[0][2:5], lines[0][5:14]
    birth_date, gender, expiry_date = lines[1][:6], lines[1][7], lines[1][8:14]
    nationality, surname, given_names = lines[2][:3], lines[2][5:17].replace('<', ' ').strip(), lines[2][18:].replace('<', ' ').strip()

    return {
        'Document Type': doc_type,
        'Country Code': country_code,
        'Document Number': doc_number,
        'Birth Date': format_date(birth_date),
        'Gender': gender,
        'Expiry Date': format_date(expiry_date),
        'Nationality': nationality,
        'Surname': surname,
        'Given Names': given_names
    }
```

```
def format_date(date_str):
    year, month, day = int(date_str[:2]), int(date_str[2:4]), int(date_str[4:6])
    return f"{day:02}/{month:02}/{'20' + str(year) if year < 30 else '19' + str(year)}"

def extract_mrz_data(image_path):
    image = cv2.imread(image_path)
    mrz_region = find_mrz_region(image)
    processed_image = preprocess_image(mrz_region)
    pil_image = Image.fromarray(processed_image)
    custom_config = r'--oem 3 --psm 6 -c tesseract_char_whitelist=ABCDEFGHIJKLMNOPQRSTUVWXYZ0123456789<'
    extracted_text = pytesseract.image_to_string(pil_image, config=custom_config)
    print(f"MRZ: {extracted_text}")
    mrz_data = parse_mrz(extracted_text)
    return mrz_data

image_path = 'C:\\passport_app\\uploaded_images\\\\UAID-Be.jpg'
extracted_mrz = extract_mrz_data(image_path)
print(extracted_mrz)
```


ВИСНОВКИ

11

Для досягнення результату в роботі були виконані наступні завдання:

- розглянута класифікація державних реєстрів та даних в них;
- досліджені процеси збору, обробки, збереження та передачі інформації в державних базах даних;
- визначені основні недоліки традиційного методу обробки інформації;
- розглянуті приклади використання штучного інтелекту для обробки різних типів даних;
- визначений потенціал інтеграції штучного інтелекту в державні реєстри для використання в різних сферах діяльності;
- розроблений алгоритм зчитування машинного коду з ID-карток для швидкого отримання основної інформації про громадянина.

1. Куфтеріна С. Р., Приймак Є. О. ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИРІШЕННЯ ПОМИЛОК В ДЕРЖАВНИХ БАЗАХ ДАНИХ. «Розвиток наук в умовах нової реальності: проблеми та перспективи» : II Міжнародна наукова конференція, м. Київ, 3 травня 2024 р. С. 139–140.

2. Куфтеріна С. Р., Приймак Є. О. ПОТЕНЦІАЛ ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ЗБІЛЬШЕННЯ ШВИДКОСТІ ТА ТОЧНОСТІ ОБРОБКИ ДАНИХ В ДЕРЖАВНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ. «Діджиталізація науки як виклик сьогодення» : VI Міжнародна студентська конференція, м. Суми, 3 травня 2024 р. С. 135–136.