

КАРГІН А.О., д.т.н, професор

(Український державний університет залізничного транспорту)

ТОРГОНСЬКИЙ А.В., PhD студент

(Український державний університет залізничного транспорту)



Щодо можливостей застосування штучного інтелекту для створення баз даних

Інформаційна технологія створення баз знань є частиною системи керування знаннями, яка дозволяє автоматизувати процеси збору, аналізу та управління знаннями. Ця технологія використовує сервіси штучного інтелекту (ШІ) для вибору релевантної інформації та її інтеграції в базу знань. Дослідження та розвиток технології ШІ з метою автоматизації створення баз даних є актуальним, так як спектр завдань що пов'язані зі створенням баз даних швидко поширюється у різних галузях. В роботі запропоновано метод переходу від традиційних способів збору даних до автоматизованих методів на основі ШІ. Проведено серію експериментів, з створення, нормалізації та наповнювання інформацією баз даних. Наведені результати порівняння трьох моделей штучного інтелекту з точки зору здатності створювати базу даних, нормалізувати її, знайти та заповнити базу даних, а також швидкість виконання запитів.

Ключові слова: бази даних, система керування знаннями, сервіси штучного інтелекту.

Вступ

Штучний інтелект (ШІ) швидко інтегрується в різні сфери діяльності, змінюючи їхні структури та функціональні можливості завдяки своїм здатностям аналізувати дані, автоматизувати процеси та приймати рішення. Від охорони здоров'я до державного управління, ШІ демонструє значний потенціал у підвищенні ефективності, точності та персоналізації послуг. У медичній сфері ШІ використовується для діагностики захворювань, персоналізованої медицини та роботизованої хірургії, що сприяє підвищенню рівня медичної допомоги та зниженню ризиків для пацієнтів [1]. В рітейлі та електронній комерції ШІ допомагає створювати персоналізовані рекомендації, оптимізувати управління запасами та підтримувати клієнтів за допомогою чат-ботів [2]. У сфері логістики та транспорту ШІ оптимізує маршрути, забезпечує безпеку безпілотних транспортних засобів та прогнозує попит [3].

Маркетингова індустрія використовує ШІ для аналізу даних, створення персоналізованого контенту та оптимізації рекламних кампаній [4]. У виробництві ШІ сприяє прогновному обслуговуванню, автоматизації процесів та контролю якості, що підвищує продуктивність і знижує витрати [5].

В освіті системи ШІ адаптують навчальні матеріали до потреб студентів, автоматизують оцінювання та забезпечують додаткову підтримку через віртуальних репетиторів [6]. У сфері розваг ШІ використовується для рекомендацій медіа-контенту, створення нового контенту та розвитку інтелектуальних ігор [7]. Нарешті, державне управління застосовує ШІ для аналізу даних, забезпечення безпеки та автоматизації адміністративних процесів, підвищуючи ефективність державних служб [8]. Сучасні роботи як універсальні машини, включаючи всі види автономних безпілотних машин [9], призначені для надання різноманітних послуг за допомогою ШІ. Щоб подолати проблему налаштування роботів на певні послуги, які зазвичай досягаються шляхом навчання, і консолідувати зусилля роботизованих лабораторій по всьому світу, було запущено проєкт RT-X [10]. Мета проєкту — створити загальний робототехнічний мозок (General Robots Brain, GBR) шляхом збору даних, ресурсів і коду, що стосується навичок, яким роботи вже навчені [11]. GenAI відіграє провідну роль, так як збір й створення знань про семантичні зв'язки між об'єктами в оточенні роботів покладають на AI.

Зважаючи на широкий спектр завдань у різних галузях що пов'язані зі створенням та підтримкою баз даних є актуальним дослідження та розвиток технології ШІ з метою автоматизації створення баз даних [12-14].

Основна мета даної статті є аналіз можливостей застосування технологій ШІ для автоматизації процесів створення баз даних, визначення їх переваг, а також розробка методу створення баз знань за допомогою ШІ.

Обговорення проблеми

Для пояснення основних мотивів створення баз даних, а саме - технологія створення баз знань за допомогою сервісів штучного інтелекту, можна розглянути просту задачу з створення бази даних, та її наповнення, з тривалість життя у різних регіонах планети.

Процес проектування та створення баз даних (БД) на сьогоднішній день є складним і багатоступеневим, вимагаючи використання сучасних методів, інструментів та технологій. Основні етапи цього процесу включають [15]:

1) Збір вимог. На початковому етапі визначаються вимоги користувачів та бізнесу до БД. Це включає проведення інтерв'ю з користувачами, аналіз бізнес-процесів та документів, а також чітке формулювання цілей і задач, які повинна вирішувати БД. 2) Концептуальне проектування. Створюються ER-діаграми (Entity-Relationship diagrams), що відображають об'єкти (сутності) та їх взаємозв'язки, а також визначаються властивості кожної сутності та взаємозв'язків між ними. 3) Логічне проектування. На цьому етапі відбувається нормалізація даних для усунення надлишковості та забезпечення цілісності даних. Логічне моделювання бази даних включає створення схем таблиць, первинних і зовнішніх ключів. 4) Фізичне проектування. Вибір системи управління базами даних (СУБД), наприклад, MySQL, PostgreSQL, Oracle або Microsoft SQL Server, є ключовим аспектом фізичного проектування. Крім того, здійснюється оптимізація фізичної структури БД для забезпечення ефективності та швидкодії, налаштовуються індекси, визначаються партиції та інші параметри. 5) Впровадження та тестування. На цьому етапі відбувається міграція існуючих даних у нову БД, якщо це необхідно. Перевірка БД на наявність помилок та відповідність вимогам включає функціональне тестування, тестування продуктивності та безпеки. 6) Експлуатація та підтримка. Постійний моніторинг продуктивності та безпеки БД, а також регулярне обслуговування, включаючи резервне копіювання, оновлення та оптимізацію БД, є важливими аспектами цього етапу. Також забезпечується безпека даних через аудит доступу та управління правами користувачів.

Проектування бази даних є складним і багатогранним процесом, під час якого можна зіткнутися з багатьма проблемами. Незрозумілі або

неповні вимоги від користувачів можуть призвести до проектування бази даних, яка не відповідає реальним потребам, або з часом вимоги користувача можуть змінюватись. Погана нормалізація, надмірна нормалізація ускладнює запити і знижує продуктивність через численні з'єднання, низька нормалізація призводить до надмірності даних та аномалій при оновленні. Неправильне визначення сутностей, їх атрибутів і зв'язків між ними може спричинити серйозні проблеми в майбутньому. Зважаючи на усі ці обставини, помилка може коштувати дуже багато часу, і в залежності від того на якому етапі виникла помилка тим більше часу потрібно на її виправлення. Не зважаючи вже на те що інколи помилки можуть бути настільки критичними, а бо вимоги можуть змінитись настільки що простіше перепроєктувати базу даних з нуля чим виправляти те що є.

Технологія створення баз знань за допомогою сервісів штучного інтелекту. Виконавши аналіз вищезазначених підходів, і врахувавши їх недоліки можна створювати бази даних, а саме метод створення баз знань за допомогою сервісів штучного інтелекту [12, 15-17].

Це означає створити технологію яка зможе автоматично виконувати певні етапи проектування бази даних, виконати концептуальне проектування бази даних, де вирішити які об'єкти будуть створенні, та які зв'язки будуть між цими об'єктами. Логічне проектування, де буде виконана нормалізація бази даних, фізичне проектування де буде написано SQL запит на створення таблиць, впровадження цієї бази, та експлуатація – система зможе сама оновлювати дані, а також переробляти структуру таблиць при необхідності [18].

Необхідно дослідити сучасні системи штучного інтелекту на здатність розуміти запити, змогу створювати SQL запит на створення бази даних та таблиць до цієї бази даних, а також здатність існуючих систем ШІ знаходити інформацію та заповнювати таблиці у базі даних.

Для вирішення задачі пропонується дослідити такі системи штучного інтелекту як Chagt GPT 3.5, Microsoft Copilot (Microsoft Bing Chat), Opera Aria [12, 17]. Пропонується створити базу даних з заздалегідь відомими полями, це буде база даних у якій буде зберігатись середня тривалість життя людини у різних країнах. Критеріями оцінки буде: здатність ШІ створювати базу даних та таблиці до цієї бази даних, здатність ШІ нормалізувати базу даних, та наповнення інформацією бази даних, час виконання запитів, кількість додаткових запитів та

наявність API (Application Programming Interface). Для прозорого результату запити до штучного інтелекту будуть ставитись однакові. Запити які

будуть задані III поділені на 3 частини: Створення бази даних, нормалізація бази даних, та заповнення бази даних. Усі результати тестів зведені у таблиці 1.

Таблиця 1. – Результати тестування III

	ChatGPT 3.5	Opera Aria	Microsoft Copilot
Створення бази даних таблиць	+	+	+
Нормалізація 1NF	+	+	+
Нормалізація 2NF	-/+	-/+	+/-
Нормалізація 3NF	+	+	+
Час виконання запитів	~ 3 секунд	~1.5 хвилини	~3 хвилин
Додаткові запити	1	3	1
Наявність API	+	-	-

Експерименти із виявлення здатності створювати бази даних сучасних за допомогою III. Розглянемо систему штучного інтелекту ChatGPT 3.5 на здатність створювати бази даних. Для цього почнемо зі створення бази даних. Поставимо системі III запит: «Створи базу даних на мові T-SQL з полями: унікальне поле, країна, стать, рік, середня тривалість життя». III ChatGPT 3.5 впорався з цією задачею, та написав SQL-запит на мові T-SQL з цією базою даних, який можна побачити на рисунку 1.

```
-- Створення бази даних
CREATE DATABASE LifeExpectancyDB;
GO
-- Використання новоствореної бази даних
USE LifeExpectancyDB;
GO
-- Створення таблиці LifeExpectancy
CREATE TABLE LifeExpectancy (
    ID INT PRIMARY KEY IDENTITY(1,1), --
    Унікальне поле з автоматичним інкрементом
    Country NVARCHAR(100) NOT NULL, --
    Країна
    Gender NVARCHAR(10) NOT NULL, --
    Стать
    Year INT NOT NULL, -- Рік
    LifeExpectancy FLOAT NOT NULL --
    Середня тривалість життя
);
GO
```

Рисунок 1. – Створення бази даних III «ChatGPT 3.5»

Також ЗIII ChatGPT 3.5 надав опис полів, на рисунку 2.

- ID: Унікальне поле (індекс), яке автоматично збільшується при додаванні нових записів.
- Country: Поле для зберігання назви країни, тип NVARCHAR(100) дозволяє зберігати до 100 символів, включаючи юнікодні символи.
- Gender: Поле для зберігання статі, тип NVARCHAR(10) дозволяє зберігати до 10 символів.
- Year: Поле для зберігання року, тип INT.
- LifeExpectancy: Поле для зберігання середньої тривалості життя, тип FLOAT.

Рисунок 2. – Опис полів у базі даних сгенерованій III «ChatGPT 3.5»

Наступним кроком була нормалізація бази даних. На запит: «Нормалізуй створену базу даних » був результат відпрацювання запити на рисунку 3.

```
CREATE TABLE Countries (
    CountryID INT PRIMARY KEY IDENTITY(1,1),
    CountryName NVARCHAR(50) NOT NULL
);
CREATE TABLE Genders (
    GenderID INT PRIMARY KEY IDENTITY(1,1),
    GenderName NVARCHAR(10) NOT NULL
);
CREATE TABLE People (
    ID INT PRIMARY KEY IDENTITY(1,1),
    CountryID INT NOT NULL,
    GenderID INT NOT NULL,
    Year INT NOT NULL,
    LifeExpectancy FLOAT NOT NULL,
    FOREIGN KEY (CountryID) REFERENCES Countries(CountryID),
```

```
FOREIGN KEY (GenderID) REFERENCES
Genders(GenderID)
);
```

Рисунок 3. – Часткова нормалізація бази даних виконаний ІІІ «ChatGPT 3.5»

ChatGPT виконав запит на нормалізацію бази даних частково. Він зміг винести в окремі таблиці поля Country та Gender, але залишив поле Year у базі, що є помилкою. Після вказівки на цю помилку запитом «Поле Year теж можна винести у окрему таблицю», ChatGPT 3.5 повністю нормалізував базу даних і написав SQL-запит на створення такої бази даних на рисунку 4.

```
-- Створення бази даних
CREATE DATABASE LifeExpectancyDB;
GO
-- Використання новоствореної бази даних
USE LifeExpectancyDB;
GO
-- Створення таблиці країн
CREATE TABLE Country (
    CountryID INT PRIMARY KEY IDENTITY(1,1), --
    Унікальний ідентифікатор країни
    CountryName NVARCHAR(100) NOT NULL --
    Назва країни
);
GO
-- Створення таблиці статей
CREATE TABLE Gender (
    GenderID INT PRIMARY KEY IDENTITY(1,1), --
    Унікальний ідентифікатор статі
    GenderName NVARCHAR(10) NOT NULL --
    Назва статі (наприклад, "Чоловік", "Жінка")
);
GO
-- Створення таблиці років
CREATE TABLE Year (
    YearID INT PRIMARY KEY IDENTITY(1,1), --
    Унікальний ідентифікатор року
    YearValue INT NOT NULL -- Значення
    року
);
GO
-- Створення таблиці тривалості життя
CREATE TABLE LifeExpectancy (
    LifeExpectancyID INT PRIMARY KEY
    IDENTITY(1,1), -- Унікальне поле з автоматичним
    інкрементом
    CountryID INT NOT NULL, --
    Зовнішній ключ на таблицю Country
    GenderID INT NOT NULL, --
    Зовнішній ключ на таблицю Gender
```

```
YearID INT NOT NULL, --
Зовнішній ключ на таблицю Year
LifeExpectancy FLOAT NOT NULL, --
Середня тривалість життя
FOREIGN KEY (CountryID) REFERENCES
Country(CountryID),
FOREIGN KEY (GenderID) REFERENCES
Gender(GenderID),
FOREIGN KEY (YearID) REFERENCES
Year(YearID)
);
GO
```

Рисунок 4. – Повна нормалізація бази даних виконана ІІІ «ChatGPT 3.5»

Тепер база даних повністю нормалізована, наступна спроба була змусити систему ІІІ ChatGPT 3.5 знайти будь-яку інформацію та перетворити цю інформацію на SQL запит, який заповнить цю базу даних. Для цього виконано запит: «Найди дані для цієї таблиці в інтернеті та переформотуй ці дані у SQL запит який би заповнював цю таблицю». Результат такого запиту ChatGPT 3.5 на рисунку 5.

```
-- Вставляємо дані в таблицю Country
INSERT INTO Country (CountryName)
VALUES ('USA');
INSERT INTO Country (CountryName)
VALUES ('China');
INSERT INTO Country (CountryName)
VALUES ('India');
-- Вставляємо дані в таблицю Gender
INSERT INTO Gender (GenderName)
VALUES ('Male');
INSERT INTO Gender (GenderName)
VALUES ('Female');
-- Вставляємо дані в таблицю Year
INSERT INTO Year (YearValue) VALUES
(2020);
-- Вставляємо дані в таблицю
LifeExpectancy
INSERT INTO LifeExpectancy (CountryID,
GenderID, YearID, LifeExpectancy)
VALUES
((SELECT CountryID FROM Country
WHERE CountryName = 'USA'),
(SELECT GenderID FROM Gender WHERE
GenderName = 'Male'),
(SELECT YearID FROM Year WHERE
YearValue = 2020),
76.3),
((SELECT CountryID FROM Country
WHERE CountryName = 'USA'),
(SELECT GenderID FROM Gender WHERE
GenderName = 'Female'),
```

```

(SELECT YearID FROM Year WHERE
YearValue = 2020),
81.2),
((SELECT CountryID FROM Country
WHERE CountryName = 'China'),
(SELECT GenderID FROM Gender WHERE
GenderName = 'Male'),
(SELECT YearID FROM Year WHERE
YearValue = 2020),
72.4),
((SELECT CountryID FROM Country
WHERE CountryName = 'China'),
(SELECT GenderID FROM Gender WHERE
GenderName = 'Female'),
(SELECT YearID FROM Year WHERE
YearValue = 2020),
77.6),
((SELECT CountryID FROM Country
WHERE CountryName = 'India'),
(SELECT GenderID FROM Gender WHERE
GenderName = 'Male'),
(SELECT YearID FROM Year WHERE
YearValue = 2020),
68.9),
((SELECT CountryID FROM Country
WHERE CountryName = 'India'),
(SELECT GenderID FROM Gender WHERE
GenderName = 'Female'),
(SELECT YearID FROM Year WHERE
YearValue = 2020),
71.6);

```

Рисунок 5. – Заповнення бази даних за допомогою ШІ «ChatGPT 3.5»

В цілому ChatGPT 3.5 впорався з поставленими задачами, але ChatGPT 3.5 має доступ лише до внутрішніх баз даних, які були створені ще в 2022 році, тому ChatGPT 3.5 не може надати актуальну інформацію, також ChatGPT 3.5 не завжди знаходить інформацію, інколи на запит про пошук інформації видає відповідь: «На жаль, я не можу надати поточної статистики щодо тривалості життя у різних країнах, оскільки моя база даних не містить свіжої інформації. Однак, ви можете знайти цю інформацію в Інтернеті на сайтах статистичних агентств, таких як Всесвітня організація охорони здоров'я (ВООЗ), Центри контролю та профілактики захворювань (CDC) США, Європейська служба статистики (Eurostat) та інші. Також є різні дослідження та звіти, які аналізують тривалість життя в різних країнах та регіонах. Ці дані можуть бути корисними для порівняння тривалості життя в різних частинах світу.»

Усі SQL запити які згенерував ChatGPT 3.5 синтаксично були вірними, та при запуску цих

запитів на SQL сервері ніяких помилок не виникло. На рисунку 6 можна переглянути діаграму створеної бази даних.

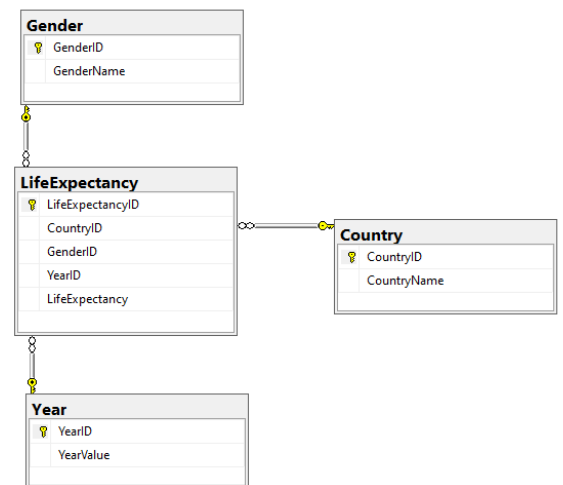


Рисунок 6. – Діаграма бази даних створенню за допомогою «ChatGPT 3.5»

Наступною системою ШІ яку пропонується розглянути буде система штучного інтелекту Microsoft Copilot (Microsoft Bing Chat). Було виконано всі запити які виконували для ChatGPT 3.5. Microsoft Copilot впорався з цією задачею, та видав відповідь у вигляді невеликого SQL запиту та пояснювання до полів, який можна побачити на рисунку 7.

```

CREATE TABLE Demographics (
  ID INT PRIMARY KEY IDENTITY(1,1),
  Country NVARCHAR(100),
  Gender NVARCHAR(10),
  Year INT,
  AverageLifeExpectancy FLOAT
);

```

Цей код створює таблицю Demographics з наступними полями:

ID: унікальне поле, яке автоматично збільшується на 1 з кожним новим записом.

Country: поле для зберігання назви країни.

Gender: поле для зберігання статі.

Year: поле для зберігання року.

AverageLifeExpectancy: поле для зберігання середньої тривалості життя.

Рисунок 7. – Створення бази даних, та опис полів ШІ «Microsoft Copilot»

Наступним кроком для Microsoft Copilot став нормалізація бази даних. Виконано запит «Нормалізуй створену базу даних базу даних». Microsoft Copilot впорався з цією задачею та надав відповідь у вигляді SQL запиту на створення нормалізованої бази даних, зображеному на рисунку 8.

```
CREATE TABLE Countries (
    CountryID INT PRIMARY KEY
    IDENTITY(1,1),
    Country NVARCHAR(100)
);

CREATE TABLE Genders (
    GenderID INT PRIMARY KEY
    IDENTITY(1,1),
    Gender NVARCHAR(10)
);

CREATE TABLE Demographics (
    ID INT PRIMARY KEY IDENTITY(1,1),
    CountryID INT FOREIGN KEY
    REFERENCES Countries(CountryID),
    GenderID INT FOREIGN KEY
    REFERENCES Genders(GenderID),
    Year INT,
    AverageLifeExpectancy FLOAT
);
```

Рисунок 8. – Нормалізація бази даних виконана ШІ «Microsoft Copilot»

Також була спроба знайти інформацію для заповнення цієї бази даних та перефармувати її у SQL запит щоб заповнити цими даними таблиці виконавши запит: «Найди данні для цієї таблиці в інтернеті та переформотуй ці дані у SQL запит який би заповнював цю таблицю». У відповідь на цей запит Microsoft Copilot надає гіпотетичні дані (іншими словами він їх вигидав). Відповідь можливо побачити на рисунку 9.

Вибачте, але я не можу знайти конкретні дані в Інтернеті та перетворити їх у SQL-запити. Однак я можу показати вам, як ви можете створити SQL-запити для заповнення таблиці на основі прикладу даних. Припустимо, у вас є наступні дані:
 Країна: Україна
 Стать: Чоловіча
 Рік: 2024
 Середня тривалість життя: 72 роки
 Ви можете використати наступні SQL-запити для заповнення таблиці:

```
INSERT INTO Countries (Country) VALUES
('Україна');
INSERT INTO Genders (Gender) VALUES
('Чоловіча');
INSERT INTO Demographics (CountryID, GenderID,
Year, AverageLifeExpectancy) VALUES ((SELECT
CountryID FROM Countries WHERE Country =
'Україна'), (SELECT GenderID FROM Genders
WHERE Gender = 'Чоловіча'), 2024, 72);
```

Рисунок 9. – Заповнення бази даних ШІ «Microsoft Copilot»

Microsoft Copilot впорався з поставленими задачами частково, усі SQL запити робочі і виконуються без помилок, але не зміг знайти дані у відкритому доступі. Також Microsoft Copilot відзначився дуже повільною роботою, генерація відповідей може займати близько 5ти хвилин. На рисунку 10 можливо побачити діаграму бази даних.

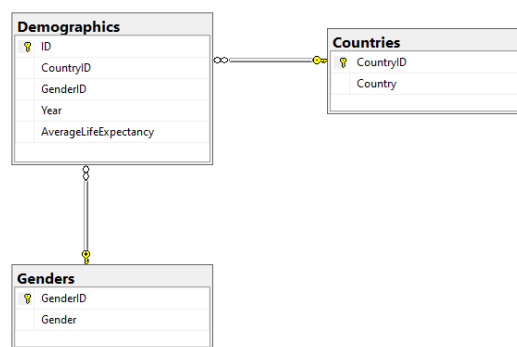


Рисунок 10. - Діаграма бази даних створенню за допомогою Microsoft Copilot

Останньою системою ШІ була Opera Aria. При запиті «Створи базу даних на мові T-SQL з полями: унікальне поле, країна, стать, рік, середня тривалість життя» Opera Aria генерує SQL запит і дає невелике пояснення що до полів в таблиці, відповідь можна побачити на рисунку 11.

```
CREATE TABLE Демографічні_дані (
    ID INT PRIMARY KEY IDENTITY(1,1),
    Країна NVARCHAR(50),
    Стать NVARCHAR(10),
    Рік INT,
    Середня_тривалість_життя FLOAT
);
```

У цій таблиці ми маємо наступні поля:
 ID: унікальний ідентифікатор запису
 Країна: назва країни
 Стать: стать осіб (чоловіки, жінки тощо)
 Рік: рік, до якого відносяться демографічні

дані

Середня_тривалість_життя: середня тривалість життя населення у вказаному році

Рисунок 11. – Створення та опис полів згенеровані ШІ «Opera Aria»

Наступним запитом був нормалізувати базу даних. Результат роботи системи ШІ можна побачити на рисунку 12.

```
CREATE TABLE Країни (
  ID INT PRIMARY KEY IDENTITY(1,1),
  Назва NVARCHAR(50) UNIQUE
);
CREATE TABLE Стати (
  ID INT PRIMARY KEY IDENTITY(1,1),
  Назва NVARCHAR(10) UNIQUE
);
CREATE TABLE Демографічні_дані (
  ID INT PRIMARY KEY IDENTITY(1,1),
  Країна_ID INT,
  Стать_ID INT,
  Рік INT,
  Середня_тривалість_життя FLOAT,
  FOREIGN KEY (Країна_ID)
REFERENCES Країни(ID),
FOREIGN KEY (Стать_ID) REFERENCES
Стати(ID)
);
```

Рисунок 12. – Нормалізація бази даних виконана ШІ «Opera Aria»

Opera Aria також впорався з цією задачею. На запит: «Найди данні для цієї таблиці в інтернеті та переформатуй ці дані у SQL запит який би заповнював цю таблицю» ШІ Opera Aria відповів таким чином: «На жаль, я не можу знайти дані для цієї таблиці в реальному часі. Будь ласка, спробуйте переформулювати своє запитання або надати більше деталей, щоб я міг допомогти вам.»

Запит іншим чином був переформульован: «Надай інформацію про середню тривалість життя у будь-якій країні та переформатуй цю інформацію у SQL запит який буде заповнювати базу даних яку ти щойно створив». Opera Aria зміг сформулювати SQL запит, рисунок 13.

```
INSERT INTO Країни (Назва) VALUES ('Монако');
INSERT INTO Країни (Назва) VALUES ('Сінгапур');
INSERT INTO Країни (Назва) VALUES ('Макао
(Китай)');
INSERT INTO Країни (Назва) VALUES ('Японія');
```

```
INSERT INTO Демографічні_дані (Країна_ID, Стать,
Рік, Середня_тривалість_життя) VALUES
```

```
((SELECT ID FROM Країни WHERE Назва =
'Монако'), 'Чоловіки', 2024, 89.52),
((SELECT ID FROM Країни WHERE Назва =
'Монако'), 'Жінки', 2024, 89.52),
((SELECT ID FROM Країни WHERE Назва =
'Сінгапур'), 'Чоловіки', 2024, 86.35),
((SELECT ID FROM Країни WHERE Назва =
'Сінгапур'), 'Жінки', 2024, 86.35),
((SELECT ID FROM Країни WHERE Назва =
'Макао (Китай)'), 'Чоловіки', 2024, 84.98),
((SELECT ID FROM Країни WHERE Назва =
'Макао (Китай)'), 'Жінки', 2024, 84.98),
((SELECT ID FROM Країни WHERE Назва =
'Японія'), 'Чоловіки', 2024, 84.83),
((SELECT ID FROM Країни WHERE Назва =
'Японія'), 'Жінки', 2024, 84.83);
```

Рисунок 13. – Заповнення бази даних за допомогою ШІ «Opera Aria»

Але цей запит не вірний. У згенерованому запиті передбачається що в базі даних є 2 таблиці: таблиця «Країни» та таблиця «Демографічні_дані», але кроком вище Opera Aria створив базу даних із 3-х таблиць, а саме: «Країни», «Стать», «Демографічні_дані». Тому згенерований SQL запит не відповідає структурі бази даних, та не може бути виконаний. З рештою через декілька ітерацій та підказок Opera Aria зміг згенерувати запит на заповнення бази даних, на рисунку 14

```
INSERT INTO Країни (Назва) VALUES ('Україна');
INSERT INTO Стати (Назва) VALUES ('Чоловіки'),
('Жінки');
INSERT INTO Демографічні_дані (Країна_ID,
Стать_ID, Рік, Середня_тривалість_життя) VALUES
((SELECT ID FROM Країни WHERE Назва =
'Україна'), (SELECT ID FROM Стати WHERE Назва
= 'Чоловіки'), 2021, 68.2),
((SELECT ID FROM Країни WHERE Назва =
'Україна'), (SELECT ID FROM Стати WHERE Назва
= 'Жінки'), 2021, 75.1);
```

Рисунок 14. – Виправлений запит на заповнення бази даних згенерований ШІ «Opera Aria»

Приведенні SQL запити являються вірними і виконались на SQL сервері без помилок. Однак у Opera Aria не завжди розуміє що потрібно зробити і потребує декілька уточнюючих запитів. Також має проблеми продуктивністю і довго обробляє запити.

Висновки

Основоючись на отриманих даних, усі вказані системи штучного інтелекту з рештою впорались з поставленими задачами. ChatGPT 3.5 показав себе як найшвидша система ШІ, достатньо

точними виявились системи ChatGPT 3.5 та Microsoft Copilot. Однак системи III Microsoft Copilot та Opera Aria виявились занадто повільними. Запити до цих систем можуть довго оброблятися (~ 5 хвилин). Також Opera Aria не завжди розуміє що від нього потребують, і не може зразу дати відповідь на питання, а через декілька уточнюючих запитів. Враховуючи ці факти, а також те що ChatGPT від компанії OpenAI має свій API цнй III має переваги для створень баз даних.

Планується перевірити ChatGPT на можливість створення спеціалізованих структур знань для FAI (Feeling Artificial Intelligence). ChatGPT версії 4.0o як більш досконала версія ChatGPT 3.5 що має можливість працювати з файлами та дозволяє генерувати SQL запити на створення баз даних буде протестовано щодо створення нових моделей III [19,20].

СПИСОК ЛІТЕРАТУРИ

1. Topol, E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again / E. Topol. – New York: Basic Books, Mar. 2019. – pp.75–120 ISBN-13 : 978-1541644632
2. Davenport, T. H., Kalakota, R. "The potential for artificial intelligence in healthcare" / T. H. Davenport, R. Kalakota. – Future Healthcare Journal, 6(2), 2019. – pp. 94-98. doi: 10.7861/futurehosp.6-2-94
3. WEF. The Future of Jobs Report 2018 / World Economic Forum. – Geneva: World Economic Forum, 2018. – pp. 5-6.
4. Davenport, T., Guha, A. The AI Advantage: How to Put the Artificial Intelligence Revolution to Work / T. Davenport, A. Guha. – Cambridge: MIT Press, 2018. – pp.23–39. ISBN 9780262039178
5. Bessen, J. E. "AI and Jobs: The Role of Demand" / J. E. Bessen. – NBER Working Paper, No. 24235. – National Bureau of Economic Research, 2019. – pp. 14-15. doi 10.3386/w24235
6. Luckin, R., Holmes, W., Griffiths, M., Forcier, L. B. Intelligence Unleashed: An Argument for AI in Education / R. Luckin, W. Holmes, M. Griffiths, L. B. Forcier. – London: Pearson, 2016. – pp. 18 – 21. ISBN-13: 9780992424886
7. McKinsey. Notes from the AI frontier: Applications and value of deep learning / McKinsey Global Institute. – McKinsey & Company, 2018. – pp. 7 – 16
8. Eggers, W. D., Schatsky, D., Viechnicki, P. AI-augmented government / W. D. Eggers, D. Schatsky, P. Viechnicki. – Deloitte Insights, 2017. – pp. 7–14.
9. E. Guizzo, Types of Robots. Categories frequently used to classify robots. Robotsguide.com. <https://robotsguide.com/learn/types-of-robots> (accessed Jun. 15, 2024).
10. Open X-Embodiment: Robotic Learning Datasets and RT-X Models Open X-Embodiment Collaboration, 2024. [Online]. Available: <https://robotics-transformer-x.github.io/>
11. S. Levine, K. Hausman, "The global project to make a general robotic brain," IEEE Spectrum, Jan 2024. [Online]. Available: <https://spectrum.ieee.org/global-robotic-brain>
12. MIT researchers introduce generative AI for databases: <https://news.mit.edu/2024/mit-researchers-introduce-generative-ai-databases-0708> (accessed Nov. 20, 2024)
13. End-to-end tools to productionize AI and data science. Domo. <https://www.domo.com/data-integration/data-science> (accessed Nov. 20, 2024).
14. David Bunting. How to use GenAI for database query optimization and natural language analysis. <https://www.chaossearch.io/blog/genai-database-query-natural-language> (accessed Nov. 20, 2024).
15. "Database System Concepts" by Abraham Silberschatz, Henry Korth, S. Sudarshan / Abraham Silberschatz, Henry Korth, S. Sudarshan. 7th edition (February 19, 2019). pp. 241–453. ISBN 978-0078022159
16. GPT-4. OpenAI. <https://openai.com/index/gpt-4-research/> (accessed Nov. 20, 2024).
17. 11 Best Generative AI Tools and Platforms. <https://www.turing.com/resources/generative-ai-tools> (accessed Nov. 20, 2024).
18. Guiding Principles for Working with GenAI and SQL Data. <https://www.clearobject.com/genai-guiding-principles/> (accessed Nov. 20, 2024).
19. A. Kargin and T. Petrenko, "Feeling Artificial Intelligence for AI-Enabled Autonomous Systems," in Proc. of 2022 IEEE Global Conf. on Artificial Intelligence and Internet of Things (GCAIoT) Alamein New City, Egypt, Dec.18, 2022, pp.88-93.
20. A. Kargin and T. Petrenko, "Knowledge Distillation for Autonomous Intelligent Unmanned System," in W. Pedrycz, S.-M. Chen. Advancements in Knowledge Distillation: Towards New Horizons of

Intelligent Systems. Studies in Comput. Intell., vol. 1100. Springer International Publishing, 2023, pp. 193-230, doi: 10.1007/978-3-031-32095-8_7.

Kargin A.O., Torhonskyi A.V. Regarding the possibilities of using artificial intelligence to create databases

Information technology for creating knowledge bases is part of the knowledge management system, which allows you to automate the processes of knowledge collection, analysis, and management. This technology uses Artificial Intelligence (AI) services to select the most relevant information and integrate it into the knowledge bases. The study of the method of creating knowledge bases with the help of AI allows the development of a new knowledge management algorithm, which will significantly improve the efficiency and accuracy of the system. This paper proposes a new method of transition from traditional techniques of database creation and data collection to automated ones based on AI. The results of the comparison of three artificial intelligence models in terms of the ability to create a database, normalize it, find and fill the database, as well as the speed of query execution are presented.

Keywords: databases, knowledge management system, artificial intelligence services.

Каргін Анатолій Олексійович, доктор технічних наук, професор, завідувач кафедри інформаційних технологій, Український державний університет залізничного транспорту, Харків, Україна. E-mail: kargin@kart.edu.ua, ORCID ID <http://orcid.org/0000-0003-2885-9071>.

Торгонський Антон Вадимович, аспірант кафедри інформаційних технологій, Український державний університет залізничного транспорту, Харків, Україна. E-mail: torgonskiy@kart.edu.ua, ORCID ID <https://orcid.org/0009-0008-4309-8032>.

Anatolii Kargin, Doctor of Technical Science, professor, department of Information Technology, Ukrainian State University of Railway Transport, Kharkiv, Ukraine. E-mail: kargin@kart.edu.ua, ORCID ID <http://orcid.org/0000-0003-2885-9071>.

Torghonskyi Anton, PhD student, department of Information Technology, Ukrainian State University of Railway Transport, Kharkiv, Ukraine. E-mail: torgonskiy@kart.edu.ua, ORCID ID <https://orcid.org/0009-0008-4309-8032>.