

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
КАФЕДРА МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ ТА АНАЛІЗУ
ДАНИХ

«До захисту допущено»

В. о. завідувача кафедри

І. М. Терещенко

(підпис)

(ініціали, прізвище)

Дипломна робота
на здобуття ступеня бакалавра

зі спеціальності 113 «Прикладна математика»

(код і назва)

на тему: **Розробка голосового інтерфейсу для створення діаграм**

Виконав: студент 4 курсу, групи ФІ-92

Поночевний Назар Юрійович

(прізвище, ім'я, по батькові)

(підпис)

Науковий керівник д.т.н., професор Куссуль Н. М.

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

(підпис)

Консультант

(номер розділу)

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

(підпис)

Рецензент

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

(підпис)

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студент

(підпис)

Національний Технічний Університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Навчально-науковий Фізико-технічний інститут
Кафедра математичного моделювання та аналізу даних

Рівень вищої освіти — перший (бакалаврський)

Спеціальність (освітня програма) — 113 «Прикладна математика»

ОПП — Математичні методи моделювання, розпізнавання образів та комп'ютерного зору

«ЗАТВЕРДЖЕНО»

В. о. завідувача кафедри

І. М. Терещенко

(підпис)

(ініціали, прізвище)

« ___ » _____ 2023 р.

ЗАВДАННЯ

на бакалаврську роботу студенту

Поночевному Назару Юрійовичу

(прізвище, ім'я, по батькові)

(підпис)

1. Тема роботи: Розробка голосового інтерфейсу для створення діаграм, науковий керівник роботи _____ д.т.н., професор Куссуль Н. М. _____, (прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від

2. Термін подання студентом роботи
3. Об'єкт дослідження: Голосові інтерфейси для створення діаграм.
4. Предмет дослідження: Моделі машинного навчання для аналізу відкритого голосового вводу для створення діаграм.
5. Перелік завдань, які потрібно розробити: Провести аналіз існуючих публікацій по даній темі, провести дослідження користувачів для знаходження структури голосових описів, згенерувати датасет та побудувати моделі машинного навчання для створення діаграм на основі голосового вводу.
6. Орієнтовний перелік ілюстративного матеріалу: Презентація доповіді

9 Дата видачі завдання « 16 » травня 2022 р.

Календарний план

| № з/п | Назва етапів виконання бакалаврської дисертації | Термін виконання етапів проекту | Примітка |
|-------|---|---------------------------------|----------|
| 1. | Опрацювання літератури за темою | 16.05.2022 - 01.09.2022 | Виконано |
| 2. | Проведення дослідження користувачів | 02.09.2022 - 28.12.2022 | Виконано |
| 3. | Генерація датасету | 09.01.2023 - 28.04.2023 | Виконано |
| 4. | Реалізація моделі | 01.05.2023 - 14.05.2023 | Виконано |
| 5. | Аналіз отриманих результатів | 15.05.2023 - 31.05.2023 | Виконано |

Студент

(підпис)

Н. Ю. Поночевний

(ініціали, прізвище)

Науковий керівник роботи

(підпис)

Н. М. Куссуль

(ініціали, прізвище)

РЕФЕРАТ

Пояснювальна записка дипломної роботи за обсягом становить 63 сторінки, містить 4 таблиці та 6 рисунків. Для дослідження було використано 64 бібліографічних найменувань.

Люди все частіше приймають рішення на основі аналізу даних і одним з популярних способів комунікації фактів є візуалізація даних. Проте процес створення графіків вручну є довгоємним процесом і потребує вміння візуалізації даних, тому цей процес може бути частково автоматизований за допомогою голосового інтерфейсу. До цього часу вже було проведено багато досліджень щодо створення графіків за допомогою коротких вказівок, але все ще залишаються питання щодо використання вільного голосового вводу.

Ми провели розвідувальне дослідження, щоб визначити структуру голосової взаємодії, охарактеризувати висловлювання відповідно до їхніх фраз та інформації, яку вони містять, а також виявити прогалини. Крім того, ми згенерували датасет голосових описів графіків і виконали порівняльний аналіз моделей машинного навчання на основі архітектури Transformer для генерації коду графіків, щоб розробити прототип розмовного інтерфейсу користувача для створення графіків за допомогою голосового введення.

Отримані результати дозволяють розробляти більш ефективні голосові розмовні інтерфейси користувача та застосовувати їх у більш складних контекстах аналізу даних.

Основний внесок цієї дипломної роботи полягає у формулюванні структури і генерації датасету відкритих голосових описів графіків. Це необхідно для тренування і покращення моделей машинного навчання, які генерують код графіків на основі відкритого голосового введення.

Ключові слова: *машинне навчання, візуалізація даних, голосовий інтерфейс, трансформери, обробка природної мови.*

SUMMARY

The diploma work explanatory note includes 63 pages of the text, 4 tables and 6 illustrations. At the problem modern state analysis, overall 64 references were used.

People are increasingly making decisions based on data analysis, and one of the most popular ways to communicate facts is through data visualization. However, the process of creating graphs manually is time-consuming and requires data visualization skills, so this process can be partially automated using a voice interface. So far, a lot of research has been done on creating graphs using short utterances, but there are questions remaining about the use of free-form voice input.

We conducted an exploratory study to determine the structure of voice interaction, characterize utterances according to their phrases and the information they contain, and identify gaps. In addition, we generated a dataset of voice descriptions of bar charts and performed a comparative analysis of machine learning models based on the Transformer architecture for generating chart code to develop a prototype conversational user interface for creating charts using voice input.

The results will allow for the development of more effective voice conversational user interfaces and their application in more complex data analysis contexts.

The main contribution of this thesis is the formulation of a structure and generation of a dataset of open voice chart descriptions. This is necessary for training and improving machine learning models that generate chart code based on open voice input.

Keywords: *machine learning, data visualization, voice interface, transformers, natural language processing.*

ЗМІСТ

| | |
|--|----|
| ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ. | 8 |
| ВСТУП | 9 |
| РОЗДІЛ 1. Теоретичні відомості | 12 |
| 1.1. Доступні медіа та інтерфейси взаємодії людини з комп'ютером . | 12 |
| 1.2. Ієрархії природної мови та візуалізації. | 13 |
| 1.3. Автоматичні методи створення візуалізації | 15 |
| 1.4. Висновки до розділу 1 | 18 |
| РОЗДІЛ 2. Побудова датасету | 20 |
| 2.1. Емпіричне дослідження голосових інструкцій у довільній формі. | 20 |
| 2.1.1. Метод | 20 |
| 2.1.2. Результати | 26 |
| 2.2. Синтетичне розширення набору даних | 35 |
| 2.2.1. Метод | 35 |
| 2.2.2. Впровадження. | 36 |
| 2.3. Висновки до розділу 2 | 38 |
| РОЗДІЛ 3. Система для генерації діаграм на основі голосового вводу . . . | 42 |
| 3.1. Вимоги до системи | 42 |
| 3.2. Побудова моделей машинного навчання | 44 |
| 3.2.1. Few-shot Learning GPT-J-6B модель | 44 |
| 3.2.2. Few-shot Learning GPT-3 Davinci 175B модель. | 47 |
| 3.2.3. Fine-tuned GPT-3 Davinci 175B модель | 47 |
| 3.3. Тестування розроблених моделей | 47 |
| 3.3.1. Метод | 47 |
| 3.3.2. Результати | 48 |
| 3.4. Функціонал кінцевого інтерфейсу | 49 |
| 3.4.1. Опис діаграми голосом | 49 |
| 3.4.2. Експорт візуалізації. | 50 |

| | |
|---|----|
| 3.4.3. Реалізація | 51 |
| 3.5. Висновки до розділу 3 | 51 |
| ВИСНОВКИ. | 53 |
| ПЕРЕЛІК ПОСИЛАНЬ | 54 |
| ДОДАТКИ. | 62 |
| Додаток А. Матеріали для дослідження користувачів | 62 |

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ML (МН) — machine learning (машинне навчання)

CUI — conversational user interface (розмовний інтерфейс користувача)

HCI — human-computer interaction (взаємодія людини з комп'ютером)

CV — computer vision (комп'ютерний зір)

NLP — natural language processing (обробка природної мови)

CNN — convolutional neural network (згорткова нейронна мережа)

LSTM — long short-term memory (нейронна мережа з довгою короткочасною пам'яттю)

RF — random forest (метод машинного навчання випадковий ліс)

ВСТУП

Актуальність дослідження. Через експоненціальне зростання обсягів даних люди все частіше використовують інтерактивні візуалізації для прийняття рішень на основі даних [1]. Глибина наукових знань у різних галузях неухильно розширюється з кожним роком, що посилює тиск на аналітиків, дослідників і підприємців у створенні та інтерпретації візуалізацій у таких складних сферах, як забруднення, безпека, медицина і природничі науки [2]. Створення графіків зазвичай забирає багато часу, включає багато рутинних процесів [3] і має проблеми з доступністю для людей з вадами зору [4]. Одним із шляхів вирішення цих проблем, що вивчається в дослідженнях на сьогоднішній день, є використання діалогових інтерфейсів користувача (CUI) для створення графіків. Дослідники вивчали використання CUI [5] і продемонстрували, що вони мають певний потенціал для підвищення ефективності [6]. Крім того, голосові інтерфейси можуть покращити доступність і дозволити використання декількох пристроїв [7].

Ефективний дизайн системи залежить від розуміння структури розмовного введення, яке природно надається людьми. Описи діаграм мають ієрархічну структуру з чотирьох семантичних рівнів: (1) елементарні та закодовані властивості; (2) статистичні концепції та взаємозв'язки; (3) перцептивні та когнітивні явища; (4) контекстні та специфічні для домену уявлення [8]. З цих рівнів ми можемо дізнатися, що ці описи включають фактичні елементи та інтерпретації. Так само були зроблені перші спроби вивчити структуру розмовного введення для розробки діаграм [9].

Однак, хоча дослідники зосередилися на коротких висловлюваннях, залишилися питання щодо використання відкритих голосових описів. Деякі елементи повинні бути включені в описи, такі як атрибути даних, зв'язки між ними, кодування і параметри дизайну. Досі не визначено, чи будуть люди систематично включати їх. Таким чином, щоб розробити математичну модель,

яка спиратиметься на описи у вільній формі, надані користувачами для генерації коду діаграми, нам потрібно розуміння структури та набір даних природних голосових описів. Це дозволить розробляти більш ефективні голосові діалогові інтерфейси користувача та застосовувати їх у більш складних контекстах аналізу даних.

Метою дослідження є розробка ефективного голосового інтерфейсу для створення діаграм, який відповідає потребам користувачів.

Завданням дослідження є побудова датасету відкритих голосових описів діаграм і реалізувати модель машинного навчання для генерації коду діаграм на основі голосового вводу. Для цього необхідно:

- 1) Проаналізувати літературу у сфері голосових інтерфейсів для створення діаграм
- 2) Побудувати набір даних відкритих голосових описів діаграм
- 3) Реалізувати та порівняти моделі машинного навчання для генерації коду діаграм на основі відкритого голосового вводу
- 4) Порівняти метрики для оцінки моделей генерації коду
- 5) Зробити загальні висновки до отриманих результатів

Об'єктом дослідження є голосові інтерфейси для створення діаграм.

Предметом дослідження є моделі машинного навчання для аналізу відкритого голосового вводу для створення діаграм.

Методи дослідження: тематичний якісний аналіз, методи математичної статистики, і методи побудови та навчання глибоких нейронних мереж.

Наукова новизна полягає у дослідженні голосових взаємодій у вільній формі для візуалізації даних, що є майже незвіданою територією порівняно з використанням коротких голосових команд. Також був представлений унікальний набір даних відкритих голосових описів діаграм для навчання моделей машинного навчання і порівнюються моделі на основі Transformer для перекладу голосу в код діаграми.

Практичне значення полягає в оптимізації процесу створення візуалізацій даних та покращенні користувацького досвіду. Завдяки розробці діалогового інтерфейсу користувача, який перетворює голосові команди у довільній формі на графіки, процес створення візуалізацій даних стає більш доступним та інтуїтивно зрозумілим, заощаджуючи час та зусилля. Більше того, унікальний набір даних згенерованих голосових описів діаграм може стати цінним інструментом для вдосконалення моделей машинного навчання для застосунків з перетворенням голосу в код. Таким чином, це дослідження розширює межі голосових технологій, відкриваючи нові можливості для більш складних, природних взаємодій з голосовими системами.

Апробація результатів полягає в проведенні дослідження в колаборації з The COoKIE Group HCI лабораторією в Університеті Торонто. Також, планується публікація цієї наукової роботи на The ACM Conference on Human Factors in Computing Systems (CHI) у Гонолулу, США.

РОЗДІЛ 1.

ТЕОРЕТИЧНІ ВІДОМОСТІ

Багато дослідників вивчали взаємодію між візуалізаціями та їхніми описами природною мовою. Однак до завдання розробки голосового інтерфейсу для створення діаграм дослідники зазвичай підходили як до інженерної задачі. Для того, щоб отримати цілісну картину, ми розглянемо літературу, включаючи доступні медіа-інтерфейси, взаємодію людини з комп'ютером (HCI), ефективне дослідження даних, структури взаємодії тексту та діаграм, комп'ютерний зір (CV) та обробку природної мови (NLP).

1.1. Доступні медіа та інтерфейси взаємодії людини з комп'ютером

Мультимодальні інтерфейси часто допомагають людям з вадами зору покращити доступ до інформації, наприклад, шляхом інтеграції мультисенсорного зворотного зв'язку для вирівнювання доступу та можливості вилучення сенсу [10].

Цей підхід може застосовуватися майже до будь-якого типу модальності. Наприклад, для розуміння зображень 6 з 10 людей з вадами зору заявили, що надають перевагу аудіоопису для GIF-файлів [11]. Більше того, універсальні описи є більш доречними, оскільки люди схильні звертати увагу на різні деталі зображення в різних контекстах [12]. Можливим рішенням є використання багатошарової сенсорної системи дослідження зображень, яка допоможе людям з вадами зору визначити відносне розташування об'єктів один до одного, а також ієрархічно розкрити внутрішню структуру об'єктів [13]. У сфері створення візуалізацій дослідники також використовують мультимодальні інтерфейси, такі як тангенціально-звукові інтерфейси [14], щоб дозволити людям з вадами

зору використовувати очевидні інструменти, такі як Microsoft® Excel® [7], а також автоматично створювати доступні аудіо-сенсорні описи для малюнків.

Крім того, інтеграція різних модальностей сприяє розумінню даних навіть зрячими людьми. Наприклад, 19 з 24 учасників дослідження [15] надали перевагу статтям на основі даних, що прокручуються, з посиланнями на текстові діаграми, ніж звичайним статтям без посилань на текстові діаграми, оскільки посилання на текстові діаграми допомагають спрямовувати читання, додають більше інформації та покращують потік розповіді. Аналогічно, використання текстової модальності для уточнення деталей і швидкого перегляду документів [16] і візуалізацій [17; 18] має потенціал для підвищення продуктивності користувачів. Продуктивність та ефективність є важливими для дослідження даних. Наприклад, розуміння великого набору даних з невідомими структурами і отримання значущих висновків для аналітика даних займає багато часу. Тому дослідники розробляють специфікації візуалізації [19], системи дослідження даних з підтримкою запитів природною мовою [20] і системи для інтерактивних рекомендацій візуалізації на основі завдань [21] і вподобань користувачів [22; 23].

1.2. Ієрархії природної мови та візуалізації

Перш ніж розробляти інтерфейс для мультимодальної взаємодії, необхідно зрозуміти структуру взаємодії кожної модальності.

Дослідники з групи візуалізації MIT використовували методологію обґрунтованої теорії для розробки і застосування чотирирівневої моделі семантичного змісту [8], щоб оцінити, як люди з вадами зору і зрячі оцінюють корисність текстових описів діаграм на різних рівнях. В іншому дослідженні [9], 102 учасники отримали десять пар таблиць і діаграм і попросили написати текстовий опис діаграми, щоб зрозуміти, як і яку інформацію люди схильні вказувати. Крім

того, ми хотіли б дослідити, як користувачі описують графіки за допомогою голосу. Для опису графіків для комп'ютерних систем дослідники розробили мову Vega-Lite [24] з інтерпретатором, який дозволяє перевести опис Vega-Lite у візуалізаційне зображення.

Для створення змістовних і практичних діаграм дослідники розробляють настанови [25] з рекомендаціями щодо структури взаємодії та графічних елементів, які сприяють більш чіткому сприйняттю візуальних образів. Крім того, К. С. Shahira та А. Ліїуа дослідили [26] існуючу літературу, присвячену розумінню графіків та вилученню з них візуального кодування. Наприклад, у 2000 році було розроблено алгоритм машинного навчання для розпізнавання рукописних гістограм, але без вилучення даних [27]. З появою глибокого навчання [28], у 2021 році з'явилася система, що складається зі згорткової нейронної мережі (CNN) [29] та нейронної мережі з довгою короткочасною пам'яттю (LSTM) [30] з механізмом уваги, яка генерує одиниці підписів на одиночних гістограмах [31]. У 2022 році було розроблено модель на основі Transformer [32] (рис. 1.1) для розв'язання загальної задачі перетворення діаграми в текст [33].

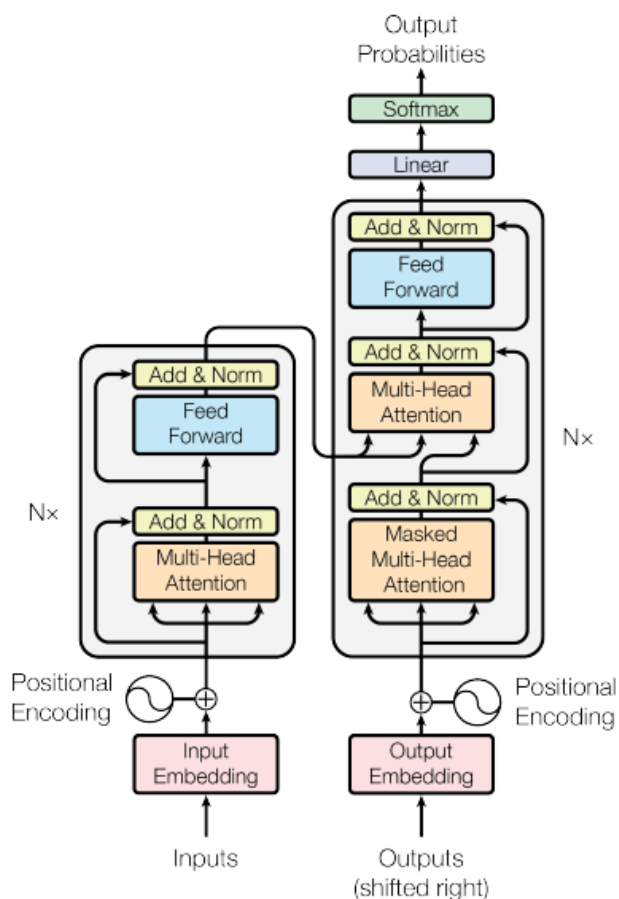


Рис. 1.1. Архітектура моделі типу трансформер

1.3. Автоматичні методи створення візуалізації

Для того, щоб створити автоматичні методи візуалізації з природної мови, вчені спочатку зібрали набори даних і метрики для навчання та оцінки моделей.

Наприклад, набір даних SciCap: Scientific Figures [34], створений з використанням понад двох мільйонів зображень з більш ніж 290 000 наукових статей, зібраних і опублікованих на arXiv, і набір даних nvBench [35], що складається з 25 750 пар описів і візуалізацій, для оцінки моделей в 105 різних областях і семи типів візуалізацій. Метрика Bilingual Evaluation Understudy (BLEU) [36] - це широко використовувана метрика для оцінювання якості машинного контенту. Вона порівнює машинний контент з одним або кількома еталон-

ним контентом, створеним людиною. Формула оцінки BLEU (1.1) і BP (1.2) виглядає наступним чином:

$$BLEU = BP \cdot \exp \left(\sum_{n=1}^N w_n \log p_n \right) \quad (1.1)$$

$$BP = \begin{cases} 1 & \text{if } c > r \\ \exp \left(1 - \frac{r}{c} \right) & \text{if } c \leq r \end{cases} \quad (1.2)$$

де

- $BLEU$ - оцінка BLEU
- BP - фактор короткого штрафу (brevity penalty)
- w_n - вага для кожної n -грами (зазвичай вибирається так, що сума всіх w_n дорівнює 1)
- p_n - точність n -грами
- c - довжина машинно-генерованого тексту
- r - довжина еталонного тексту
- N - максимальний розмір n -грами (зазвичай 4 в задачах машинного перекладу)

Метрика Recall-Oriented Understudy for Gisting Evaluation - Longest Common Subsequence (ROUGE-L) [37] - це широко використовувана метрика для оцінювання якості автоматичного резюмування тексту. Вона порівнює автоматично згенероване резюме з одним або кількома еталонними резюме, створеними людиною. Формула оцінки ROUGE-L (F) (1.3) виглядає наступним чином:

$$ROUGE - L(F) = \frac{(1 + \beta^2) \cdot R \cdot P}{R + \beta^2 \cdot P} \quad (1.3)$$

де

- $ROUGE - L(F)$ - оцінка ROUGE-L (F)

- R - повернення (recall), яке вимірює відсоток n -грам в еталонному резюме, які зустрічаються в автоматичному резюме
- P - точність (precision), яка вимірює відсоток n -грам в автоматичному резюме, які зустрічаються в еталонному резюме
- β - вага, яка контролює баланс між точністю та поверненням (recall). Зазвичай вибирається $\beta = 1$, щоб надати однакову важливість точності та поверненню.

Метрика BERTScore [38] - це новітня метрика для оцінювання якості машинного контенту. Вона використовує BERT [39], модель переносу навчання машинного навчання, для визначення семантичної подібності між машинно-згенерованим текстом та одним або кількома еталонними текстами, створеними людиною. Формула оцінки BERTScore (1.4) виглядає наступним чином:

$$BERTScore = \frac{1}{N} \sum_{n=1}^N \cos(T_n, C_n) \quad (1.4)$$

де

- $BERTScore$ - оцінка BERTScore
- $\cos(T_n, C_n)$ - косинусна подібність між BERT вбудовуваннями для n -го токена в еталонному тексті T_n і n -го токена в кандидатському тексті C_n
- N - кількість токенів в тексті

Після цього з'явилися системи для автоматичного створення специфікацій даних за допомогою текстових інструкцій [40], системи для створення діаграм на основі табличних даних і запитань природною мовою [41], а також системи для автоматичного створення інфографіки на основі текстового запиту [42]. Для того, щоб допомогти користувачеві правильно скласти текстовий запит, було розроблено метод автоматичного підказування на основі даних користувача [43].

Еволюція роботи в задачі перетворення тексту в діаграму була описана Шеном та ін. у 2021 році [5]. Прототип інтерфейсу візуалізації та природної

мови був створений у 2001 році [44], але він приймав лише добре структуровані запити. З розвитком нейронних мереж машинного навчання та глибокого навчання у 2021 році було розроблено підтвердження концепції, що складається з моделей Random Forest (RF) [45] та LSTM, які запускаються одна за одною. Однак набір даних був меншим за 500 зразків [46]. З подальшим масштабуванням мовних моделей було показано, що ці моделі можуть досягти покращеної продуктивності в діагностиці завдань за кілька спроб (few-shot learning). Зокрема, модель GPT-3, авторегресійна мовна модель зі 175 мільярдами параметрів, продемонструвала високу продуктивність у численних завданнях NLP за кілька спроб без градієнтного оновлення або тонкого налаштування (fine-tuning). До таких завдань належать переклад, відповіді на запитання, а також завдання, що вимагають міркування на льоту або адаптації до предметної області [47]. Однак ця модель все ще стикається з проблемами з певними наборами даних. У 2022 році було створено інтерфейс Sevi: Speech to Visualization, що складався з моделі на основі Transformer [6]. Однак, хоча дослідники зосередилися на коротких надрукованих висловлюваннях, залишалися питання щодо використання відкритих голосових описів. Тому в цьому дослідженні ми зосередилися на підході, орієнтованому на користувача.

1.4. Висновки до розділу 1

Розділ надав важливу інформацію щодо використання мульти-модальних інтерфейсів, структуру описів і візуалізацій, а також аналіз автоматичних методів створення візуалізацій.

У секції 1.1 був представлений потенціал мульти-модальних інтерфейсів для ефективної взаємодії з візуалізаціями.

В секції 1.2 було описано структуру описів і візуалізацій, а також еволюцію методів машинного навчання для розпізнавання графіків.

Далі, у секції 1.3 були розглянуті деякі набори даних з візуалізаціями, метрики для оцінки моделей генерації тексту, і еволюція методів машинного навчання для генерації графіків.

Цей розділ надав важливу інформацію про використання голосових інтерфейсів, деякі наявні набори даних, еволюцію методів машинного навчання для генерації графіків, а також реалізацію метрик для оцінки генерації тексту. В наступному розділі ми опишемо метод генерації датасету, базуючися на аналізі літератури.

РОЗДІЛ 2.

ПОБУДОВА ДАТАСЕТУ

Для побудови датасету, треба спочатку визначити які типи елементів користувачі як правило включають під час відкритого голосового вводу. Для цього, було проведено дослідження користувачів.

2.1. Емпіричне дослідження голосових інструкцій у довільній формі

Метою дослідження є аналіз структури природних голосових описів, що використовуються для створення діаграм, та виявлення прогалин у цьому процесі. Питання дослідження полягає у наступному: Які типи елементів учасники зазвичай включають у свої голосові описи для створення діаграм?

Щоб відповісти на це дослідницьке питання, ми провели розвідувальне користувацьке дослідження, в якому 24 учасникам представили текстові історії для забезпечення контексту. Потім їх попросили вибрати випадкову частину інформації, уявити собі діаграму та описати її голосом у зручний для них спосіб.

2.1.1. Метод

У наступних розділах ми розповімо про процес набору учасників, учасників та процедуру дослідження. Ми також надамо детальне пояснення методів аналізу даних, що використовувалися в цьому процесі.

2.1.1.1. Учасники

Ми набрали 24 учасники через списки електронної пошти студентів, дослідників та інших людей, які часто створюють діаграми у своїй роботі. Єдиною вимогою для участі було володіння англійською мовою та повнолітній вік. Учасники не отримували жодної компенсації. Серед 24 учасників 54% вказали свою стать як чоловічу (n=13) і 46% як жіночу (n=11). Ми класифікували рівень досвіду, використовуючи наступний протокол:

- Ніякого: Немає попереднього досвіду або знань про створення діаграм і графіків.
- Початківець: Базові знання про створення простих діаграм, таких як гістограми та кругові діаграми, за допомогою таких інструментів, як Excel та Google Таблиці.
- Середній рівень: Досвід створення більш складних діаграм і графіків та використання таких інструментів, як PowerBI і Tableau. Знання основних методів візуалізації даних та попереднього аналізу даних.
- Просунутий: Великий досвід у створенні складних і нестандартних діаграм і графіків за допомогою таких мов програмування, як Python і R, з використанням сучасних інструментів візуалізації даних, таких як matplotlib і seaborn. Знання складних методів візуалізації даних та досвід у дослідницькому аналізі даних.

На основі цієї схеми ми визначили, що 29% мають початковий (n=7), 50% - середній (n=12) і 21% - просунутий (n=5) досвід або знання у створенні діаграм. Для досвіду роботи з віртуальними помічниками ми використовували наступну схему:

- Жодного: Люди, які не користуються віртуальними помічниками або через технічні обмеження, або вважають за краще отримувати доступ до інформації вручну.

- Нечасто: Люди, які користуються віртуальними помічниками час від часу, здебільшого для розваг або коли не можуть отримати доступ до свого телефону для виконання основних завдань.
- Регулярно: Люди, які користуються віртуальними помічниками щодня для виконання простих завдань, таких як налаштування таймерів, визначення погоди і часу або відтворення музики.
- Експерт: люди, які є експертами в технології віртуальних помічників, користуються ними щодня і часто тестують їхні можливості та обмеження.

Ми визначили 42% як таких, що не мають досвіду (n=10), 12,5% як таких, що користуються віртуальними помічниками нечасто (n=3), 33% як таких, що користуються ними регулярно (n=8), і 12,5% як таких, що мають досвід або знання про віртуальних помічників на рівні експерта (n=3). Розподіл учасників за статтю, діаграмою та досвідом роботи з віртуальними асистентами можна знайти на рис. 2.1.

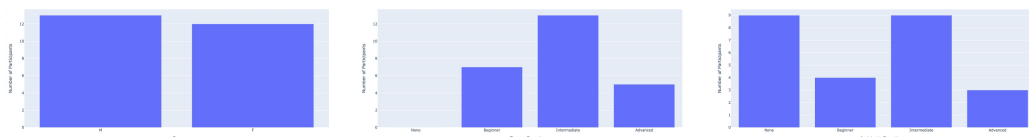


Рис. 2.1. Демографічні дані користувача

Учасники повідомили, що створюють діаграми за допомогою різних інструментів, включаючи бібліотеки Python (У12, У25, У2), онлайн-інструменти (У18), Excel (У17), Google Таблиці (У8) та Google Data Studio (У8). Деякі учасники надають перевагу бібліотекам Python, оскільки вони надають більше можливостей для кастомізації та є більш зручними (У12, У25, У2). Інші використовують онлайн-інструменти, Excel або Google Sheets, для простоти і зручності використання (У18, У17, У8). Деякі учасники використовують ці інструменти для побудови діаграм в академічних цілях (У17), а інші - для ілюстрації ідей та розподілу даних у своїй роботі інженерами машинного на-

вчання (У2). Більшість учасників використовують шаблони для створення діаграм (У2), причому тип діаграми залежить від мети.

З 24 учасників десять не використовують віртуальних помічників. Серед 14, які користуються віртуальними помічниками, більшість використовують їх для простих завдань, таких як встановлення таймерів, перевірка прогнозу погоди, відтворення музики та здійснення дзвінків. Кілька людей також використовують їх для більш складних завдань, таких як керування світлом і домашніми пристроями. Дехто використовує кілька віртуальних помічників і перемикається між ними залежно від потреб. Однак деякі люди мають проблеми з конфіденційністю або вважають, що віртуальні асистенти заважають, і вимикають їх. Дехто також вважає, що голосові помічники лише іноді бувають точними, і це не дає їм змоги регулярно ними користуватися. Крім того, деякі люди не користуються голосовими помічниками в громадських місцях, бо вважають їх безглуздими.

2.1.1.2. Процедура

Щоб відобразити весь шлях користувача, ми вирішили не показувати учасникам заздалегідь розроблені діаграми і таблиці даних. Замість цього ми надали загальну текстову історію для контексту і попросили користувачів вирішити, яку частину інформації вони хотіли б візуалізувати і в який спосіб. Таким чином, ми почали збирати дані з пошуку текстових матеріалів, які часто потребують візуалізації, наприклад, статей, публікацій та звітів з економіки, географії, статистики, демографії та новин. Щоб запобігти засмученню користувачів невідповідною або невідомою інформацією, важливо було відбирати лише текстовий контент на нейтральні теми. Таким чином, ми зібрали корпус з восьми текстів довжиною від 54 до 100 слів (див. таблицю 2.1) з ресурсу Statista [48]. Тексти містили звичайну загальновідому інформацію, яку можна було доповнити візуалізацією - наприклад, числовими значеннями або порівнянням понять. Додаток А містить повні тексти, які були використані в дослідженні.

| Текст | Кількість слів |
|---------|----------------|
| Текст 1 | 96 |
| Текст 2 | 82 |
| Текст 3 | 80 |
| Текст 4 | 65 |
| Текст 5 | 81 |
| Текст 6 | 100 |
| Текст 7 | 54 |
| Текст 8 | 57 |

Таблиця 2.1

Кількість слів для кожного тексту

Приблизно 30-хвилинне дослідження проводилося онлайн через Zoom. Після вступного слова та інструкцій до дослідження учасники отримували чотири випадково вибрані тексти з корпусу з восьми текстів. Завдання полягало в тому, щоб уважно прочитати текст і подумати, яку частину тексту можна було б підкріпити відповідною візуалізацією. Наприклад, якщо в тексті йдеться про результати команд на міжнародному турнірі, то сприйняття цієї інформації можна покращити за допомогою візуалізації у вигляді відсортованої гістограми. Потім ми попросили користувачів створити відповідну діаграму за допомогою голосу, ніби вони просили колегу створити для них малюнок. Дослідники ставили уточнюючі запитання про дії учасників протягом усього процесу. Процес повторювався для всіх чотирьох текстів. Дослідження завершилося напівструктурованим дебрифінг-інтерв'ю про проблеми учасників під час виконання вправи та їхній досвід роботи з фігурами і голосовими помічниками.

2.1.1.3. Аналіз даних

Інтерв'ю були записані на аудіо та розшифровані. Усі дані, включно із записами та нотатками інтерв'юера, зберігалися на захищеному сервері. Всі дані були захищені паролем, зашифровані та анонімізовані.

Кількість відповідей на кожен текст варіювалася, оскільки кожному учаснику випадковим чином було запропоновано чотири тексти. Однак середня кількість відповідей на кожен текст становила приблизно $12,5 \pm 4,5$, як показано на рис. 2.2. Середня довжина слова відповіді для кожного тексту становила

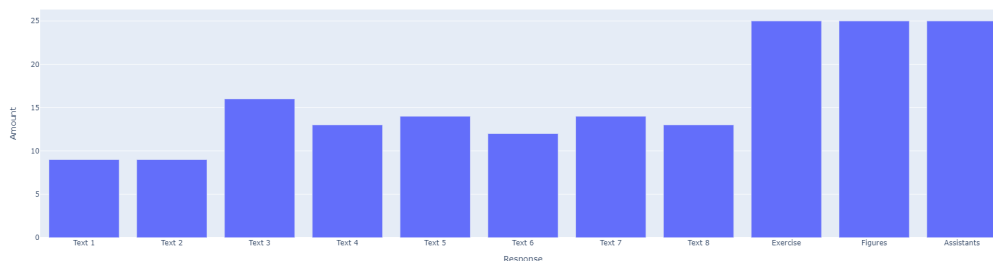


Рис. 2.2. Кількість відповідей на кожен текст та питання

приблизно 173.4 ± 34.7 , як показано на рис. 2.3. Для того, щоб отримати

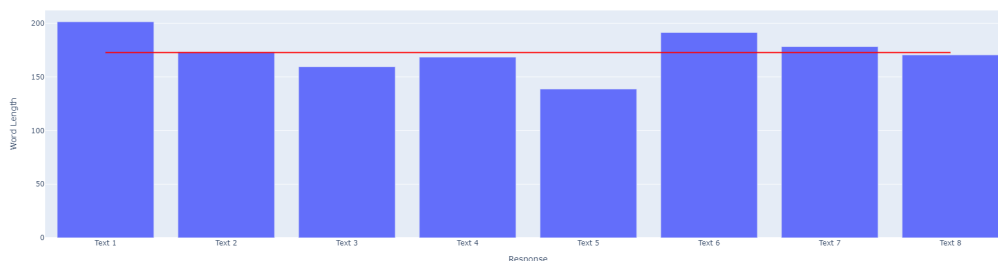


Рис. 2.3. Середня довжина слова у відповідях для кожного тексту

неупереджені описи гістограм довільної форми, ми навмисно утрималися від явного обмеження типу діаграми в інструкції до дослідження. Після збору 100 описів діаграм ми проаналізували дані і виявили, що лише 50 відповідей містили опис гістограми або невизначеного типу діаграми. Варто зазначити, що кожен текст у нашому дослідженні можна було б описати як гістограму. Детальний розподіл типів графіків можна побачити на рис. 2.4. З огляду на те, що в цьому дослідженні ми зосередилися на гістограмах, ми звузили наш аналіз до 50 описів діаграм, які відповідали цьому критерію, зменшивши кількість учасників з 25 до 24, оскільки лише один учасник описував виключно негістографічні типи діаграм.

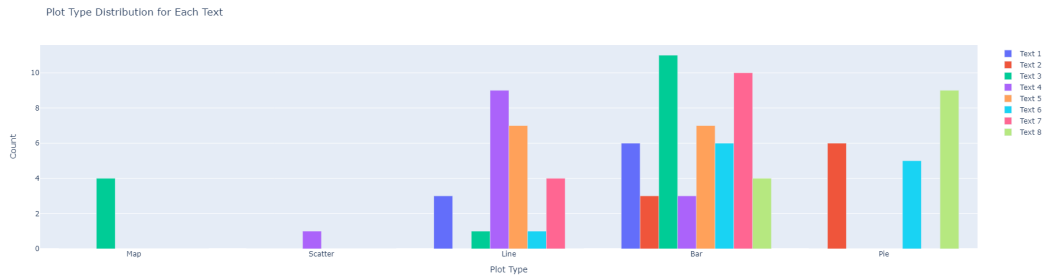


Рис. 2.4. Розподіл типів графіків для кожного тексту

Нарешті, ми провели тематичний аналіз даних з використанням індуктивного підходу [49]. Один дослідник переглянув усі інтерв'ю та стенограми, щоб розробити початковий набір кодів, які були уточнені під час наступних проходів через дані. Ми використовували систему відкритого кодування, щоб забезпечити гнучкість і вдосконалення кодів, дозволяючи з'являтися новим кодам і розвивати існуючі коди за необхідності [50]. Ми визначили теми, вивчивши взаємозв'язки між кодами, і уточнили їх за необхідності. Ми організували теми в послідовний набір, який відображав досвід учасників щодо наших дослідницьких запитань. Виявлені теми вплинули на результати, описані нижче.

2.1.2. Результати

2.1.2.1. Які типи елементів учасники зазвичай включають у свої голосові описи для створення діаграм?

Учасники, як правило, включають певні типи елементів у свої голосові описи. Ці елементи можна умовно поділити на п'ять груп: елементи діаграми, формат команд, характеристики елементів, організація елементів та мовні особливості (див. таблицю 2.2). Наш аналіз не виявив жодних значущих закономірностей у порядку, в якому учасники надавали різні типи елементів опису діаграми.

Елементи діаграми - це основні компоненти гістограми, зокрема назва, тип, підпис, вісь, масштаб, одиниці виміру, стовпчики, значення, форма і

підпис. Іноді учасники детально описують ці елементи, надаючи інформацію про їхні характеристики та атрибути.

Назва. Елемент заголовка на гістограмах учасники описують двома різними способами: неявним (11 входжень) та явним (20 входжень). У неявному підході учасники описують тему діаграми таким чином, що її можна інтерпретувати як заголовок. Наприклад, учасник зазначив: "Отже, третій графік, я б назвав його, я б назвав його базою користувачів LinkedIn у США, які використовують Android. Або давайте використаємо іншу назву: Android, база користувачів LinkedIn у США"(Текст 8, У10). В експліцитному підході учасники безпосередньо вказують елемент заголовка в описі діаграми. Наприклад, учасник зазначив: "Гм, так я б зробив. Я уявляю, що це буде стовпчикова діаграма, вертикальна стовпчикова діаграма з заголовком, що говорить про гендерний розрив в Еквадорі"(Текст 7, У1).

Тип діаграми. Учасники описували тип діаграми як стовпчикову, але інколи вказували на її орієнтацію і називали її гістограмою. Орієнтація гістограми була вказана в 6 з 50 випадків. Наприклад, учасник сказав: "Гаразд, цього разу давайте зробимо горизонтальну гістограму"(Текст 1, У1). З іншого боку, 7 з 50 учасників назвали гістограму гістограмою. Наприклад, один учасник сказав: "Я так розумію, що ви можете побудувати гістограму"(Текст 3, У12).

Мітка для осі. Зазвичай учасники неявно вказують мітку для осі, що траплялося у 77 з 85 випадків. Наприклад, учасник зазначив: "Отже, на осі X ми відкладемо роки, які показують, як дані змінювалися з часом. А на осі Y ми матимемо відсоток жінок, відсоток жінок, відсоток того, наскільки жінки мають менше або більше шансів, ніж чоловіки, мати рівні можливості"(Текст 7, У6).

Вісь. Учасники, як правило, надають інформацію про вісь двома способами: або вказуючи осі X та Y (65 випадків), або орієнтацію (11 випадків), хоча орієнтація зустрічається рідше. Наприклад, учасник зазначив: "Створіть

графік, де на осі Y буде відкладено оцінку гендерного розриву, а на осі X - рік”(Текст 7, У21), щоб вказати осі X та Y. З іншого боку, учасник використав орієнтацію для опису осі, сказавши: ”ї вертикальна вісь буде для хвилин, а горизонтальна вісь буде для відсотків американських лікарів і їхніх витрат часу, так”(Текст 6, У14).

Шкала та одиниці виміру. Учасники нечасто включали інформацію про масштаб (6 з 50) та одиниці виміру (13 з 50) осі в опис діаграми. Наприклад, один учасник зазначив: ”Гм, але у нас буде шкала не від нуля до одиниці, а від нуля до 100, щоб зробити її більш, гм, візуалізованою”(Текст 7, У12) щодо масштабу осі. Інший учасник зазначив, що ”вісь X повинна мати напис ”роки а вісь Y може мати напис ”кількість користувачів”або ”кількість користувачів Facebook у мільярдах”(Текст 1, У19) щодо одиниць виміру осі.

Стовпчик і кількість. Учасники надали опис для елемента ”стовпчик”двома різними способами (55). Перший підхід - кількісний (30), коли учасники просять створити або один (13), або декілька стовпчиків (Кількість - 17). Другий підхід - послідовний (25), коли учасники створюють відрізки крок за кроком. Наприклад, у кількісному підході учасник зазначив: ”Я хотів би побудувати або хотів би намалювати просту стовпчикову діаграму, гм, три, з трьома стовпчиками”(Текст 1, У4). При послідовному підході учасник зазначив: ”Отже, перший стовпчик - округ Колумбія, 13 доларів за годину. Другий бар - Вашингтон, 13,50 доларів за годину. І Каліфорнія, 13 доларів за годину. І остання планка, гм, Сполучені Штати, \$7.25 за годину”(Текст 3, У9).

Значення. Учасники в описах графіків часто подають значення двома різними способами: явно (200 входжень) або неявно (4 входження). При явному підході учасники безпосередньо вказують значення для кожного стовпчика. Наприклад, у тексті 1, У1, учасник зазначив: ”Тоді наступним баром буде другий квартал 2020 року зі значенням 2,7”. При неявному підході учасники використовують нечіткі описи, такі як ”останній квартал”або ”безпосередньо перед/після пандемії щоб вказати значення. Як видно з тексту 1, У1, учасник

зазначив: "І останній, третій був би протягом останнього звітнього кварталу". Хоча це трапляється рідко, учасники також можуть звертатися до самого тексту, щоб отримати значення, як зазначено в тексті 5, У7, де учасник зазначив: "Отже, як на осі X, у вас буде 19 вересня, 20 липня та листопад 2020 року. Так, і відповідні числа".

Інша форма. Учасники в описах діаграм часто використовують різні фігури для представлення гістограми (13 випадків). Ці фігури включають лінії (8 випадків), зірочки, стрілки та зображення монет. Наприклад, у тексті 3, У4, учасник зазначив: "Я хотів би додати ще одну лінію, напевно. І щоб це була просто вертикальна лінія, яка б вказувала на мінімальний розмір мінімальної заробітної плати". У тексті 1, У16, інший учасник зазначив: "А потім, можливо, просто для, можливо, просто для ясності, я також хочу додати стрілку зверху, щоб показати прогноз, що це зростає з часом".

Підпис. Учасники запропонували три типи підписів до діаграми (15). Перший тип - це визначення терміну або використаного методу підрахунку (5). Наприклад, один учасник зазначив: "і, можливо, внизу я також додам виноску, щоб сказати, що щомісячні активні користувачі визначаються як, знаєте, десь є визначення, тож я додам визначення і туди"(Текст 1, У16). Другий тип - джерело даних (7). Наприклад, одна з учасниць зазначила: "О, ми могли б додати, ми могли б додати опис, що цей експеримент триває з 2020 року. І ми також могли б використати додаток App Are, який обчислював інтенсивність використання LinkedIn для різних аудиторій, щоб ми могли додати деякі дані, щоб бути більш достовірними для глядача цієї діаграми"(Текст 8, У10). Останній тип - це короткий опис, який можна розмістити під діаграмою або потенційно інтерпретувати як імпліцитний заголовок (3). Як зазначив один з учасників: "Було б добре мати невеликий опис під графіком, власне те, що він описує"(Текст 6, У5).

Формат команди стосується способу, в який учасники структурують свої голосові команди. Учасники використовують ітеративні команди, щоб ви-

конати дію для кожного елемента, звернутися до елемента діаграми, створеного раніше, і попросити поради.

Посилання на попередній(і) елемент(и). Учасники часто звертаються до раніше визначеного елемента діаграми (71 випадок), щоб змінити його характеристики чотирма різними способами: за порядком (33), за типом елемента (18), за значенням (17) і за розміром (3). Наприклад, деякі учасники визначають елементи на основі їхнього порядку, як показано в тексті 6, У3: "а потім буде гістограма, яка довша за попередню гістограму, тому що вона на 33%". Інші учасники змінюють елементи залежно від їхнього типу, як показано в тексті 1, У19: "на осі X буде мітка, мітка, пов'язана з віссю X, може бути роком, а мітка, пов'язана з віссю Y, може бути кількістю користувачів Facebook у мільярдах". Крім того, учасники можуть вказати елемент за його значенням, як показано в тексті 3, У4: "Тоді наступним, е-е, стовпчиком, який буде зображений трохи нижче, буде Каліфорнія". Нарешті, учасники можуть змінювати елементи за їхнім розміром, як у тексті 8, У2: "Буде чотири колонки, найбільша з них - неактивні користувачі LinkedIn, оскільки вони становлять 45% від усіх користувачів".

Ітеративна команда. Ітеративні команди, які передбачають створення кількох стовпчиків або додавання значень чи кольорів до кожного стовпчика, нечасто використовувалися учасниками в описі гістограм (6 випадків). Наприклад, у тексті 1, У7 пропонується використовувати ітеративні команди для додавання значень на кожен смужку: "Я б на осі Y відклав числа, і у вас буде три стовпчикові діаграми, на кожній з яких я б дописав додаткові числа, щоб було зрозуміло, що це за числа".

Попросіть поради. До елемента опису діаграми, пов'язаного з проханням звернутися за порадою до системи, учасники зверталися неявно і рідко - лише 7 разів. Серед них 2 стосувалися кількості даних, а 5 - найкращого способу побудови або розуміння діаграми. Наприклад, у тексті 7 учасниця 22 висловила зацікавленість у створенні графіка, який показує зміну гендерного розриву з

роками, з 28% до 24%, але не була впевнена, як відобразити бали та відсотки разом. Учасниця зазначила: "Найцікавіше в цьому графіку те, що на ньому є бали і є відсотки. Я не знаю, як показати це на графіку".

Характеристика елемента стосується атрибутів елементів діаграми, таких як орієнтація, колір і розмір. Учасники вказують ці характеристики, щоб налаштувати зовнішній вигляд діаграми.

Орієнтація. Учасники описали елемент орієнтації гистограми у 26 випадках. Зазвичай вони вказували орієнтацію для стовпчиків, форми та осі. Наприклад, у тексті 6, У3, учасник зазначив: "Я б поставив горизонтальну смугу, яка б відповідала відсотку американських лікарів, які витрачають стільки-то часу на розмови з пацієнтами". Це свідчить про те, що учасник зорієнтувався, сказавши, що стовпчики будуть горизонтальними.

Колір. Колірний елемент гистограми згадувався учасниками у 22 випадках. Зазвичай учасники вказували колір стовпчиків, і було два способи, якими вони це робили. У 10 випадках учасники прямо вказували колір, наприклад, зелений, червоний і синій. Як у тексті 5, У13, де учасник сказав: "Я б сказав, нехай це буде просто заради експерименту, нехай це буде смарагдового кольору, нагадує мені чомусь Minecraft, в основному зеленкуваті, е-е, кольорові смужки". У 12 випадках учасники описували колір неявно, наприклад, "інший" колір або "якийсь відтінок або посилалися на кольорову палітру конкретних об'єктів. Як у тексті 3, У22, де учасник сказав: "Тож я думаю, що, можливо, кольори Сполучених Штатів, щоб сподобатися, щоб, знаєте, люди дивилися на це і розуміли, що це про Сполучені Штати. Тому що всі дані, які ми збираємося розглянути, стосуються Сполучених Штатів".

Розмір. Елемент розміру на гистограмі згадувався учасниками 19 разів. Учасники, як правило, вказували розмір стовпчиків і зазначали, що він є відносно інших стовпчиків. Наприклад, у тексті 8, У20, учасник сказав: "І якщо поставити середніх користувачів, то це, гм, 7,1%, і ця частина знаходиться на правах попередньої, і вона трохи, трохи вища, ніж та, що представляє,

гм, важких користувачів”. Це свідчить про те, що учасник вказав розмір, зазначивши, що смужка, яка представляє середніх користувачів, трохи вища за смужку, яка представляє інтенсивних користувачів.

Організація елементів вказує на те, як учасники розташували елементи діаграми, зокрема їхнє розташування, порядок та кількість. Іноді учасники надають конкретні вказівки щодо того, як розташувати елементи діаграми, щоб досягти бажаного ефекту.

Позиція. Елемент позиції в описах діаграм часто згадується учасниками (33 входження). Учасники зазвичай вказують відносно розташування значень, смуг і підписів. Наприклад, один учасник зазначив: ”Отже, перший стовпчик я б намалював близько до початку координат X та Y, де вони перетинаються. Тому я б намалював смужку досить близько до лівої сторони, де близько до осі Y”(Текст 1, У16). Цей учасник вказує конкретне місце розташування першої смужки відносно осей x та y.

Порядок. Елемент порядку в описах діаграм також згадується учасниками (8 входжень). У всіх 8 випадках учасники вказували порядок розташування стовпчиків на графіку. Наприклад, один з учасників зазначив: ”Я хотів би побудувати графік зверху вниз. І я б попросив вас почати з найнижчої заробітної плати і побудувати діаграму з найменшою заробітною платою в першу чергу”(Текст 3, У4). Цей учасник надає чітку інструкцію щодо порядку розташування стовпчиків на графіку.

Лінгвістичні особливості - це способи, якими учасники використовують мову для опису елементів діаграми. Сюди входить виправлення раніше описаних елементів, повторення описів раніше згаданих елементів, саморефлексія шляхом надання метаописів та завершення опису гістограми.

Виправлення. Учасники, описуючи виправлення на діаграмах, зазвичай використовують два підходи - або виправляють помилку у значенні, або змінюють свою думку про те, що саме потрібно описати. У першому випадку (11 випадків) учасники визначають помилку і вказують правильне значення.

Наприклад, учасник зазначив: "Гістограма і горизонтальна для місяців і року, починаючи з вересня 2018 року, закінчуючи листопадом, ні, 20 липня 2020 року? Ні, листопадом 2020 року"(Текст 5, У14), виправивши помилку в кінцевій даті. У другому підході (6 випадків) учасники змінюють свою думку про те, як описати графік. Наприклад, учасник зазначив: "А потім ми будемо використовувати різні кольори для різних гістограм. Е-е, ні. Насправді, хм. Якщо ми будемо використовувати різні кольори, це може означати різні речі. Насправді, це одне й те саме, тому ми можемо використовувати один і той самий колір"(Текст 7, У25). Спочатку учасник запропонував використовувати різні кольори для різних гістограм, але потім змінив свою думку і запропонував використовувати один колір для всіх.

Повторення. Учасники, описуючи діаграми, повторюють раніше описані елементи, як правило, з трьох причин - щоб підсумувати опис, тому що вони забули, що вже описували його, або щоб написати аббревіатуру. У першому випадку (18 випадків) учасники надають стислий опис елемента, який вони вже описували раніше. Наприклад, учасник сказав: "Е-е, так, е-е, ну, ще раз, е-е, перший стовпчик описує третій квартал 2012 року, і, е-е, за цей час було, е-е, 1 мільярд користувачів, е-е, на Facebook"(Текст 1, У4), повторивши "Протягом третього кварталу 2012 року було 1 мільярд користувачів Facebook" як підсумок свого попереднього опису. У другому випадку (2 випадки) учасники повторюють елемент, тому що забули, що вже описували його. Наприклад, учасниця зазначила: "Отже, це буде стовпчикова діаграма, яка матиме рік як вісь Y, гендерний розрив у відсотках як вісь Y і рік як вісь X... додайте підпис на осі Y.... А потім на осі Y, на осі X, я б просто сказала рік"(Текст 7, У13), повторюючи свій опис позначки на осі X у стовпчиковій діаграмі, хоча вони вже надавали цей опис раніше. У третьому випадку (1 випадок) учасники повторюють елемент, щоб пояснити аббревіатуру. Наприклад, учасник пояснив: "Отже, ми хочемо показати гру, яка називається Player, вибачте, вона називається Battlegrounds або Player Unknowns Battlegrounds, так. Це аббревіатура р Р,

U B G. Так вона називається”(Текст 2, У22), повторюючи назву гри по буквах абрєвіатури.

Метаопис. Після специфікації елементів (56) учасники надали три типи метаописів: обґрунтування (21), плани на майбутнє (16), коментарі та відчуття користувачів (11). У підході обґрунтування учасники виправдовували свої дії, пояснюючи, чому вони включили певні елементи в опис діаграми. Наприклад, учасник зазначив: ”Ми також можемо намалювати лінію 7,25, гм, гм, в, і використовувати інший колір, щоб контрастувати, гм, гм, мінімальну заробітну плату, гм, з іншими”(Текст 3, У12). У підході до майбутніх планів учасники пояснювали, що вони планують описати далі. Наприклад, учасник зазначив: ”а на іншому, х, ми, ми мали б, так, тільки три, е, тільки час, е, і, е, три різні бари відповідно. Гм, так. Тож я можу перейти до більш детальної інформації, наприклад, до цифр, і тоді я закінчу”(Текст 1, У4). У підході ”коментарі та почуття користувачів”учасники ділилися своїми думками та емоціями щодо даних, таких як пропущені дані, тенденції, якість даних та їхнє використання. Наприклад, учасник прокоментував: ”У нас 7,1% середніх користувачів, 40,5% легких користувачів і 45,7% неактивних. О, якби я був LinkedIn, я б турбувався про ці останні 45 на 7% неактивних, тому що це дуже погано”(Текст 8, У22).

Закінчити фразу. Учасники цього дослідження зазвичай не додавали формальної фінальної фрази, яка б сигналізувала про завершення опису діаграми (35/50). Однак у деяких випадках (15/50) вони використовували звичайні розмовні фрази, такі як ”ось і все”або ”так, це все”. Наприклад, один учасник сказав: ”І остання планка буде трохи більшою за попередню... і за цей час у Facebook було понад 3 мільярди активних користувачів, так. Так що це має бути, так, ось так”. (Текст 1, У4)

2.2. Синтетичне розширення набору даних

Після проведення дослідження користувачів ми отримали 50 відкритих описів стовпчикових діаграм від 24 учасників дослідження. Згідно з документацією OpenAI [51], для тонкого налаштування моделі необхідний датасет розміром у декілька сотень прикладів, який має відкритий опис та відповідний код візуалізації.

У наступних розділах ми представляємо огляд використаних метод для синтетичного розширення датасету. Ми також надаємо детальну інформацію про особливості впровадження.

2.2.1. Метод

В першу чергу нам було потрібно вручну створити діаграми до кожного з 50 відкритих описів, щоб отримати 50 пар відкритого голосового опису та відповідних кодів діаграм. Далі ми випадковим чином розділили цей набір на 25 пар для тренування і 25 пар для тестування, щоб уникнути витіку тренувальних даних у тестовий набір.

Для розширення і балансування датасету зібраних описів, ми вирішили додати нові домени та елементи діаграм. Для розширення доменів було використано стовпчикові діаграми з набору даних NVBench [35]. Специфікація Plotly бібліотеки [52] використовувалася для охоплення 20 різних елементів гістограм. Генеруючи нові пари опису та коду, було важливо синхронізувати інформацію в описі та коді. Через те, що код легше контролювати, спочатку була запущена процедура для генерації нового коду з новими даними та елементами. Ця процедура випадковим чином обирала дані з NVBench і до 5 елементів діаграми зі специфікації Plotly бібліотеки, після чого у режимі

few-shot learning використовувалася велика мовна модель для генерації коду, який включатиме нові обрані дані та елементи.

Після цього на отриманих кодах була запущена процедура генерації опису. Ця процедура випадковим чином обирала відкритий голосовий опис з 25 описів для тренування, після чого у режимі few-shot learning використовувалася велика мовна модель для генерації нового опису, який зберігатиме структуру опису-прикладу, але при цьому буде описувати новий графік.

Останнім кроком була аугментація деяких описів, обраних випадковим чином, за допомогою зворотного перекладу, для стійкості набору даних до синонімів. Після чого згенеровані і реальні тренувальні дані було об'єднано, перемішано випадковим чином, і 20% даних було випадковим чином відділено для валідації. Таким чином було зібрано збалансований датасет (див. таблицю 2.3):

- тренувальний датасет: 412 пар описів і коду діаграм
- валідаційний датасет: 103 пар описів і коду діаграм
- тестовий датасет: 25 пар описів і коду діаграм

2.2.2. Впровадження

Зі специфікації Plotly ми обрали наступні елементи:

- orientation
- hover
- marker color
- background color
- grid lines color
- font color
- chart type
- text on bars

- caption
- plus signs pattern fill
- crosses pattern fill
- slashes pattern fill
- dots pattern fill
- font size
- rotated labels
- spacing
- base
- category order
- annotations
- range

Для кожного елементу ми прописали один або декілька прикладів Python Plotly коду в один рядок для промптингу моделі. Для генерації коду використовувалася модель GPT-3.5-turbo [53] і наступний шаблон промпу:

Act as a Visualization Expert. Based on the following data:

```
"x_data": [<values for the x axis>],  
"y_data": [<values for the y axis>],  
"classify": [<values for the classification>],  
"describe": <values for the grouping>,  
"nl_query": <short typed natural language query>
```

Generate a single one-line Python code that combines the <list of the elements> bar chart properties from these example codes:

```
<code example for each element>
```

Return only a single one-line Python code without your explanations, "import", and ".show()". Use `<if go.Figure in any example, then go.Figure, else px.bar>` method:

Для генерації опису використовувалася модель GPT-3.5-turbo [53] і наступний шаблон промпту:

Act as a person who describes the chart using voice. For example, this person has read this bar chart code:

`<bar chart code example from the training set of 25>`

And returned this example description:

`<bar chart free-form description example from the training set of 25>`

Repeat the structure, the format, and the style of the example description with disfluencies and filler words, but transform it to describe a new bar chart code:

`<new code generated from the previous step>`

Return only a new bar chart description, which should be almost like the example description, but for the new bar chart data and properties. Mention the `<list of the elements>` bar chart properties from the new bar chart code in the style of this person:

2.3. Висновки до розділу 2

Розділ надав важливу інформацію щодо побудови датасету відкритих голосових описів і відповідних до них гістограм у вигляді коду.

В секції 2.1 було описано дослідження користувачів для визначення структури відкритого голосового опису гістограми.

Далі, у секції 2.2 було описано методи синтетичного збільшення набору даних, який був отриманий з дослідження користувачів.

Цей розділ надав важливу інформацію про генерацію набору даних відкритих голосових описів, який можна використовувати для тренування моделей машинного навчання що генерують гістограми на основі відкритого голосового опису. В наступному розділі ми опишемо тренування моделей машинного навчання на отриманому датасеті.

| Назва | Тип |
|---------------------------------------|--------------------------|
| Назва | Елемент діаграми |
| Тип діаграми | Елемент діаграми |
| Мітка для осі | Елемент діаграми |
| Вісь | Елемент діаграми |
| Шкала | Елемент діаграми |
| Одиниці виміру | Елемент діаграми |
| Стовпчик | Елемент діаграми |
| Значення | Елемент діаграми |
| Інша форма | Елемент діаграми |
| Підпис | Елемент діаграми |
| Посилання на попередній(і) елемент(и) | Формат команди |
| Ітеративна команда | Формат команди |
| Попросіть поради | Формат команди |
| Орієнтація | Характеристика елемента |
| Колір | Характеристика елемента |
| Розмір | Характеристика елемента |
| Позиція | Організація елементів |
| Порядок | Організація елементів |
| Кількість | Організація елементів |
| Виправлення | Лінгвістична особливість |
| Повторення | Лінгвістична особливість |
| Метаопис | Лінгвістична особливість |
| Закінчити фразу | Лінгвістична особливість |

Таблиця 2.2

Структура опису діаграми у довільній формі

| Контекст | Опис | Код |
|--|--|---|
| x_data: ['Steven', 'Alexander', ... | So, um, so yeah, so I see a plot ... | go.Figure(data=[go.Bar(x= ['Steven', ... |
| x_data: ['Chesham', 'Amersham', ... | So in this, we can have a plot ... | go.Figure(data=[go.Bar(x= ['Chesham', ... |
| x_data: ['Lisa', 'Sarah'] ... | Okay, so we have a vertical bar chart ... | go.Figure(data=[go.Bar(x= ['Lisa', ... |
| x_data: ['Andrew', 'Charles', ... | Okay, so we have a vertical bar chart ... | go.Figure(data=[go.Bar(name= "Player's ... |
| How many users does Facebook ... | Okay, this time let's do a horizontal bar chart ... | px.bar(y= ['Q3 2012', 'Q2 2020', ... |
| x_data: ['50 Marshall', '71 St. Joseph', ... | Okay, in this chart, let's do a vertical bar chart ... | go.Figure(data=[go.Bar(x= ['50 Marshall', ... |
| ... | ... | ... |

Таблиця 2.3

Побудований датасет

РОЗДІЛ 3.

СИСТЕМА ДЛЯ ГЕНЕРАЦІЇ ДІАГРАМ НА ОСНОВІ ГОЛОСОВОГО ВВОДУ

Для впровадження ефективного голосового інтерфейсу для створення діаграм, який відповідає потребам користувачів, необхідно спершу визначити вимоги до системи, побудувати і навчити моделі машинного навчання для генерації діаграм на основі відкритого опису, протестувати моделі і порівняти метрики на побудованому раніше датасеті, який відображає вимоги користувачів, а також розробити кінцевий функціонал інтерфейсу.

3.1. Вимоги до системи

Для того, щоб задовольнити запити користувачів на основі результатів дослідження користувачів, були визначені наступні вимоги до системи генерації діаграм:

1) Система повинна бути здатна розпізнавати неявні описи та інтерпретувати їх у контексті з конкретними елементами та відкритими даними, включаючи назву, мітку, одиниці виміру, значення та елементи підписів.

2) Система повинна мати великий словник синонімів для назв елементів діаграм, таких як гістограма/діаграма, стовпчик/колонка/категорія тощо. Це необхідно для того, щоб врахувати різні стратегії користувачів при визначенні елементів діаграми.

3) Система повинна мати можливість автоматично визначати, чи шукає користувач пораду щодо правильного дизайну та маркування діаграми у випадках, коли користувач не надав цієї інформації, наприклад, щодо орієнтації, підпису, масштабу та одиниць виміру.

4) Система повинна мати можливість підраховувати кількість і запам'ятовувати порядок, значення і розмір елементів, наведених в описі, щоб зрозуміти конкретний елемент, на який посилається користувач, наприклад, з першим стовпчиком або значеннями, які збільшилися майже вдвічі.

5) Система повинна бути здатна отримувати контекстну інформацію з висновку та загальнодоступних даних в Інтернеті, таких як поточний квартал або початок пандемії, щоб підвищити точність і релевантність діаграми.

6) Система повинна дозволяти використовувати різні форми при створенні гістограм, такі як лінії, стрілки та малюнки. Це необхідно для того, щоб врахувати вподобання користувача та його творчі здібності.

7) Система повинна мати можливість розпізнавати виправлення та повторення в описах і відповідно коригувати дизайн діаграми.

8) Система повинна вміти розрізняти метаописи та звичайні описи. Це важливо для забезпечення належного маркування діаграм і розрізнення описів, які надають важливу інформацію про діаграму, від тих, що її не надають.

9) Система повинна дозволяти користувачам організовувати розташування елементів відносно один одного. Це важливо для того, щоб користувачі могли налаштовувати діаграму відповідно до своїх уподобань.

10) Система повинна підтримувати ітеративні команди, щоб користувачі могли змінювати та коригувати елементи діаграми за потреби.

Разом ці вимоги гарантують, що система генерації діаграм буде здатна розпізнавати та ефективно інтерпретувати вхідні дані користувача, а також забезпечувати гнучкість та кастомізацію, необхідні для задоволення різноманітних користувацьких потреб. Згенерований датасет у попередньому розділі відображає ці вимоги.

3.2. Побудова моделей машинного навчання

Захоплення семантичного контексту опису діаграми може бути складним завданням, головним чином, коли маємо справу із зашумленими і неоднозначними запитами і доменами, не включеними до набору даних. Танг та ін. [6] підкреслили це обмеження, припускаючи, що розширення набору даних може бути потенційним рішенням. Однак альтернативою може бути підхід навчання з кількох спроб, як, наприклад, у прикладі Plotly Dash [54].

Використовуючи велику мовну модель (LLM), навчену на широкому наборі даних з різних областей, ми можемо генерувати код діаграм на основі простих описів діаграм у режимі навчання за кілька пострілів. LLM також може бути точно налаштована на конкретні домени або запити за допомогою декількох сотень прикладів пар <опис, код діаграми> [55].

Ми розробили три великих мовних моделей машинного навчання на основі архітектури Transformer:

- 1) Few-shot Learning GPT-J-6B модель
- 2) Few-shot Learning GPT-3 Davinci 175B модель
- 3) Fine-tuned GPT-3 Davinci 175B модель

3.2.1. Few-shot Learning GPT-J-6B модель

Після того, як користувач задоволений транскрибованим описом діаграми, LLM перекладає опис у код, який будує і будує діаграму. Однак, емпірично ми виявили, що LLM іноді потрібно наздоганяти елементи діаграми, наприклад, заголовок та назви осей. Щоб вирішити цю проблему, ми вирішили використати додаткову модель запитань-відповідей [56] в описі діаграми, щоб уточнити конкретні значення та інтегрувати їх у згенерований код з LLM.

На основі чотирирівневої моделі семантичного змісту [8], ми використали наступні питання для уточнення конкретних значень:

- 1) "Яка найкраща назва для цієї діаграми, описаної в тексті?"
- 2) "Що відкладено на горизонтальній осі x на цій діаграмі, описаній у тексті?"
- 3) "Що відкладено на горизонтальній осі x на цій діаграмі, описаній у тексті?"
- 4) "Що відкладено на вертикальній осі y цієї діаграми, описаної в тексті?"
- 5) "Що відкладено на вертикальній осі y цієї діаграми, описаної в тексті?"
- 6) "Який діапазон значень на горизонтальній осі x для цієї діаграми, описаної в тексті?"
- 7) "Який діапазон значень на горизонтальній осі x для цієї діаграми, описаної в тексті?"
- 8) "Який діапазон значень (від до) на вертикальній осі y для цієї діаграми, описаної в тексті?"
- 9) "Який діапазон значень (від до) на вертикальній осі y для цієї діаграми, описаної в тексті?"
- 10) "Які кольори найкраще підходять для цієї діаграми, описаної в тексті?"
- 11) "Яке середнє значення на вертикальній осі y для цієї діаграми?"
- 12) "Яке середнє або середнє значення даних цієї діаграми, описаної в тексті?"
- 13) "Яке середнє або середнє значення даних цієї діаграми, описаної в тексті?"
- 14) "Яке максимальне або найбільше значення даних цієї діаграми, описаної в тексті?"
- 15) "Яке мінімальне або найбільше значення даних цієї діаграми, описаної в тексті?"
- 16) "Яке мінімальне або мінімальне значення даних цієї діаграми, описаної в тексті?"

17) "Яке максимальне або мінімальне значення даних цієї діаграми, описаної в тексті?"

18) "Який тип кореляції або коли збільшуються або зменшуються дані на цьому графіку, описаному в тексті?"

19) "Який тип кореляції або коли збільшуються або зменшуються дані на цьому графіку, описаному в тексті?"

20) "Які тенденції, закономірності, винятки або концепції можна знайти на цій діаграмі, описаній у тексті?"

21) "Які тенденції, закономірності, винятки або концепції можна знайти на цій діаграмі?"

22) "Які специфічні для домену уявлення, поточні події, соціальний та політичний контекст і пояснення, пов'язані з цією діаграмою, описано в тексті?"

Щоб інтегрувати конкретні значення, з'ясовані за допомогою моделі запитань-відповідей, у згенерований код з LLM, ми можемо використовувати умовні оператори, які перевіряють наявність кожного значення і відповідно включають його в код. Наприклад, якщо відповідь на перше запитання "Яка найкраща назва для цієї діаграми, описаної в тексті?" "Показники продажів за регіонами" ми можемо використати оператор `if`, щоб перевірити, чи вже було додано назву до коду діаграми. Якщо ні, ми можемо додати назву до коду, використовуючи рядок "Показники продажів за регіонами" як назву діаграми.

Крім того, для автоматичного виправлення граматичних і пунктуаційних помилок використовуються граматичні та пунктуаційні моделі [57; 58]. Для завдання "запитання-відповідь" використовується модель `roberta-base-squad2` [56]. Для генерації коду Plotly [59] використовується велика модель мови GPT-J-6B, що навчається за кілька пострілів [60].

3.2.2. Few-shot Learning GPT-3 Davinci 175B модель

Для генерації коду Plotly [59] використовується велика модель мови GPT-3 Davinci (175B параметрів), що навчається за кілька пострілів [61].

3.2.3. Fine-tuned GPT-3 Davinci 175B модель

Для генерації коду була натренована велика модель мови GPT-3 Davinci (175B параметрів) [61]. Тренування проходило на тренувальному наборі з валідацією на валідаційному наборі згенерованих даних з попереднього розділу. Гіперпараметри були підібрані емпіричним шляхом:

- Epoch = 1
- Batch size = 1
- Learning Rate = 0.1

Отримана модель не потребує навчання за кілька пострілів. Замість цього достатньо задати контекст і опис за наступним шаблоном:

```
<context>\n\nDescription:\n<description>\n\nCode:\n
```

Навчена модель продовжить відповідати згенерованим кодом до цього опису і завершить генерацію за досягнення стоп-слова:

```
)\n\n\n
```

3.3. Тестування розроблених моделей

3.3.1. Метод

Для оцінки ефективності моделей ми порахували наступні метрики на тестовому наборі даних:

- BLEU
- ROUGE-L (F)
- BERTScore
- Час

Метрики, які ми обрали для оцінки ефективності наших моделей, використовуються в загальних випадках для порівняння різних моделей машинного навчання, що вирішують задачі генерації тексту або перекладу.

3.3.2. Результати

Результати оцінки показали, що Fine-tuned GPT-3 Davinci 175B модель перевершила як GPT-J-6B, так і GPT-3 Davinci 175B за показниками BLEU та BERTScore, а також мала досить конкурентоспроможний час відгуку (див. таблицю 3.1).

| Метрика | GPT-J-6B | GPT-3 Davinci 175B | Fine-tuned Davinci |
|---------------|---------------|--------------------|--------------------|
| BLEU ↑ | 0.0292 | 0.0058 | 0.077 |
| ROUGE-L (F) ↑ | 0.1325 | 0.1409 | 0.2659 |
| BERTScore ↑ | 0.8839 | 0.8495 | 0.8867 |
| Time ↓ | 4.3641 | 14.1058 | 5.6765 |

Таблиця 3.1

Порівняння моделей

У Fine-tuned Davinci найвищий показник BLEU - 0,077, тоді як у GPT-J-6B - 0,0292, а у GPT-3 Davinci 175B - 0,0058. Крім того, Fine-tuned Davinci отримав найвищий показник BERTScore - 0,8867, порівняно з GPT-J-6B - 0,8839 та GPT-3 Davinci 175B - 0,8495. За показником ROUGE-L (F) GPT-3 Davinci 175B перевершив GPT-J-6B з результатом 0,1409 проти 0,1325. Однак обидва були перевершені Fine-tuned GPT-3 Davinci 175B моделлю, яка досягла показника ROUGE-L (F) 0,2659.

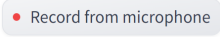
Що стосується часу відгуку, то GPT-J-6B виявився найшвидшим із середнім часом 4,3641 секунди, тоді як Fine-tuned Davinci мав середній час 5,6765 секунди, а GPT-3 Davinci 175B - 14,1058 секунди. Цікаво, що, незважаючи на високі показники, Fine-tuned GPT-3 Davinci 175B був трохи повільнішим за GPT-J-6B, але все ж значно швидшим за GPT-3 Davinci 175B.

3.4. Функціонал кінцевого інтерфейсу

На основі результатів дослідження користувачів і побудованої моделі, ми оновили розробили прототип діалогового інтерфейсу, який дозволяє користувачам створювати діаграми за допомогою голосового введення. Основна мета цієї системи - функціонувати як тестова платформа для оцінки голосового підходу до створення діаграм.

3.4.1. Опис діаграми голосом

У робочому процесі системи людина може легко записати голосовий опис діаграми за допомогою мікрофону (див. рис. 3.1). Запис автоматично перетворюється на текст, який може бути відредагований користувачем. Якщо потрібно, користувач також може перезаписати опис діаграми.

Ми вирішили використати голос як інтерфейс для нашої системи створення діаграм через його потенційні переваги, про які свідчать існуючі дослідження. Наприклад, Такер та ін. [62] пояснюють, що мова має потенціал для покращення зручності використання комп'ютерних систем. Більше того, голосовий інтерфейс також використовувався в системі Sevi [6], що підкреслює його ефективність. Наш інтерфейс спроектовано в мінімалістичному стилі, він складається з кнопки запису , анімації запису та текстового поля.

The Voice Interface for Chart Creation

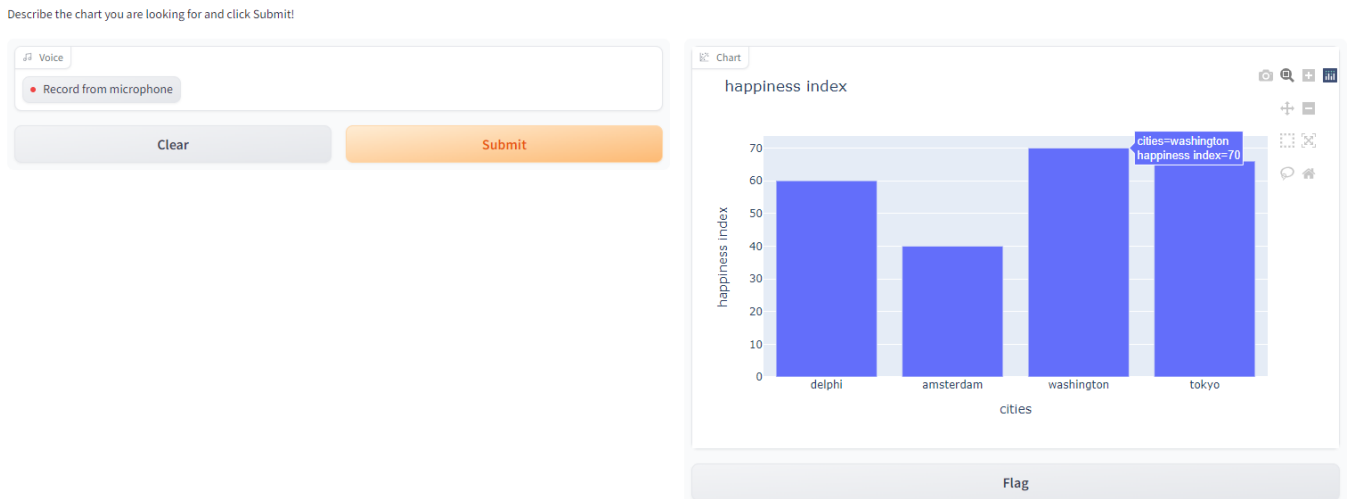


Рис. 3.1. Погляд користувача на функції запису голосу та інтерактивної діаграми

Таким чином, він не повинен вимагати попереднього навчання для користувачів, які мають досвід роботи з диктофонами та сучасними веб-сайтами.

Хоча поточний прототип підтримує лише описи англійською мовою, з появою надійних багатомовних моделей розпізнавання голосу існує можливість його екстраполяції на інші мови.

3.4.2. Експорт візуалізації

За прикладом Plotly Dash [54], запропонована система представляє користувачам інтерактивну діаграму, яка дозволяє маніпулювати масштабом діаграми, наводити курсор на точки даних для отримання конкретних значень, а також зберігати діаграму у вигляді растрового зображення. Інтерактивна діаграма будується шляхом виконання попередньо згенерованого коду, а отримана діаграма відображається у графічному інтерфейсі користувача системи.

3.4.3. Реалізація

Система використовує Conformer-1 API [32; 63] для автоматичного розпізнавання англійської мови. Для генерації коду Plotly [59] використовується Fine-tuned GPT-3 Davinci 175B модель. Інтерактивний інтерфейс Gradio [64] використовується для зручного аудіозапису та представлення вихідного графіка у вигляді коду візуалізації графіків JavaScript, згенерованого рушієм візуалізації Plotly. Час обробки завдання генерації графіка становить до 60 секунд, залежно від тривалості аудіозапису. Система реалізована мовою програмування Python. Система може розпізнавати голос, розуміти контекст опису, генерувати діаграми з відповідними даними, елементами, заголовками та назвами осей, а також експортувати інтерактивний графік.

3.5. Висновки до розділу 3

Розділ описав створення системи для генерації діаграм на основі голосового вводу, а саме вимоги до системи на основі дослідження користувачів, побудову математичних моделей машинного навчання для задоволення поставлених вимог, оцінку побудованих моделей, та функціонал кінцевого інтерфейсу.

У секції 3.1 було представлено вимоги до системи на основі дослідження користувачів. Задовільнивши поставлені вимоги, ми розробимо голосовий інтерфейс для створення діаграм, який відповідає потребам користувачів.

В секції 3.2 була розглянута побудова і навчання моделей машинного навчання на основі архітектури Transformer для генерації коду діаграми на основі голосового опису.

Далі, у секції 3.3 було проведено тестування розроблених моделей на побудованому раніше датасеті, який відображає вимоги користувачів. Результати

тестування показали що навчена модель на побудованому датасеті має кращі метричні показники у порівнянні з іншими моделями.

Нарешті, у секції 3.4 було представлено функціонал кінцевого інтерфейсу системи для генерації діаграм на основі голосового вводу.

Цей розділ показав як результати дослідження користувачів і методи машинного навчання були використані для створення системи для генерації діаграм на основі голосового вводу. Отримані результати демонструють що побудована система є ефективним, згідно з метриками, голосовим інтерфейсом для створення діаграм, який відповідає потребам користувачів. В наступному розділі ми опишемо висновки до всієї роботи.

ВИСНОВКИ

У даній роботі було зроблено важливий внесок у розуміння та реалізацію голосових інтерфейсів для генерації діаграм.

У першому розділі було проаналізовано існуючі голосові інтерфейси, дасети та методи машинного навчання для генерації графіків. Також були детально розглянуті метрики для оцінки генерації тексту, які можна використовувати для оцінки роботи розроблених моделей.

У другому розділі було проведено розвідувальне дослідження, щоб визначити структуру голосової взаємодії, охарактеризувати висловлювання відповідно до їхніх фраз та інформації, яку вони містять, а також виявити прогалини. Крім того, було створено новий набір даних відкритих голосових описів, який потім був використаний для тренування моделей машинного навчання. Це дало можливість розробити моделі, які можуть генерувати діаграми на основі відкритого голосового опису.

У третьому розділі було створено прототип розмовного інтерфейсу користувача для створення графіків за допомогою голосового введення, що використовує результати дослідження користувачів і результати тренування моделей машинного навчання для генерації діаграм. Оцінка за допомогою метрик підтвердили ефективність такої системи.

Висновок з роботи полягає в тому, що система для генерації діаграм на основі голосового вводу, створена з використанням методів машинного навчання, відповідає потребам користувачів і має високу ефективність згідно з метриками. Побудована система може бути використана для створення діаграм у різних областях, що значно полегшує процес візуалізації даних за допомогою голосового інтерфейсу.

ПЕРЕЛІК ПОСИЛАНЬ

1. Battle L., Scheidegger C. A Structured Review of Data Management Technology for Interactive Visualization and Analysis // IEEE Transactions on Visualization and Computer Graphics. — 2021. — Т. 27, № 2. — С. 1128—1138. — DOI: 10.1109/TVCG.2020.3028891.
2. Amyrotos C. Adaptive Visualizations for Enhanced Data Understanding and Interpretation // Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization. — Utrecht, Netherlands : Association for Computing Machinery, 2021. — С. 291—297. — (UMAP '21). — ISBN 9781450383660. — DOI: 10.1145/3450613.3459657.
3. Myers B. A., Goldstein J., Goldberg M. A. Creating Charts by Demonstration // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. — Boston, Massachusetts, USA : Association for Computing Machinery, 1994. — С. 106—111. — (CHI '94). — ISBN 0897916506. — DOI: 10.1145/191666.191715.
4. Communicating Visualizations without Visuals: Investigation of Visualization Alternative Text for People with Visual Impairments / C. Jung [та ін.] // IEEE Transactions on Visualization and Computer Graphics. — 2022. — Т. 28, № 1. — С. 1095—1105. — DOI: 10.1109/TVCG.2021.3114846.
5. Towards Natural Language Interfaces for Data Visualization: A Survey / L. Shen [та ін.] // IEEE Transactions on Visualization and Computer Graphics. — 2022. — С. 1—1. — DOI: 10.1109/TVCG.2022.3148007.
6. Sevi: Speech-to-Visualization through Neural Machine Translation / J. Tang [та ін.] // Proceedings of the 2022 International Conference on Management of Data. — Philadelphia, PA, USA : Association for Computing Machinery, 2022. — С. 2353—2356. — (SIGMOD '22). — ISBN 9781450392495. — DOI: 10.1145/3514221.3520150.

7. Multimodal Presentation of Two-Dimensional Charts: An Investigation Using Open Office XML and Microsoft Excel / I. Abu Doush [та ін.] // ACM Trans. Access. Comput. — New York, NY, USA, 2010. — Листоп. — Т. 3, № 2. — ISSN 1936-7228. — DOI: 10.1145/1857920.1857925.
8. Lundgard A., Satyanarayan A. Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content // IEEE Transactions on Visualization and Computer Graphics. — 2022. — Т. 28, № 1. — С. 1073—1083. — DOI: 10.1109/TVCG.2021.3114770.
9. Collecting and Characterizing Natural Language Utterances for Specifying Data Visualizations / A. Srinivasan [та ін.] // Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. — 2021.
10. Robots for Inclusive Play: Co-Designing an Educational Game With Visually Impaired and Sighted Children / O. Metatla [та ін.] // Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. — New York, NY, USA : Association for Computing Machinery, 2020. — С. 1—13. — ISBN 9781450367080.
11. Making GIFs Accessible / C. Gleason [та ін.] // The 22nd International ACM SIGACCESS Conference on Computers and Accessibility. — Virtual Event, Greece : Association for Computing Machinery, 2020. — (ASSETS '20). — ISBN 9781450371032. — DOI: 10.1145/3373625.3417027.
12. Stangl A., Morris M. R., Gurari D. "Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions // Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. — New York, NY, USA : Association for Computing Machinery, 2020. — С. 1—13. — ISBN 9781450367080.
13. Image Explorer: Multi-Layered Touch Exploration to Make Images Accessible / J. Lee [та ін.] // The 23rd International ACM SIGACCESS Conference

on Computers and Accessibility. — Virtual Event, USA : Association for Computing Machinery, 2021. — (ASSETS '21). — ISBN 9781450383066. — DOI: 10.1145/3441852.3476548.

14. Engel C., Müller E. F., Weber G. SVGPlott: An Accessible Tool to Generate Highly Adaptable, Accessible Audio-Tactile Charts for and from Blind and Visually Impaired People // Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments. — Rhodes, Greece : Association for Computing Machinery, 2019. — C. 186—195. — (PETRA '19). — ISBN 9781450362320. — DOI: 10.1145/3316782.3316793.
15. Leveraging Text-Chart Links to Support Authoring of Data-Driven Articles with VizFlow / N. Sultanum [та ил.] // Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. — Yokohama, Japan : Association for Computing Machinery, 2021. — (CHI '21). — ISBN 9781450380966. — DOI: 10.1145/3411764.3445354.
16. Conversations with Documents: An Exploration of Document-Centered Assistance / M. ter Hoeve [та ил.] // Proceedings of the 2020 Conference on Human Information Interaction and Retrieval. — New York, NY, USA : Association for Computing Machinery, 2020. — C. 43—52. — ISBN 9781450368926.
17. Kim D. H., Hoque E., Agrawala M. Answering Questions about Charts and Generating Visual Explanations // Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. — New York, NY, USA : Association for Computing Machinery, 2020. — C. 1—13. — ISBN 9781450367080.
18. Answering Questions about Data Visualizations using Efficient Bimodal Fusion / K. Kafle [та ил.] // 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). — 2020. — C. 1487—1496.

19. Making Data Visualization More Efficient and Effective: A Survey / X. Qin [та ін.] // The VLDB Journal. — Berlin, Heidelberg, 2020. — Січ. — Т. 29, № 1. — С. 93—117. — ISSN 1066-8888. — DOI: 10.1007/s00778-019-00588-3.
20. Yu B., Silva C. T. FlowSense: A Natural Language Interface for Visual Data Exploration within a Dataflow System // IEEE Transactions on Visualization and Computer Graphics. — 2020. — Т. 26. — С. 1—11.
21. TaskVis: Task-oriented Visualization Recommendation / L. Shen [та ін.] // EuroVis 2021 - Short Papers / за ред. М. Agus, С. Garth, А. Kerren. — The Eurographics Association, 2021. — ISBN 978-3-03868-143-4. — DOI: 10.2312/evs.20211061.
22. VisGuide: User-Oriented Recommendations for Data Event Extraction / Y.-R. Cao [та ін.] // Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. — New Orleans, LA, USA : Association for Computing Machinery, 2022. — (CHI '22). — ISBN 9781450391573. — DOI: 10.1145/3491102.3517648.
23. VisGNN: Personalized Visualization Recommendation via Graph Neural Networks / F. Ojo [та ін.] // Proceedings of the ACM Web Conference 2022. — Virtual Event, Lyon, France : Association for Computing Machinery, 2022. — С. 2810—2818. — (WWW '22). — ISBN 9781450390965. — DOI: 10.1145/3485447.3512001.
24. Vega-Lite: A Grammar of Interactive Graphics / A. Satyanarayan [та ін.] // IEEE Transactions on Visualization and Computer Graphics. — USA, 2017. — Січ. — Т. 23, № 1. — С. 341—350. — ISSN 1077-2626. — DOI: 10.1109/TVCG.2016.2599030.
25. Midway S. R. Principles of Effective Data Visualization // Patterns. — 2020. — Т. 1.

26. Shahira K. C., Lijiya A. Towards Assisting the Visually Impaired: A Review on Techniques for Decoding the Visual Data From Chart Images // IEEE Access. — 2021. — Т. 9. — С. 52926—52943. — DOI: 10.1109/ACCESS.2021.3069205.
27. Zhou Y. P., Tan C. L. Hough technique for bar charts detection and recognition in document images // Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101). Т. 2. — 2000. — 605—608 vol.2. — DOI: 10.1109/ICIP.2000.899506.
28. LeCun Y., Bengio Y., Hinton G. Deep Learning // Nature. — 2015. — Трaв. — Т. 521. — С. 436—44. — DOI: 10.1038/nature14539.
29. LeCun Y., Bengio Y. Convolutional Networks for Images, Speech, and Time Series // The Handbook of Brain Theory and Neural Networks. — Cambridge, MA, USA : MIT Press, 1998. — С. 255—258. — ISBN 0262511029.
30. Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. — 1997. — Т. 9. — С. 1735—1780.
31. Generating Accurate Caption Units for Figure Captioning / X. Qian [та ил.] // Proceedings of the Web Conference 2021. — Ljubljana, Slovenia : Association for Computing Machinery, 2021. — С. 2792—2804. — (WWW '21). — ISBN 9781450383127. — DOI: 10.1145/3442381.3449923.
32. Attention is All You Need / A. Vaswani [та ил.] // Proceedings of the 31st International Conference on Neural Information Processing Systems. — Long Beach, California, USA : Curran Associates Inc., 2017. — С. 6000—6010. — (NIPS'17). — ISBN 9781510860964.
33. Chart-to-Text: A Large-Scale Benchmark for Chart Summarization / S. Kanthara [та ил.]. — 2022. — Бep.

34. Hsu T.-Y., Giles C. L., Huang T.-H. SciCap: Generating Captions for Scientific Figures // Findings of the Association for Computational Linguistics: EMNLP 2021. — Punta Cana, Dominican Republic : Association for Computational Linguistics, 11.2021. — C. 3258—3264.
35. Luo Y., Tang J., Li G. nvBench: A Large-Scale Synthesized Dataset for Cross-Domain Natural Language to Visualization Task // ArXiv. — 2021. — T. abs/2112.12926.
36. Bleu: a Method for Automatic Evaluation of Machine Translation / K. Papineni [та ит.] // Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. — Philadelphia, Pennsylvania, USA : Association for Computational Linguistics, 07.2002. — C. 311—318. — DOI: 10.3115/1073083.1073135.
37. Lin C.-Y. ROUGE: A Package for Automatic Evaluation of Summaries // Text Summarization Branches Out. — Barcelona, Spain : Association for Computational Linguistics, 07.2004. — C. 74—81.
38. BERTScore: Evaluating Text Generation with BERT / T. Zhang [та ит.]. — 2020. — arXiv: 1904.09675 [cs.CL].
39. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding J. Devlin [та ит.]. — 2019. — arXiv: 1810.04805 [cs.CL].
40. Narechania A., Srinivasan A., Stasko J. T. NL4DV: A Toolkit for Generating Analytic Specifications for Data Visualization from Natural Language Queries // IEEE Transactions on Visualization and Computer Graphics. — 2021. — T. 27. — C. 369—379.
41. ADVISor: Automatic Visualization Answer for Natural-Language Question on Tabular Data / C. Liu [та ит.] // 2021 IEEE 14th Pacific Visualization Symposium (PacificVis). — 2021. — C. 11—20.

42. Text-to-Viz: Automatic Generation of Infographics from Proportion-Related Natural Language Statements / W. Cui [та ін.] // IEEE Transactions on Visualization and Computer Graphics. — 2020. — Т. 26. — С. 906—916.
43. Srinivasan A., Setlur V. Snowy: Recommending Utterances for Conversational Visual Analysis // The 34th Annual ACM Symposium on User Interface Software and Technology. — Virtual Event, USA : Association for Computing Machinery, 2021. — С. 864—880. — (UIST '21). — ISBN 9781450386357. — DOI: 10.1145/3472749.3474792.
44. A Multi-Modal Natural Language Interface to an Information Visualization Environment / К. Cox [та ін.] // International Journal of Speech Technology. — 2001. — Лип. — Т. 4. — С. 297—314. — DOI: 10.1023/A:1011368926479.
45. Breiman L. Random forests // Machine learning. — 2001. — Т. 45, № 1. — С. 5—32.
46. Text2Chart: A Multi-Staged Chart Generator from Natural Language Text / M. M. Rashid [та ін.] // PAKDD. — 2022.
47. Language Models are Few-Shot Learners / Т. B. Brown [та ін.]. — 2020. — arXiv: 2005.14165 [cs.CL].
48. Statista. Statista - The Statistics Portal for Market Data, Market Research and Market Studies. — 2007. — [Accessed: April 17, 2023]. <https://www.statista.com>.
49. Braun V., Clarke V. Using thematic analysis in psychology // Qualitative research in psychology. — 2006. — Т. 3, № 2. — С. 77—101.
50. Corbin J., Strauss A. Basics of qualitative research: Techniques and procedures for developing grounded theory. — Sage publications, 2015.
51. OpenAI. Fine-Tuning Guide. — 2023. — Accessed: 2023-06-13.
52. Plotly. Bar Charts in Python. — 2023. — Accessed: 2023-06-13.

53. OpenAI. GPT-3.5 Turbo. — 2023. — Accessed: 2023-06-13.
54. Inc. P. T. Dash GPT-3 Bar Charts Generation. — 2020. — URL: <https://github.com/plotly/dash-sample-apps/tree/main/apps/dash-gpt3-bars>.
55. OpenAI. Fine-Tuning GPT-3. — 2021. — [Accessed: April 15, 2023]. <https://platform.openai.com/docs/guides/fine-tuning>.
56. RoBERTa: A Robustly Optimized BERT Pretraining Approach / Y. Liu [та ін.] // CoRR. — 2019. — Т. abs/1907.11692. — arXiv: 1907.11692.
57. developers L. LanguageTool - Online Grammar, Style Spell Checker. — 2023. — Accessed April 15, 2023. <https://languagetool.org/>.
58. FullStop: Multilingual Deep Models for Punctuation Prediction / O. Guhr [та ін.]. — 2021. — Черв.
59. Inc. P. T. Collaborative data science. — 2015. — URL: <https://plot.ly>.
60. Wang B., Komatsuzaki A. GPT-J-6B: A 6 Billion Parameter Autoregressive Language Model. — 05.2021. — <https://github.com/kingoflolz/mesh-transformer-jax>.
61. OpenAI. GPT-3 Model Documentation. — 2023.
62. Tucker P., Jones D. M. Voice as interface: An overview // International Journal of Human-Computer Interaction. — 1991. — Т. 3, № 2. — С. 145—170. — DOI: 10.1080/10447319109526002. — eprint: <https://doi.org/10.1080/10447319109526002>.
63. Wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations / A. Baevski [та ін.] // Proceedings of the 34th International Conference on Neural Information Processing Systems. — Vancouver, BC, Canada : Curran Associates Inc., 2020. — (NIPS'20). — ISBN 9781713829546.
64. Gradio. Build Share Delightful Machine Learning Apps. — 2021. — URL: <https://gradio.app>.

ДОДАТКИ

Додаток А. Матеріали для дослідження користувачів

1. How many users does Facebook have? How many users does Facebook have? With over 2.7 billion monthly active users as of the second quarter of 2020, Facebook is the biggest social network worldwide. In the third quarter of 2012, the number of active Facebook users surpassed one billion, making it the first social network ever to do so. Active users are those who have logged in to Facebook during the last 30 days. During the last reported quarter, the company stated that 3.14 billion people were using at least one of the company's core products (Facebook, WhatsApp, Instagram, or Messenger) each month.

2. PUBG Players Map First released at the start of 2017, PlayerUnknown's Battlegrounds (PUBG) has quickly gained a legion of followers across the world, most notably in the United States. Almost one-quarter of PUBG's players base was in the United States, with a further 19 percent originating in China and 6 percent in Germany. The Battle Royale game, developed by Bluehole, pits up to 100 online players against each other in a fight to the death until only one player or one team is left standing.

3. How much is minimum wage? How much is minimum wage? The federally mandated minimum wage in the United States is 7.25 U.S. dollars per hour, although the minimum wage varies from state to state. As of January 1, 2020, the District of Columbia had the highest minimum wage in the U.S., at 14 U.S. dollars per hour. This was followed by Washington state, which had 13.5 U.S. dollars per hour and California state, which had 13 U.S. dollars per hour as the state minimum wage.

4. Import trend in India In fiscal 2020, India imported over 2.8 trillion Indian rupees worth of electrical and non-electrical equipment. In 2011, the amount was more than 1.2 trillion Indian rupees. Since fiscal 2011, there has been an upward

trend in the value of imports in this sector every year: 1.6 trillion Indian rupees in 2013, 1.9 trillion Indian rupees in 2016, and 2.2 trillion Indian rupees in 2018.

5. Twitch streams about Minecraft First developed by Markus Persson, Minecraft is a sandbox video game which officially released in November 2011. The game immerses players in a dynamically created world composed of blocks representing different materials such as earth, sand, stone, water, lava, or minerals forming various structures and animals or monsters. Minecraft's events on the video streaming service Twitch were watched by an average of 44.52 thousand viewers in September 2019, 73.11 thousand viewers in July 2020, and 106.29 thousand viewers in November 2020.

6. Time spent with the patient About 33 percent of U.S. physicians spent 17 - 24 minutes, 29 percent of U.S. physicians spent 13 - 16 minutes, and 22 percent of U.S. physicians spent 9 - 12 minutes with their patients, according to a survey conducted in 2018. Physicians are often constrained in their time directly working with patients, which could have an impact on patient care outcomes. Studies found that physicians spend almost half of their time in the office on data entry and other desk work. More sophisticated, network-enabled EHR (electronic health records) systems for physicians could probably be a step towards more time directly with patients.

7. Gender gap in Ecuador In 2020, Ecuador scored 0.76, which shows a gender gap of approximately 24 percent (women are 24 percent less likely than men to have equal opportunities). In 2018, Ecuador scored 0.73, which shows a gender gap of approximately 27 percent. In 2017, Ecuador scored 0.72, which shows a gender gap of approximately 28 percent.

8. LinkedIn Android app users As of June 2020, it was found that only 6.5 percent of LinkedIn app users in the United States were heavy users, accessing the app more than 2/3 of a month. According to App Ape, 7.1 percent of U.S. LinkedIn audiences via Android were middle users, 40.5 percent were light users, and 45.7 percent were inactive.