

УДК 004

Бєлов К.В.¹, Степаненко О.О.²

¹ студ. гр. КНТ-111 НУ «Запорізька політехніка»

² доц. НУ «Запорізька політехніка»

ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ГОЛОСОВОГО ПОМІЧНИКА ДЛЯ РОБОТИ З СЕРВІСАМИ ШТУЧНОГО ІНТЕЛЕКТУ

Об'єкт дослідження: процес розробки програмного забезпечення для взаємодії з сервісами штучного інтелекту та операційною системою комп'ютера.

Предмет дослідження: програмні засоби, що забезпечують зручність використання функціоналу операційної системи та взаємодію з нейронними мережами.

Мета роботи: створення програмного забезпечення, що спрощує доступ до можливостей операційної системи та нейронних мереж шляхом реалізації голосового помічника.

У процесі виконання дослідження було детально проаналізовано розвиток настільних голосових асистентів і розглянуто сучасні технологічні фактори в цій галузі. Робота показує, що операційні системи можуть стати більш зручними для користувача, якщо користувач керуватиме системою за допомогою голосу, а додатки, що складаються з модулів ASR, TTS і NLP, стануть простішими і зручнішими. Було розглянуто декілька архітектур які є найбільш популярну на ринку серед існуючих рішень: Microsoft Copilot, Amazon Alexa, NVIDIA ChatRTX, Project G-Assist, після вивчення їхніх переваг і недоліків, і за аналогією було обрано найкращу архітектуру та модулі керування системою для голосового асистента Voice Helper.

Описані функціональні вимоги до програмного забезпечення голосового асистента. Було вирішено, що система повинна мати можливість розпізнавати

голосових команд з точністю 90%, надавати локальний синтез мови, застосовувати модульні функції для підвищення функціональності та відповідати вимогам конфіденційності та безпеки.

Обрано мову програмування Python для реалізації програмного забезпечення, підтримка якої відкритим вихідним кодом, коротка крива навчання та велика екосистема бібліотек роблять її придатною для швидкого прототипування, гнучкої розробки і кросплатформенного розвитку. Для обробки запитів до нейромережі використано нейронний сервіс DeepSeek, який має відкриту ліцензію і забезпечує швидке генерування відповідей без стягнення плати за API-запити.

Модуль розпізнавання мовлення автоматично реалізується за допомогою Vosk, може працювати офлайн і обробляти потокове аудіо без інтернету, забезпечуючи негайну відповідь. Генерація мовлення здійснюється за допомогою Silero TTS - невеликих локальних моделей, що гарантують природне звучання без ефекту робота з надзвичайно низькою затримкою.

Нейромережа DeepSeek була обрана для проекту як одна з найбільш кращих варіантів представлених на ринку. Це модель з малою затримкою відповіді, яка має функцію глибокого мислення над задачею. DeepSeek має можливість відправки безкоштовних запитів через ключ API, що дає змогу на постійній основі без фінансових витрат мати доступ до сучасних технологій.

Для взаємодії з найбільш важливою функцією, якою є зв'язок з мовними моделями штучного інтелекту, а саме озвучування відповіді від моделі, було обрано Silero TTS. Цей модуль має великі переваги в зручності його використання та застосування, та в якості звучання голосу. Саме ці переваги стали основними при виборі цієї моделі TTS для голосового помічника Voice Helper.

Запропоновано архітектуру програмного забезпечення, що складається з автономних програм для захоплення аудіо, STT, обробки NLP, TTS і взаємодії з операційною системою. Враховано роботу системи: запуск і закриття додатків, контроль гучності, відтворення і налаштування звукових файлів, розподіл веб-ресурсів, запити до нейронної мережі, а також показ вибіркового вихідного коду та журналів із системними повідомленнями.

Розроблено інтерфейс користувача. Програмне забезпечення було протестоване і відповідало визначеним вимогам, а операції були стабільними під час інтерактивної роботи. Результати тестування підтвердили належну реалізацію всіх функціональних випадків і забезпечили потокове відтворення мовлення та розпізнавання команд у реальному часі.

Розроблений Voice Helper забезпечує ефективну автоматизацію рутинних завдань на ПК, підвищує зручність користування системою та відкриває перспективи подальшого розвитку через інтеграцію додаткових модулів і адаптивне навчання під індивідуальні потреби користувача.

Розроблено голосового помічника, який дозволяє керувати операційною системою та отримувати доступ до функцій штучного інтелекту за допомогою голосових команд.

Практична цінність роботи: створена програма може використовуватись як персональний інтерфейс для полегшення взаємодії користувача з комп'ютером, що особливо корисно в контексті роботи з нейронними мережами, запуску програм та автоматизації задач.

Галузь використання: інтелектуальні користувацькі інтерфейси, системи автоматизації, навчальні платформи, персональні асистенти