

Автор:

Сторчак Андрій Олександрович
студент 4 курсу, 41 КН групи,

Науковий керівник:

Франчук Василь Михайлович,
доктор педагогічних наук, доцент,
завідувач кафедри комп'ютерної та програмної
інженерії

РОЗРОБЛЕННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ РОЗПІЗНАВАННЯ ГОЛОСОВИХ І ТЕКСТОВИХ КОМАНД ІЗ ВИКОРИСТАННЯМ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ

Анотація: Метою дослідження є створення програмного забезпечення для розпізнавання голосових і текстових команд із використанням моделей штучного інтелекту, що забезпечує високу точність розпізнавання, швидкість опрацювання та можливість самонавчання для підвищення ефективності та зручності використання комп'ютера. Завданням роботи є розроблення архітектури локального голосового асистента та імплементація інтелектуального ядра з інтеграцією ASR, NLU, LLM та Vision AI. Об'єктом дослідження є процес розпізнавання та опрацювання голосових і текстових команд користувача. Предметом дослідження є методи та технології автоматичного розпізнавання мови, опрацювання природної мови, великих мовних моделей та комп'ютерного “зору”. У дослідженні використано методи системного аналізу, експериментальні та статистичні методи. Результатом є функціонуючий прототип системи з гібридним NLU-рушієм, локальним опрацюванням даних та вебінтерфейсом на базі сучасних технологій.

Ключові слова: голосовий асистент, розпізнавання мови, ASR, NLU, штучний інтелект, LLM, Whisper, Ollama, комп'ютерний “зір”.

Вступ: У сучасному світі технології штучного інтелекту стрімко проникають у всі сфери людського життя. За статистикою, 85% українців користуються інструментами штучного інтелекту постійно або час від часу [8]. Проблема в тому, що всі асистенти призначені лише для смартфонів та жоден не має повної чи офіційної підтримки персонального комп'ютера (ПК). Водночас більшість існуючих рішень є закритими хмарними системами, що викликає занепокоєння щодо приватності даних та обмежує можливість кастомізації.

Постановка задачі: Існуючі системи голосового управління часто не забезпечують достатнього рівня конфіденційності, мають затримку через залежність від інтернету та погано адаптуються до потреб конкретного користувача. Виникає необхідність створення локального, приватного асистента на ПК з оптимізацією під українську мову, який би гарантував швидкий відгук та мав можливість мультимодальної “взаємодії” (“розумів” і голос, і зображення на екрані) [6].

Мета роботи: Розробити програмне забезпечення для розпізнавання голосових і текстових команд із використанням моделей штучного інтелекту, використання яких забезпечує високу точність розпізнавання, швидкість опрацювання та можливість самонавчання.

Основна частина: Система побудована на базі модульної монолітної архітектури. Для розпізнавання мови (ASR) інтегровано модель Whisper (base) від OpenAI [2]; завдяки оптимізаціям (квантизація int8, налаштування VAD) використання оперативної пам'яті було зменшено з 600 МБ до 150 МБ, а час розпізнавання команди складає ~1.2 сек. Модуль опрацювання природної мови (NLU) [7] використовує гібридний підхід: швидкі rule-based алгоритми для простих команд (опрацюють ~80% запитів) та велику мовну модель для складніших інтеракцій (20% запитів) [1].

Локальний штучний інтелект (AI Brain) працює на базі моделі LLaMA 3.2 [4], розгорнутої локально через платформу Ollama [3]. Backend системи розроблено мовою

Python із застосуванням фреймворку FastAPI (підтримує REST API та WebSocket) [5], а Frontend-дашборд у вигляді PWA реалізовано на React 18. Унікальною особливістю системи є підсистема самонавчання, що складається з *Analytics Engine*, *Habit Learner* та *Feedback System*; за допомогою цієї системи аналізується рутинна робота користувача (наприклад, щоденні послідовності запуску програм) та формує проактивні пропозиції. Крім того, додано Vision Module, який застосовує OCR (Tesseract) для аналізу вмісту екрану.

Висновки: Розроблений асистент є ефективним інструментом, що повністю розв'язує задачу локального опрацювання даних без передачі голосу в хмару. Впроваджений гібридний NLU-підхід у поєднанні з механізмом кешування відповідей LLM дозволив досягти часу опрацювання команди менш ніж 0.1 сек. Запропонований дизайн, орієнтований на модульність і використання систем самонавчання, робить застосунок гнучким, швидким і здатним адаптуватися до індивідуальних звичок і потреб користувача.

Список використаних джерел

1. Vaswani A. et al. Attention Is All You Need / Advances in Neural Information Processing Systems. – 2017. – [Фундаментальна архітектура Transformer для обробки природної мови]. [Електронний ресурс] – URL: <https://arxiv.org/pdf/1706.03762>.
2. Radford A. et al. Robust Speech Recognition via Large-Scale Weak Supervision / OpenAI. – 2022. – [Опис архітектури та навчання моделі Whisper]. [Електронний ресурс] – URL: <https://arxiv.org/pdf/2212.04356>.
3. Ollama: Get up and running with large language models locally. [Електронний ресурс]. – URL: <https://docs.ollama.com/>
4. Touvron H. et al. LLaMA: Open and Efficient Foundation Language Models / Meta AI. – 2023. – [Електронний ресурс] – URL: <https://arxiv.org/abs/2302.13971>
5. Офіційна документація FastAPI [Електронний ресурс]. – URL: <https://fastapi.tiangolo.com/>.
6. Франчук В. М. Захист інформаційних ресурсів: криптографічні та стеганографічні методи захисту даних / В. М. Франчук. – Київ, 2012
7. What is natural language understanding (NLU)? [Електронний ресурс]. – URL: <https://www.ibm.com/think/topics/natural-language-understanding>.
8. Gradus Research. 85% українців використовують ШІ, тільки 6% за нього платять // Спеціальне дослідження компанії Gradus Research. – 2026. – [Статистика використання штучного інтелекту в Україні]. [Електронний ресурс] – URL: <https://gradus.app/uk/open-reports/85-ukrainians-use-ai-only-6-pay-it/>.