

ОСОБЛИВОСТІ РОЗРОБКИ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ СТВОРЕННЯ ІНТЕРАКТИВНОГО ЕКРАННОГО ПОМІЧНИКА

Вінницький національний технічний університет

Анотація

У роботі розглянуто особливості розробки інформаційної технології інтерактивного екранного помічника з перспективою інтеграції ChatGPT. Подано переваги використання DirectX для рендерингу сцени (3D-моделі), особливості візуальної взаємодії помічника з користувачем у вигляді анімацій, інтеграцію штучного інтелекту на основі ChatGPT API, а також голосове введення та синтез мовлення (TTS), що сприяють підвищенню продуктивності роботи та комунікації з інтерактивним екранним помічником. Кінцевий результат поданий як десктопний застосунок для ОС Windows у вигляді безрамкового вікна з анімованою 3D-моделлю, яка реагує на дії користувача у вигляді анімацій з додатковими вікнами чату та налаштувань.

Ключові слова: екранний помічник, 3D-модель, DirectX, ChatGPT API, синтез мовлення, розпізнавання голосу.

Abstract

The paper considers the features of the creation and development of information technology for an interactive screen assistant with the prospect of ChatGPT integration. The advantages of using DirectX for rendering a scene (3D model), features of visual interaction between the assistant and the user in the form of animations, integration of artificial intelligence based on the ChatGPT API, as well as voice input and speech synthesis (TTS), which contribute to increasing productivity and communication with an interactive on-screen assistant, are presented. The final result is presented as a desktop application for Windows OS in the form of a frameless window with an animated 3D model that responds to user actions in the form of animations with additional chat and settings windows.

Keywords: on-screen assistant, 3D model, DirectX, ChatGPT API, speech synthesis, voice recognition.

Вступ

Нині інформаційні технології набули високого рівня розвитку, в результаті чого люди проводять багато часу за комп'ютером і як наслідок, можуть навіть доволі довго не спілкуватись з іншими людьми. Виконання рутинних задач або пошук вирішення простих проблем, здебільшого, або вимагає багато часу, або просто морально виснажує. Вирішенню таких проблем сприяє інтерактивний екранний помічник. Актуальність та доцільність створення інтерактивного екранного помічника можна аргументувати низкою фактів: поміч у вирішенні різного рівня складності задач; можливість створення запитів до ChatGPT безпосередньо з його допомогою, використовуючи текстовий чат чи голосове введення; оголошення відповідей за допомогою синтезу мовлення, а також він може стати як цікавим співрозмовником так і гарною прикрасою робочого столу.

Крім того, подібна інформаційна технологія доволі малорозвинена на сьогодні, що може свідчити про гарну перспективу для її розвитку в майбутньому, враховуючи стрімкий розвиток як продуктивності комп'ютерної техніки, так і рівня розвитку штучного інтелекту. В контексті розвитку штучного інтелекту, можна сказати, що незабаром люди зможуть створювати своїх віртуальних друзів та асистентів з його використанням, які здатні покращити швидкість і продуктивність роботи за комп'ютером з можливістю інтеграції в робочий, навчальний чи розважальний процес.

Результати дослідження

Наразі, кількість аналогів інтерактивного екранного помічника можна, буквально, перерахувати на пальцях однієї руки. Це з однієї точки зору дає доволі розпливчату картину щодо загального концепту інформаційної технології, а з іншої – широкий спектр для введення різнобічного функціоналу. В цьому дослідженні акцентовано увагу на загальних особливостях реалізації програмного забезпечення та інтеграції низки функціональних можливостей [1], які можна розділити на декілька модулів: рендеринг та анімування, інтеграція ChatGPT API, голосове введення та синтез мовлення.

Найважливішою частиною інформаційної технології виступатиме саме графічний модуль, який базується на архітектурі DirectX11 [2]. Використання «чистого» DirectX певним чином ускладнює процес розробки в порівнянні з вже готовими рушіями, проте, не створює обмежень для реалізації графіки, введення моделей та анімацій різних форматів. Графічний модуль забезпечуватиме рендеринг тривимірної моделі, її скінінг, текстуринг і шейдерні візуальні ефекти. Наступним питанням є саме виокремлення відрендереної моделі з рендертаргету та усунення рамок вікна програми для більшої естетичності та простоти сприйняття, це можна вирішити з вже існуючим для Windows вікон блендингом прозорості фону (alpha-blending) і вимкнувши рамки для вікна.

Черговим ключовим модулем виступає система анімування, яка контролюватиме програвання потрібної анімації на 3D-моделі інтерактивного помічника відповідно до дій користувача або певного триггеру [3]. Програвання скелетних анімацій, наприклад, анімації очікування, реакції на запит, відтворення мовлення чи подібні емоційні прояви забезпечуватимуть природність взаємодії користувача з помічником. Анімаційні послідовності розподілено на декілька видів, наприклад, так звані idle-анімації для постійного відтворення у разі бездіяльності та спеціальні, прив'язані до певних тригерів.

Одним із найбільш важливих етапів є саме інтеграція з програмним інтерфейсом ChatGPT API [4]. Механізм взаємодії передбачає формування текстового запиту користувачем у графічному чатовому вікні або його голосове озвучення, подальше надсилення до серверної частини, після чого, отримання відповіді та її виведення. Така архітектура дозволяє реалізувати природний діалог у форматі «питання-відповідь», в якому помічник виступає у ролі віртуального співрозмовника. При цьому, обробку запитів забезпечено з урахуванням попередніх питань, що формуватиме ефект послідовного логічного ведення діалогу.

Зокрема, вищезгаданий функціонал голосового введення є доволі актуальним для полегшення процесу взаємодії для користувача. Використання бібліотек розпізнавання мовлення [5], наприклад, Windows Speech та інших API, надає користувачеві можливість взаємодії з помічником без нудного введення запиту в чаті вручну. Після опрацювання голосового запиту результат відправляється до ChatGPT у вигляді текстового повідомлення-запиту. Виведення результату може бути подано у текстовому форматі безпосередньо в чаті або у вигляді озвучування отриманої відповіді з використанням системи синтезу мовлення, так званий Text To Speech (TTS), наприклад, з використанням TTS-сервісів, серед яких можна виокремити Microsoft Azure Speech або Google TTS.

Загалом, можна виділити основні етапи розробки інтерактивного екранного помічника, які включають такі модулі: графічний модуль (рендеринг та анімування), логічний модуль (ChatGPT API) та модуль введення-виведення (текстовий чат чи голосове введення та синтез мовлення). Для реалізації інформаційної технології та розширення її функціональних можливостей існує чимало доступних програмних бібліотек та систем, що суттєво спрощують процес її розробки.

Висновки

Згідно з проведеним дослідженням встановлено, що для створення інформаційної технології інтерактивного екранного помічника є доцільним впровадження функціональних можливостей таких як анімовані реакції помічника на дії користувача, інтеграція ChatGPT з використанням API від OpenAI, функціонал для голосового введення від користувача та синтез мовлення для отриманих від ChatGPT відповідей. Впровадження таких можливостей суттєво може спростити користувачеві процес роботи за комп'ютером, знизити рівень стресу чи, навіть, стати гарним співрозмовником або естетичною прикрасою робочого столу.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Faryal Fatima Syeda: Development of a VR/3D Hand Prosthesis Software Application: Avatar Customization and Hand Motion Control. URL: <https://odr.chalmers.se/server/api/core/bitstreams/10e08170-d09a-4511-9077-6f1f31311758/content> (дата звернення: 11.06.2025).
2. Understand the Direct3D 11 rendering pipeline. URL: <https://learn.microsoft.com/en-us/windows/win32/direct3dgetstarted/understand-the-directx-11-2-graphics-pipeline> (дата звернення: 11.06.2025).
3. Animating Characters with DirectX. URL: <https://gamedev.net/tutorials/programming/graphics/animating-characters-with-directx-r3632/> (дата звернення: 11.06.2025).
4. Як використовувати API ChatGPT. URL: <https://www.chatgptacademy.online/instrukcziyi-chatgpt/yak-vykorystovuvaty-api-chatgpt/> (дата звернення: 11.06.2025).
5. A Complete Guide to Speech Recognition. URL: <https://transkriptor.com/speech-recognition/> (дата звернення: 11.06.2025).

Ганевич Владислав Миколайович – студент групи 2КН-24м, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, email: vladislav.ganevich@gmail.com.

Крилик Людмила Вікторівна – к.т.н, доцент, доцент кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця.

Hanevych Vladyslav M. – Faculty of Intelligent Information Technology and Automation, Vinnytsia National Technical University, Vinnytsia, email: vladislav.ganevich@gmail.com

Krylik Lyudmila V. – PhD (Eng.), Associate Professor of Department for Computer Science, Vinnytsia National Technical University, Vinnytsia.