

ВИКОРИСТАННЯ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ ДЛЯ АНАЛІЗУ ВІДЕО

Кунуп Т.В., Антощук С.Г.

Національний університет «Одеська політехніка», м. Одеса

Сучасні технології штучного інтелекту (ШІ) і великі мовні моделі (LLMs, Large Language Models) активно впроваджуються в різні галузі промисловості, сприяючи автоматизації моніторингу, підвищенню продуктивності та сталого розвитку [1]. Найбільш традиційними завданнями, в яких використовуються LLMs є автоматизація створення статей, звітів, маркетингових документів, поліпшення перекладів. Для їх вирішення є рекомендації щодо формування промтів (prompt engineering), що передаються моделі у вигляді інструкцій або запитів, для налаштування її роботи з генеративним ШІ. Але впровадження LLMs, що поєднують можливості аналізу текстової та візуальної інформації, вимагає більш складного підходу, що включає додаткову обробку. Метою даного дослідження є оцінка потенціалу та ефективності застосування великих мовних моделей у поєднанні з технологіями технічного зору.

Розроблено інформаційну систему на основі Chat GPT4 (GPT4-mini), роботу якої можна умовно розділити на основні етапи:

Завантаження відео. Користувач може завантажувати відео з локального комп'ютера або вставляти посилання на YouTube. Підтримується обробка відео у різних форматах, що забезпечує гнучкість у роботі з даними.

Вилучення кадрів. Здійснюється вибір кадрів з відео через заданий інтервал часу (від 1 до 20 секунд).

Аналіз кадрів. Проводиться попередня обробка (фільтрація, контрастування тощо) і аналіз кадрів за допомогою алгоритмів машинного навчання для виявлення об'єктів, що цікавлять, і аномалій на кадрі.

Взаємодія з користувачем. Дозволяє переглядати оброблені кадри, ставити питання, уточнювати деталі та аналізувати окремі кадри з відео, а також отримати автоматично генеровані звіти на основі аналізу відео.

Загалом розроблена інформаційна система успішно виконала обробку відео, витягла ключові кадри і сформувала текстовий опис, спрощуючи аналіз відеоданих та надаючи зручний інструмент для роботи з візуальною інформацією. Слід відмітити, що хоча на даний момент використання LLMs може бути досить дорогим через високі вимоги до обчислювальних ресурсів та ліцензування, але очікується, що згодом їхня вартість буде знижуватися. Розвиток технологій та збільшення кількості LLMs з відкритим вихідним кодом розширюють доступність цих інструментів для широкої аудиторії, що дозволить компаніям та організаціям використати передові технології ШІ без значних фінансових вкладень, знижуючи бар'єри для інновацій.

Література

1. <https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/> Дата звернення 26.02.2025