

УДК 004.9

DOI: 10.31498/2225-6733.53.1.2026.359778

**СИСТЕМА РОЗПІЗНАВАННЯ ЗМІСТУ МЕДИЧНИХ АУДІОЗАПИСІВ
ДЛЯ ЗАПОВНЕННЯ МЕДИЧНОЇ ДОКУМЕНТАЦІЇ****Затуловський
Г.А.***магістр, НТУУ «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ,
e-mail: h.zatulovskyi.zk41mp.fbmi25@ill.kpi.ua;***Піднебесна Г.А.***канд. техн. наук, ст. викладач, НТУУ «Київський політехнічний інститут імені Ігоря Сікорського»; Інститут інформаційних технологій та систем НАН України, м. Київ,
ORCID: <https://orcid.org/0000-0002-5735-9861>, e-mail: halyna.pidnebesna@gmail.com*

У статті розглядається проблема ведення медичної документації лікарями, що є одним із найбільш трудомістких аспектів роботи лікаря в сучасній системі охорони здоров'я. Лікарі витрачають багато робочого часу на заповнення електронних медичних карток та іншої документації, що уповільнює та знижує якість медичної допомоги. Система автоматичної транскрипції та заповнення медичних форм усуває ці бар'єри, поєднуючи розпізнавання мовлення (ASR), вилучення сутностей (NER), генеративний аналіз (LLM) та процедурну валідацію. Метою роботи є підвищення швидкості та зручності ведення медичної документації за допомогою інтелектуальної системи автоматичної обробки аудіозаписів та напівавтоматичного заповнення стандартних форм медичних звітів. Об'єктом дослідження є процес ведення медичної документації під час консультації лікар-пацієнт. Предметом дослідження є методи автоматичної транскрипції медичних аудіозаписів та інтелектуального заповнення стандартизованих медичних форм на основі обробки природної мови. Практичне значення роботи полягає в можливості застосування отриманих результатів для підвищення ефективності медичної діагностики та обробки клінічних даних. Запропоновані методи можуть бути використані в медичних інформаційних системах для автоматизації аналізу даних, підтримки прийняття рішень та підвищення точності діагностичних процедур.

Ключові слова: транскрибація медичних аудіо; штучний інтелект; розпізнавання мовлення; обробка природної мови; медична документація; браузерне розширення.

Постановка проблеми

Адміністрування медичної документації є однією з найбільш гострих та часозатратних проблем у щоденній практиці лікаря. Численні дослідження вказують на тривожну тенденцію: медичні фахівці змушені приділяти значну частку свого робочого часу не пацієнту, а саме заповненню електронних карток та формуванню звітності. Такий дисбаланс, де адміністративні завдання починають переважати над клінічною взаємодією, є вкрай негативним. Ситуація додатково ускладнюється через невинне зростання загального навантаження на медичну систему та одночасне посилення регуляторних вимог до якості, повноти та деталізації медичних записів.

Вирішенням цієї проблеми є впровадження інтелектуальних систем, здатних автоматизувати перетворення усного мовлення лікаря та пацієнта в структуровані медичні записи. Поєднання технологій автоматичного розпізнавання мовлення (ASR) та генеративних мовних моделей (LLM) дозволяє створити інструмент, що забезпечує транскрибацію, витяг ключових клінічних сутностей та заповнення стандартизованих форм у реальному часі. Така система є необхідним кроком для цифрової трансформації медичної галузі.

Аналіз останніх досліджень та публікацій

Технології автоматичного розпізнавання мовлення. Автоматичне розпізнавання мовлення (Automatic Speech Recognition, ASR) є ключовою технологією для систем автоматизації медичної

документації. За останні роки відбувся значний прогрес у цій галузі завдяки розвитку глибокого навчання та трансформерних архітектур нейронних мереж [1]. Було оглянуто та проведено детальний аналіз їхніх переваг та недоліків декілька технологій, таких як OpenAI Whisper [2, 3], Google Cloud Speech-to-Text [4], Vosk [5], Microsoft Azure Speech Services [6]. Для системи було обрано Whisper Large v3 Turbo, оскільки вона забезпечує високу точність водночас для української мови та англійської, працює офлайн, що критично для конфіденційності медичних даних, та має оптимальний баланс між швидкістю й точністю. Крім того, модель є open-source, не потребує додаткового навчання й активно підтримується спільнотою.

Аналіз методів обробки природної мови. Після отримання текстового транскрипту з системи автоматичного розпізнавання мовлення постає задача структурування інформації та виділення ключових медичних сутностей. Ця задача є критичною для автоматичного заповнення медичних форм, оскільки неструктурований текст діалогу між лікарем та пацієнтом має бути перетворений у конкретні поля документа (ПІБ пацієнта, вік, діагноз, скарги, анамнез, призначення тощо).

Для вирішення цієї задачі застосовуються різні підходи, які можна умовно розділити на три категорії: класичні NER-моделі [7], генеративні LLM-моделі [8] та rule-based методи [9]. У розробленій нами системі для визначення медичних сутностей та структурування інформації з транскриптів використовується Google Gemini API (модель Gemini 2.0 Flash). Вибір цієї моделі

обґрунтований кількома факторами: Gemini демонструє відмінну підтримку української мови та здатність розуміти медичний контекст навіть у розмовному мовленні. Gemini підтримує генерацію структурованого JSON-виводу, що спрощує інтеграцію з медичними формами. Крім того, модель здатна до few-shot learning, тобто може навчитися витягувати необхідні поля на основі кількох прикладів у промпті, без потреби у великих анотованих датасетах.

Існуючі системи автоматизації медичної документації. На сучасному ринку представлена значна кількість рішень для автоматизації медичної документації, які відрізняються функціональністю, технологічним стеком, ціною політикою та географічним охопленням. Проаналізовано найбільш значущі системи, які представляють різні підходи до вирішення проблеми автоматизації медичного документообігу: Suki AI [10], Nuance Dragon Medical [11], DeepScribe.

Аналіз існуючих систем виявив наступні ключові обмеження, які обґрунтовують необхідність розробки власного рішення. По-перше, жодна з провідних міжнародних систем не підтримує українську мову на достатньому рівні для медичних застосувань, що робить їх непридатними для використання в українських медичних закладах без значних доопрацювань. По-друге, всі хмарні рішення передають медичні дані на сервери, розташовані за кордоном (переважно у США), що створює юридичні ризики та суперечить принципам збереження медичної таємниці.

Ці обмеження визначають основні вимоги до розробленої системи. Система має підтримувати українську мову на високому рівні, забезпечувати повністю локальну обробку даних для гарантування конфіденційності, використовувати open-source компоненти для мінімізації вартості, інтегруватися з українськими стандартами медичної документації [14] та застосовувати сучасні AI-технології для ambient documentation.

Мета статті

Метою роботи є підвищення ефективності документообігу в медичних закладах шляхом розробки та впровадження інформаційної системи для автоматизованої обробки аудіоданих консультацій та інтелектуального заповнення медичної звітності.

Об'єктом дослідження є процес ведення медичної документації під час консультацій лікаря з пацієнтом. Предмет дослідження це методи автоматичної транскрибації медичних аудіозаписів та інтелектуального заповнення стандартизованих медичних форм на основі обробки природної мови.

Виклад основного матеріалу

Архітектурні рішення системи. Розроблена інформаційна система для автоматичного заповнення медичних форм реалізована з використанням модульної архітектури, що забезпечує розділення функціональних обов'язків між компонентами та полегшує подальше

масштабування й підтримку системи. Вибір модульного підходу обумовлений необхідністю забезпечення гнучкості системи, можливості незалежного оновлення окремих компонентів без впливу на функціонування всієї системи в цілому, а також спрощення процесів тестування та налагодження окремих модулів.

Клієнтська частина системи реалізована у вигляді браузерного розширення для Google Chrome, що працює на платформі Google Meet. Розширення використовує Manifest V3 – найновішу специфікацію для розширень Chrome, яка забезпечує покращену безпеку та продуктивність. Вибір саме Manifest V3 був зумовлений тим, що це найсучасніший стандарт, який надає розширені можливості для роботи з медіапотоків, підвищені гарантії безпеки користувацьких даних та оптимізоване споживання системних ресурсів.

Використання ring-pong підходу для безперервного захоплення аудіопотоку з мінімальними втратами даних є особливістю розробленої системи. Це дозволяє уникнути втрати аудіоданих між послідовними фрагментами запису.

Модульна архітектура з чітким розділенням відповідальності між компонентами підвищує надійність і масштабованість системи: збій одного модуля не впливає на роботу всієї системи в цілому. Це забезпечує високий рівень відмовостійкості.

Серверна частина системи реалізована на базі FastAPI фреймворку, який забезпечує обробку аудіофайлів, їх транскрибацію та збереження результатів у базі даних. Архітектура серверної частини побудована за принципом шаблону проектування Facade (Фасад), що дозволяє приховати складність взаємодії між підсистемами за уніфікованим інтерфейсом. Це рішення суттєво спрощує інтеграцію серверу з клієнтською частиною системи та полегшує налагодження, тестування й подальший розвиток функціоналу. Використання патерну Facade дозволяє API-контролерам викликати високорівневі методи без необхідності розуміти внутрішню логіку роботи кожної підсистеми, що знижує зв'язаність між компонентами та підвищує загальну гнучкість архітектури.

Однією з ключових функцій серверної частини є асинхронна обробка аудіочанків, що надходять від клієнтської системи (розширення Google Meet). Асинхронність є критично важливою для забезпечення можливості обробки множини одночасних сесій запису від різних користувачів без блокування основного потоку виконання та з мінімальними затримками. Кожен аудіофрагмент обробляється незалежно від інших, що дозволяє системі працювати у режимі реального часу, обробляючи кілька потоків паралельно без взаємного впливу. Для цього використовується асинхронна функція (async def), яка гарантує ефективне використання ресурсів сервера навіть при великій кількості одночасних запитів, оскільки під час очікування завершення операцій введення-виведення процесор може перемикатися на обробку інших запитів.

Транскрибація виконується локально за допомогою моделі Whisper Large V3 Turbo (transcriber.py), що гарантує конфіденційність медичних даних і незалежність від зовнішніх сервісів. Модель завантажується з Hugging Face Hub і автоматично налаштовується під доступне обладнання. Такий підхід забезпечує баланс між швидкістю, точністю та безпекою даних.

Інтелектуальний аналіз з використанням LLM.

Для автоматичного заповнення медичних форм на основі транскрибованого тексту використовується Google Gemini API через бібліотеку google-gemai. Gemini 2.0 Flash обрано як оптимальний вибір завдяки його високій швидкості обробки запитів (до 2 секунд на типову медичну форму), розширеному контекстному вікну (1 мільйон токенів), що дозволяє обробляти навіть дуже довгі транскрипти консультацій, та відмінній якості розуміння медичної термінології українською мовою. Gemini 2.0 Flash демонструє гарні результати в задачах структурованого витягу інформації та генерації JSON, що критично важливо для точного заповнення форм.

Аналізатор форм побудований за принципом шаблонного промптінгу з використанням Jinja2: для кожного типу медичної форми створено окремий детальний шаблон промпту, що містить опис структури форми, приклади коректного заповнення, правила валідації та інструкції щодо обробки випадків, коли певна інформація відсутня в транскрипті. Використання низької температури (0.1) забезпечує детерміновану та передбачувану поведінку моделі, мінімізуючи галюцинації та забезпечуючи стабільність результатів між запусками.

Після отримання даних від Gemini API система проходить критично важливий етап валідації та нормалізації, щоб забезпечити коректність заповнення форм та відповідність медичним стандартам. Валідатор виконує комплексну перевірку отриманих даних за кількома критеріями: перевіряє наявність усіх обов'язкових полів, які визначені в конфігурації форми, контролює формат спеціальних полів, таких як телефонні номери (маска +380XXXXXXXXXX), дати (формат ДД.ММ.РРРР), реєстраційні номери облікових карток платників податків та інші ідентифікатори, а також перевіряє відповідність значень полів типу "select" допустимим опціям зі словника, щоб уникнути введення некоректних або нестандартизованих значень. У разі виявлення невідповідностей система автоматично виконує нормалізацію даних до стандартного формату: додає коди країн до телефонних номерів, приводить дати до єдиного формату, видаляє зайві пробіли та спеціальні символи з текстових полів.

Для забезпечення безперервного захоплення аудіопотоку під час медичних консультацій було розроблено спеціалізоване розширення для браузера Chrome, яке інтегрується з платформою Google Meet та надає можливість автоматичного запису діалогів. Архітектура розширення побудована за принципом чіткого розділення відповідальності між компонентами, де

кожен модуль виконує свою специфічну функцію: service worker (background.js) координує загальну роботу розширення та управляє його життєвим циклом, offscreen document здійснює безпосереднє захоплення та обробку аудіопотоків з використанням Web Audio API, а content script забезпечує візуальну взаємодію з інтерфейсом Google Meet та відображення транскрипції в реальному часі. Така архітектура повністю відповідає вимогам Manifest V3, останньої версії специфікації Chrome Extensions, і забезпечує стабільну роботу навіть при тривалих сеансах запису, що є критично важливим для медичних консультацій, які часто тривають від 30 хвилин до години і більше.

Інтерфейс розширення інтегрується безпосередньо в сторінку Google Meet у вигляді плаваючої панелі, розташованої в лівому нижньому куті екрана. Вона забезпечує зручний доступ до елементів управління без перекриття основної зони відеоконференції з відображенням учасників.

Для забезпечення зручної та ефективної роботи медичного персоналу з накопиченими транскриптами консультацій та автоматичного заповнення різноманітних медичних форм було розроблено інтуїтивно зрозумілий веб-додаток на базі сучасного фреймворку Streamlit, який поєднує простоту використання з потужним функціоналом обробки даних. Інтерфейс системи логічно структурований та складається з двох основних режимів роботи, між якими користувач може легко перемикатися: режим перегляду та редагування існуючих транскриптів з можливістю додавання голосових нотаток лікаря, а також режим автоматичного заповнення медичних форм на основі інтелектуального аналізу транскрибованих консультацій з використанням можливостей великих мовних моделей.

Головна сторінка інтерфейсу системи розроблена з акцентом на функціональність та зручність використання. Інтерфейс містить бічну панель навігації з інтуїтивно зрозумілими елементами вибору режиму роботи та потужними інструментами фільтрації записів для швидкого пошуку потрібних консультацій серед великої кількості збережених даних. Центральна частина екрану відображає основну таблицю з даними, завантаженими з бази даних MySQL, де кожен рядок представляє окрему медичну консультацію з основними метаданими: скороченим ідентифікатором запису, іменами пацієнта та лікаря, коротким фрагментом транскрипту для попереднього перегляду та датою проведення консультації. Ця таблиця є інтерактивною та дозволяє користувачу швидко переглядати доступні записи, сортувати їх за різними критеріями та вибрати конкретний запис для детального перегляду або редагування одним кліком миші. Реалізовано також можливість запису додаткової інформації лікарем (призначені ліки, діагноз, симптоми, тощо).

Після завершення процесу транскрибації, який зазвичай займає лише кілька секунд завдяки використанню оптимізованої turbo версії моделі Whisper, система автоматично відображає розпізнаний текст у

спеціальному текстовому полі під модулем запису аудіо. Це дозволяє лікарю одразу перевірити якість розпізнавання, переконатися що вся важлива інформація була коректно розпізнана, і при необхідності відредагувати текст перед збереженням.

Режим автоматичного аналізу транскриптів консультацій є найбільш інноваційною та технологічно складною функціональною можливістю всієї медичної інформаційної системи, що забезпечує інтелектуальну обробку неструктурованих текстів діалогів та автоматичне заповнення медичних форм різних типів з

мінімальним втручанням користувача. Підсистема автоматичного аналізу побудована на основі сучасної модульної архітектури з чітким розділенням відповідальності між компонентами та використанням принципів об'єктно-орієнтованого проектування. На загальній схемі взаємодії компонентів підсистеми, представлено інтерфейс вибору форми, модуль завантаження конфігурацій, клієнт Gemini API для аналізу тексту, валідатор отриманих даних та генератор вихідних документів у форматах DOCX та JSON (рис. 1).

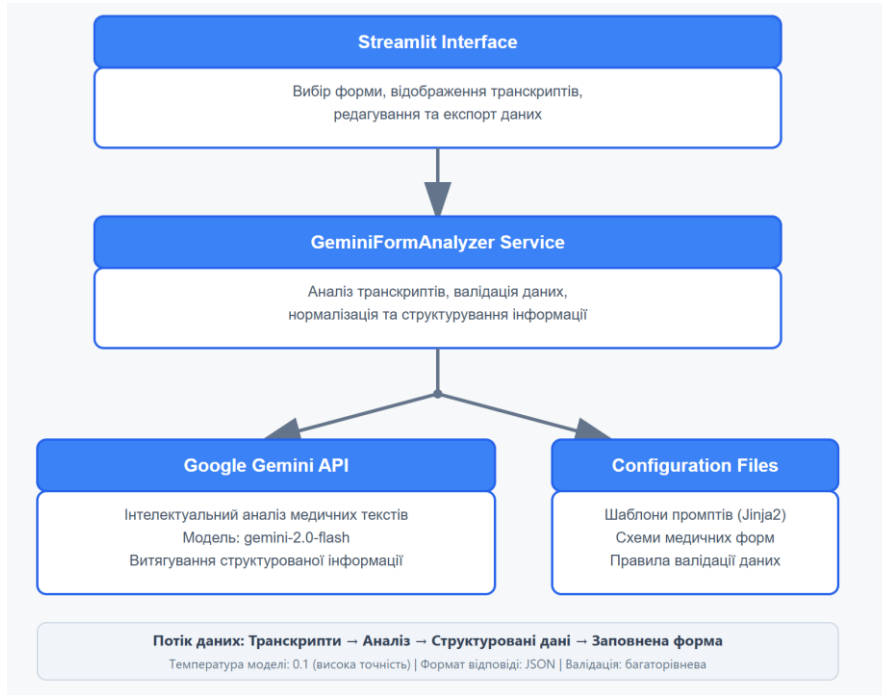


Рис. 1 – Архітектура підсистеми автоматичного аналізу

В поточній версії системи підтримуються різні типи форм медичної документації [13], включаючи форму 025/о «Медична карта амбулаторного хворого», форму 027/о «Виписка з медичної карти амбулаторного хворого» та інші стандартизовані форми МОЗ України, при цьому архітектура системи дозволяє легко додавати нові типи форм шляхом створення відповідних конфігураційних файлів та шаблонів промптів без необхідності модифікації основного програмного коду (рис. 2).

Формат JSON-файлу створюється одночасно з DOCX та містить абсолютно ідентичні дані, але представлені у машиночитаному форматі з чіткою ієрархічною структурою, де кожне поле форми має відповідний ключ та типізоване значення.

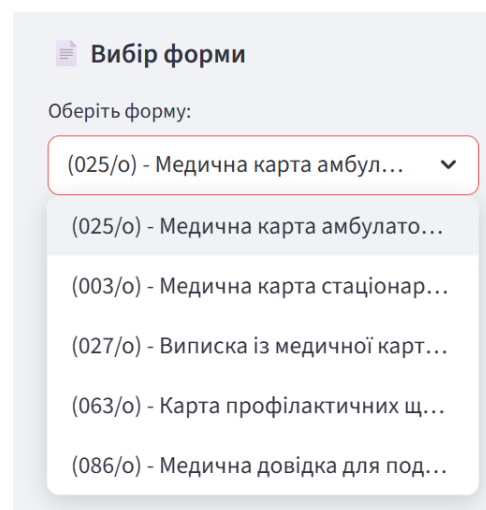


Рис. 2 – Приклад модуля для обрання медичної форми

Він особливо корисний для автоматизованої інтеграції з електронними медичними картками [16], системами електронного документообігу медичних закладів (рис. 3), регіональними та національними базами даних пацієнтів, а також для передачі даних у ЕСОЗ (Електронну систему охорони здоров'я України) через стандартизовані API без необхідності ручного перенесення інформації.



Рис. 3 – Приклад збереження у форматі DOCX

Висновки

Виконана робота направлена на вирішення гострої проблеми, пов'язаної з надмірним адміністративним навантаженням на медичний персонал в Україні, що призводить до значних часових витрат та ризиків помилок при ручному веденні документації. Головна мета полягала у створенні інтелектуальної програмної системи, здатної автоматично перетворювати аудіозаписи лікарських консультацій на структуровану звітність, адаптовану до національних стандартів.

Поставлена мета досягнута шляхом розробки модульної архітектури, яка поєднує клієнтську частину на базі браузерного розширення та високопродуктивний серверний обробник. Ключовою технічною особливістю є реалізація механізму конфіденційного та безперервного захоплення аудіопотоку завдяки унікальній «ping-pong» архітектурі. Забезпечено транскрибацію української медичної термінології через використання локально розгорнутої моделі Whisper Large v3 Turbo, що є фундаментальною перевагою над

хмарними іноземними аналогами та гарантує повну конфіденційність даних пацієнтів.

Важливим функціональним модулем системи є реалізація семантичного аналізу, який інтегрує генеративні можливості Gemini API з динамічними шаблонами Jinja2, дозволяє надійно витягувати з транскрипту ключові сутності (діагноз, скарги, анамнез) та автоматично формувати валідовані документи у форматі DOCX для офіційного використання, а також JSON-файли для інтеграції з існуючими МІС. Проведені обчислювальні експерименти повністю підтвердили як високу якість розпізнавання, так і успішне виконання всіх вимог, визначених технічним завданням.

Розроблена система має чіткі конкурентні переваги: вона є адаптованою до мовного та нормативного середовища України, забезпечує безпрецедентний рівень конфіденційності даних завдяки локалізації процесу транскрибації та прямо генерує звіти за національними стандартами.

Перелік використаних джерел

- [1] ISO 9001:2015. Quality management systems – Requirements. Geneva: International Organization for Standardization, 2015. 29 p.
- [2] LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature*. 2015. Vol. 521. Pp. 436–444. DOI: <https://doi.org/10.1038/nature14539>.
- [3] Robust Speech Recognition via Large-Scale Weak Supervision / A. Radford et al. arXiv preprint. arXiv:2212.04356. 2022. DOI: <https://doi.org/10.48550/arXiv.2212.04356>.
- [4] OpenAI Whisper Model Card. URL: <https://github.com/openai/whisper/blob/main/model-card.md> (дата звернення 07.11.2025).
- [5] Google Cloud Speech-to-Text Documentation. URL: <https://cloud.google.com/speech-to-text/docs> (дата звернення 07.11.2025).
- [6] Vosk Offline Speech Recognition API. URL: <https://alphacephei.com/vosk/> (дата звернення 10.11.2025).
- [7] Jensen P. B., Jensen L. J., Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*. 2012. Vol. 13. Pp. 395–405. DOI: <https://doi.org/10.1038/nrg3208>.
- [8] Yadav V., Bethard S. A Survey on Recent Advances in Named Entity Recognition from Deep Learning models. arXiv preprint. arXiv:1910.11470. 2019. DOI: <https://doi.org/10.48550/arXiv.1910.11470>.
- [9] LLaMA: Open and Efficient Foundation Language Models / H. Touvron et al. arXiv preprint. arXiv:2302.13971. 2023. DOI: <https://doi.org/10.48550/arXiv.2302.13971>.
- [10] ICD-11: International Classification of Diseases 11th Revision. World Health Organization, 2024. URL: <https://icd.who.int/en/> (дата звернення 04.11.2025)

- [11] Suki AI Platform Overview. URL: <https://www.suki.ai/> (дата звернення 15.11.2025)
- [12] DeepScribe: Ambient AI Scribe for Healthcare. 2024. URL: <https://www.deepscribe.ai/> (дата звернення 16.11.2025)
- [13] Електронна система охорони здоров'я в Україні. URL: <https://ehealth.gov.ua/> (дата звернення 16.11.2025)
- [14] Про затвердження форм первинної облікової документації та Інструкцій щодо їх заповнення, що використовуються у закладах охорони здоров'я незалежно від форми власності та підпорядкування : Наказ Міністерства охорони здоров'я України від 14.02.2012 № 110.

MEDICAL AUDIO RECORDING CONTENT RECORDING SYSTEM FOR COMPLETING MEDICAL DOCUMENTATION

- Zaturovskiy H.A.** M.Sc., National Technical University of Ukraine «Igor Sikorsky Kyiv polytechnic institute», Kyiv, e-mail: h.zatulovskiy.zk41mp.fbmi25@iit.kpi.ua;
- Pidnebesna H.A.** PhD (Engineering), senior lecturer, National Technical University of Ukraine «Igor Sikorsky Kyiv polytechnic institute», Kyiv; Institute of information technologies and systems of the National academy of sciences of Ukraine, Kyiv, ORCID: <https://orcid.org/0000-0002-5735-9861>, e-mail: halyna.pidnebesna@gmail.com

Maintaining medical documentation is one of the most labor-intensive aspects of a doctor's work in the modern healthcare system. Doctors spend a lot of working time filling out electronic medical records and other documentation, which slows down and reduces the quality of medical care. In most medical institutions in Ukraine, documentation is done manually, which creates a significant routine workload, increases the risk of errors, and distracts the doctor from interacting with the patient. A system for automatic transcription and medical form filling removes these barriers by combining speech recognition (ASR), entity extraction (NER), generative analysis (LLM), and procedural validation. The aim of the work is to improve the speed and convenience of medical documentation through an intelligent system for automatic processing of audio recordings and semi-automatic filling of standard medical report forms. Object of research is the process of medical documentation during doctor-patient consultations. Subject of research are methods for automatic transcription of medical audio recordings and intelligent filling of standardized medical forms based on natural language processing. To achieve the research goal, a detailed analysis of relevant modern technologies was conducted. The developed information system for automatic filling of medical forms is implemented using a modular architecture, which ensures the separation of functional responsibilities between components and facilitates further scaling and support of the system. The choice of a modular approach is due to the need to ensure system flexibility, the possibility of independent updating of individual components without affecting the functioning of the entire system as a whole, as well as simplifying the testing and debugging processes of individual modules. Modular architecture with a clear separation of responsibilities between components increases the reliability and scalability of the system: the failure of one module does not affect the operation of the entire system as a whole. This provides a high level of fault tolerance. The client part of the system is implemented as a browser extension for Google Chrome, running on the Google Meet platform. The extension uses Manifest V3, the latest specification for Chrome extensions, which provides improved security and performance. The use of a ping-pong approach for continuous audio stream capture with minimal data loss is a feature of the developed system. This avoids the loss of audio data between consecutive recording fragments. The goal was achieved by developing a modular architecture that combines a client part based on a browser extension and a high-performance server processor. The key technical feature is the implementation of a mechanism for confidential and continuous audio stream capture thanks to a unique «ping-pong» architecture. Transcription of Ukrainian medical terminology is provided through the use of a locally deployed Whisper Large v3 Turbo model, which is a fundamental advantage over cloud-based foreign analogues and guarantees complete confidentiality of patient data. An important functional module of the system is the implementation of semantic analysis, which integrates the generative capabilities of the Gemini API with dynamic Jinja2 templates, allows you to reliably extract key entities (diagnosis, complaints, anamnesis) from the transcript and automatically generate validated documents in DOCX format for official use, as well as JSON files for integration with existing MIS. The conducted computational experiments fully confirmed both the high quality of recognition and the successful implementation of all requirements specified in the technical specifications. The developed system has clear competitive advantages: it is adapted to the linguistic and regulatory environment of Ukraine, provides an unprecedented level of data confidentiality due to the localization of the transcription process, and directly generates reports according to national standards. The practical significance of the work lies in the possibility of applying the obtained results to increase the efficiency of medical diagnostics and clinical data processing. The proposed methods can be used in medical

information systems for automating data analysis, supporting decision-making, and improving the accuracy of diagnostic procedures. The developed system has clear competitive advantages: it is adapted to the linguistic and regulatory environment of Ukraine, provides an unprecedented level of data confidentiality due to the localization of the transcription process, and directly generates reports according to national standards.

Keywords: medical audio transcription; artificial intelligence; speech recognition; natural language processing; medical documentation; browser extension.

References

- [1] *Quality Management Systems – Requirements*, ISO 9001:2015, International Organization for Standardization, Geneva, Switzerland, 2015.
- [2] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015. doi: **10.1038/nature14539**.
- [3] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, “Robust Speech Recognition via Large-Scale Weak Supervision,” 2022, *arXiv:2212.04356*. doi: **10.48550/arXiv.2212.04356**.
- [4] OpenAI Whisper Model Card. [Online]. Available: <https://github.com/openai/whisper/blob/main/model-card.md>. Accessed on: November 07, 2025.
- [5] Google Cloud Speech-to-Text Documentation. [Online]. Available: <https://cloud.google.com/speech-to-text/docs>. Accessed on: November 07, 2025.
- [6] Vosk Offline Speech Recognition API. [Online]. Available: <https://alphacephei.com/vosk/>. Accessed on: November 10, 2025.
- [7] P. B. Jensen, L. J. Jensen, and S. Brunak, “Mining electronic health records: towards better research applications and clinical care,” *Nature Reviews Genetics*, vol. 13, pp. 395–405, 2012. doi: **10.1038/nrg3208**.
- [8] V. Yadav, and S. Bethard, “A Survey on Recent Advances in Named Entity Recognition from Deep Learning models,” 2019, *arXiv:1910.11470*. doi: **10.48550/arXiv.1910.11470**.
- [9] H. Touvron et al., “LLaMA: Open and Efficient Foundation Language Models,” 2023, *arXiv:2302.13971*. doi: **10.48550/arXiv.2302.13971**.
- [10] ICD-11: International Classification of Diseases 11th Revision. World Health Organization, 2024. [Online]. Available: <https://icd.who.int/en/>. Accessed on: November 04, 2025.
- [11] Suki AI Platform Overview. [Online]. Available: <https://www.suki.ai/>. Accessed on: November 15, 2025.
- [12] DeepScribe: Ambient AI Scribe for Healthcare. 2024. [Online]. Available: <https://www.deepscribe.ai/>. Accessed on: November 16, 2025.
- [13] Електронна система охорони здоров'я в Україні. [Online]. Available: <https://ehealth.gov.ua/>. Accessed on: November 16, 2025.
- [14] *Pro zatverdzhennia form pervynnoi oblikovoi dokumentatsii ta Instruksii shchodo yikh zapovnennia, sheho vykorystovuiutsia u zakladakh okhorony zdorovia nezalezhno vid formy vlasnosti ta pidporiadkuvannia* [On Approval of Forms of Primary Accounting Documentation and Instructions for Their Completion Used in Health Care Institutions Regardless of Ownership and Subordination], Order of the Ministry of Health of Ukraine no. 110, Feb. 14, 2012. (Ukr.)

Стаття надійшла 28.11.2025

Стаття прийнята 23.12.2025

Стаття опублікована 26.03.2026

Цитуйте цю статтю як: Затуловський Г. А., Піднебесна Г. А. Система розпізнавання змісту медичних аудіо-записів для заповнення медичної документації. *Вісник Приазовського державного технічного університету. Серія: Технічні науки*. 2026. Вип. 53, том 1. С. 69–75. DOI: <https://doi.org/10.31498/2225-6733.53.1.2026.359778>.