DIGITAL MEDICINE
ЦИФРОВА МЕДИЦИНА

https://doi.org/10.15407/intechsys.2026.01.043
UDC 004:621.391

H.A. ZATULOVSKYI, Master's Student,
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",
37, Beresteysky ave., Kyiv, 03056, Ukraine
https://orcid.org/0009-0005-1579-5461
joranosorog@gmail.com

H.A. PIDNEBESNA, PhD (Engineering), Senior Lecturer,
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",
37, Beresteysky ave., Kyiv, 03056, Ukraine
Senior Researcher,
Institute of Information Technologies and Systems NAS Ukraine,
40, Hlushkova Akad. ave., Kyiv, 03187, Ukraine
https://orcid.org/0000-0002-5735-9861
pidnebesna@ukr.net

SYSTEM FOR MEDICAL DOCUMENTATION FILLING BASED ON AUTOMATIC RECOGNITION OF AUDIO RECORDINGS

The investigation aims to address the pressing issue of excessive administrative burden on medical personnel in Ukraine, which leads to significant time expenditures and the risk of errors in manual documentation. As a result, an intelligent software system was created that can automatically convert audio recordings of medical consultations into structured reporting adapted to national standards. The developed system is adapted to the linguistic and regulatory environment of Ukraine, ensures an unprecedented level of data confidentiality by localizing the transcription process, and directly generates reports in accordance with national standards.

***Keywords**: intelligent software system, automatic recognition of audio recordings, medical records administration, automatic speech recognition, generative language models, structured reporting.*

Introduction

Medical record management is one of the most pressing and time-consuming challenges in physician practice. Numerous studies point to a disturbing trend: medical professionals are forced to devote a significant

Cite: Zatulovskyi H.A., Pidnebesna H.A. System For Medical Documentation Filling Based on Automatic Recognition of Audio Recordings. *Information Technologies and Systems*. 1 (7). 2026. 43—53. https://doi.org/10.15407/intechsys.2026.01.043

portion of their time to completing electronic records and generating reports, rather than to patient care. This imbalance, where administrative tasks begin to take precedence over clinical interactions, is extremely detrimental. The situation is further complicated by the ever-increasing overall burden on the medical system and the simultaneous tightening of regulatory requirements for the quality, completeness and detail of medical records [1—3].

The solution to this problem is the implementation of intelligent systems capable of automating the conversion of physician and patient speech into structured medical records. The combination of automatic speech recognition (ASR) and generative language models (LLM) technologies enables the creation of a tool that enables transcription, extraction of key clinical entities and the completion of standardized forms in real time. This system is a necessary step in the digital transformation of the medical industry.

The purpose of the paper is to improve the efficiency of document flow in medical institutions by developing and implementing an information system for automated processing of audio data from consultations and intelligent completion of medical reports.

Survey of Modern Technologies for Solving the Problem

Automatic Speech Recognition Technologies. Automatic Speech Recognition is a key technology for medical record automation systems. Significant progress has been made in this field in recent years thanks to the development of deep learning and transformative neural network architectures [4]. Several technologies, including OpenAI Whisper [5, 6], Google Cloud Speech-to-Text [7], Vosk [8], and Microsoft Azure Speech Services [9], were reviewed and their advantages and disadvantages were analyzed in detail. Whisper Large v3 Turbo was selected for the system because it provides high accuracy for both Ukrainian and English, operates offline (critical for medical data privacy), and offers an optimal balance between speed and accuracy. Furthermore, the model is open-source, requires no additional training, and is actively supported by the community.

Analysis of Natural Language Processing Methods. After receiving a text transcript from an automatic speech recognition system, the task of structuring the information and identifying key medical entities arises. This task is critical for automatically filling medical forms, as the unstructured text of the doctor-patient dialogue must be converted into specific document fields (patient name, age, diagnosis, complaints, medical history, prescription, etc.) [10].

Various approaches are used to solve this problem, which can be roughly divided into three categories: traditional NER models [11], generative LLM models [12], and rule-based methods [13]. Our system for identifying medical entities and structuring transcript information uses the

Google Gemini API (Gemini 2.0 Flash model). The choice of this model was based on several factors: Gemini demonstrates excellent support for the Ukrainian language and the ability to understand medical context even in colloquial speech. Gemini supports the generation of structured JSON output, which simplifies integration with medical forms. In addition, the model is capable of few-shot learning, meaning it can learn to extract the required fields based on a few examples in the prompt, without the need for large annotated datasets.

Existing Medical Record Automation Systems. The current market offers a significant number of medical record automation solutions, varying in functionality, technology stack, pricing and geographic coverage. The most significant systems, representing different approaches to solving the problem of automating medical record flow, were analyzed: Suki AI [14], Nuance Dragon Medical [15] and DeepScribe [16].

An analysis of existing systems revealed the following key limitations, justifying the need to develop a custom solution. First, none of the leading international systems support the Ukrainian language at a level sufficient for medical applications, making them unsuitable for use in Ukrainian medical institutions without significant modifications. Second, all cloud-based solutions transmit medical data to servers located abroad (primarily in the United States), which creates legal risks and violates confidentiality principles.

These limitations define the basic requirements for the developed system. The system must support the Ukrainian language at a high level, ensure fully local data processing to ensure confidentiality, use open-source components to minimize costs, integrate with Ukrainian medical documentation standards [17] and use modern AI technologies for ambient documentation.

System Architecture

We developed an information system for automated medical form filling. It utilizes a modular architecture that enables seamless expansion of functionality between components and facilitates scalability and customization (Fig. 1). The modular approach eliminates the need for system maintenance, allows for independent component updates without requiring the inclusion of all system functions, accelerates testing and module loading.

The client-side portion of the system is designed for viewing browser extensions for Google Chrome and runs on the Google Meet platform. Manifest V3 is a new specification for Chrome extensions that ensures uninterrupted performance. We chose Manifest V3, the highest standard, which provides advanced capabilities for working with media streams, guarantees data-free access and optimizes system resource utilization.

An alternative ping-pong solution for seamless audio streaming with minimal overhead in a custom, advanced system. This will allow

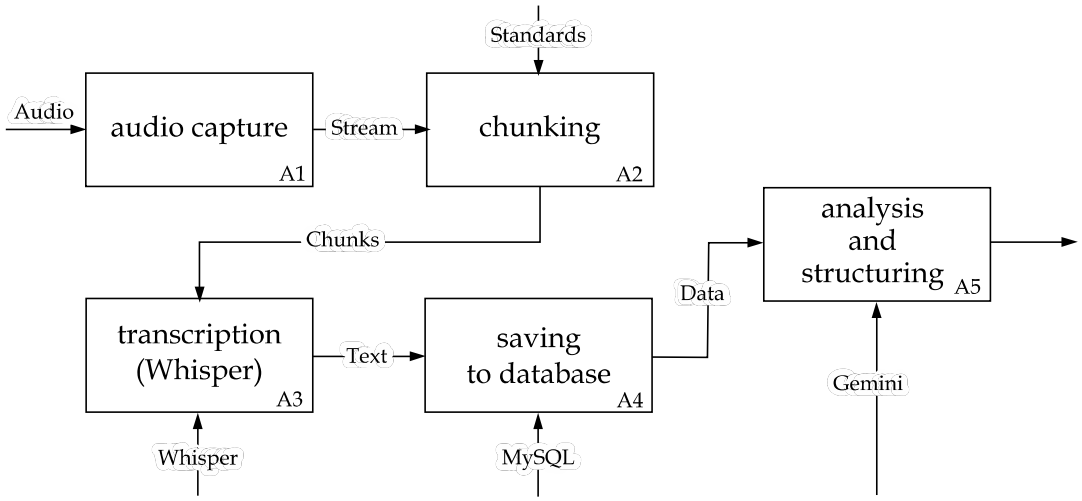

Fig 1. Main processing scheme

you to transition to audio recording with multiple sequential recording fragments.

A modular architecture with a clear separation of individual components improves the system's reliability and scalability: no single module is required to operate all systems as a whole. This ensures a high level of performance.

The server-side portion of the system is implemented using the FastAPI framework, which handles audio file processing, transcription and storing the results in a database. The server-side architecture is designed using the Facade design pattern, allowing for flexible interactions between subsystems with a unified interface. At this point, the servers are integrated with the client-side partial system, facilitating overlay, testing and further development of functions. The use of Facade API controllers enables the use of high-level methods without the need to develop internal logic for each subsystem's operations, which reduces the number of components and increases architectural complexity.

One of the key functions of the server-side is the asynchronous processing of audio chunks retrieved from the client system (a Google Meet extension). This asynchronous nature is crucial to avoid the need to process multiple one-hour recording sessions for various operations without blocking the main thread and with minimal latency. Sometimes an audio fragment is processed independently of others, allowing the system to operate in real time, processing multiple streams in parallel without interfering with each other. This function executes an asynchronous function (async def), ensuring efficient use of server resources during large numbers of simultaneous downloads. Oscillators can be recycled to handle other requests at the completion of the process creation/decomposition operation.

Transcription is performed locally for the additional Whisper Large V3 Turbo model (transcriber.py), ensuring the confidentiality of medical data and the immediate availability of additional services. The model is

loaded via the Hugging Face Hub and automatically configures itself upon access to the hardware. This approach provides a balance between speed, accuracy and unavailability.

Intelligent Analysis Using LLM

The Google Gemini API via the google-genai library is used to automatically fill medical forms based on transcribed text. Gemini 2.0 Flash was chosen as the optimal choice due to its high request processing speed (up to 2 seconds for a typical medical form), an expanded context window (1 million tokens), which allows processing even very long consultation transcripts, and excellent understanding of medical terminology in Ukrainian. Gemini 2.0 Flash demonstrates good results in structured information extraction and JSON generation, which is critical for accurate form filling.

The form parser is built on the principle of template prompting using Jinja2: for each type of medical form, a separate, detailed prompt template is created, containing a description of the form structure, examples of correct completion, validation rules and instructions for handling cases where certain information is missing from the transcript. The use of a low temperature (0.1) ensures deterministic and predictable model behavior, minimizing hallucinations and ensuring stability of results between runs. After receiving data from the Gemini API, the system undergoes a critical validation and normalization step to ensure form completion is accurate and complies with medical standards. The validator performs a comprehensive check of the received data against several criteria: it verifies the presence of all required fields defined in the form configuration, controls the format of special fields such as phone numbers (mask +380XXXXXXXXX), dates (DD.MM.PPRR format), registration card numbers and checks “select” fields against acceptable dictionary options to prevent the entry of incorrect or non-standardized values. If inconsistencies are detected, the system automatically normalizes the data to a standard format: it adds country codes to phone numbers, standardizes dates and removes extra spaces and special characters from text fields.

To ensure seamless audio stream capture during medical consultations, a specialized Chrome browser extension was developed that integrates with the Google Meet platform and provides the ability to automatically record dialogues. The extension’s architecture is built around a clear separation of responsibilities between components, with each module performing its own specific function: a service worker (background.js) coordinates the overall operation of the extension and manages its lifecycle, while an offscreen document captures and processes audio streams using the Web Audio API and displays transcripts in real time. This architecture fully complies with Manifest V3, the latest version of the Chrome Extensions specification, and ensures stable operation even during long

recording sessions, which is critical for medical consultations, which often last from 30 minutes to an hour or more.

The extension's interface integrates directly into the Google Meet page as a floating panel located in the lower left corner of the screen. It provides convenient access to controls without obscuring the main video conference area displaying participants.

To ensure convenient and efficient work for medical staff with accumulated consultation transcripts and the automatic filling of various medical forms, an intuitive web application was developed based on the modern Streamlit framework, combining ease of use with powerful data processing functionality. The system's interface is logically structured and consists of two main operating modes, which the user can easily switch between: viewing and editing existing transcripts with the ability to add doctor voice notes, and automatically filling out medical forms based on intelligent analysis of transcribed consultations using the capabilities of large language models.

The system's main interface page is designed with an emphasis on functionality and ease of use. The interface includes (Fig. 2) a side navigation bar with intuitive mode selection controls and powerful record filtering tools for quickly finding the desired consultations among the large amount of stored data. The central part of the screen displays the main data table loaded from a MySQL database, where each row represents an individual medical consultation with basic metadata: an abbreviated record identifier, patient and doctor names, a short transcript fragment for preview and the date of the consultation. This table is interactive and allows the user to quickly browse available records, sort them by various criteria and select a specific record for detailed viewing or editing with one click.

The ability for the physician to record additional information (medication prescriptions, diagnosis, symptoms, etc.) is also implemented.

After the transcription process is complete, which typically takes only a few seconds thanks to the optimized turbo version of the Whisper model, the system automatically displays the recognized text in a dedicated text field below the audio recording module. This allows the physician to immediately check the recognition quality, ensure that all important information was correctly recognize and edit the text, if necessary, before saving.

The automatic analysis mode for consultation transcripts is the most innovative and technologically sophisticated feature of the entire medical information system, providing intelligent processing of unstructured dialogue texts and automatic filling of various types of medical forms with minimal user intervention. The automatic analysis subsystem is built on a modern modular architecture with a clear separation of responsibilities between components and utilizing object-oriented design principles. The general diagram of the interaction of the subsystem components shows the form selection interface, the configuration loading module, the Gemini

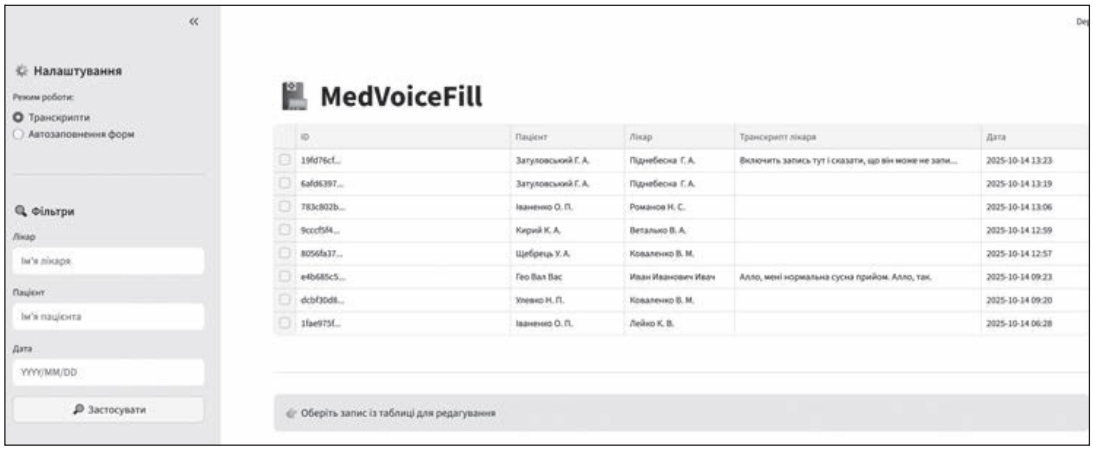

Fig 2. Main interface page with transcript mode

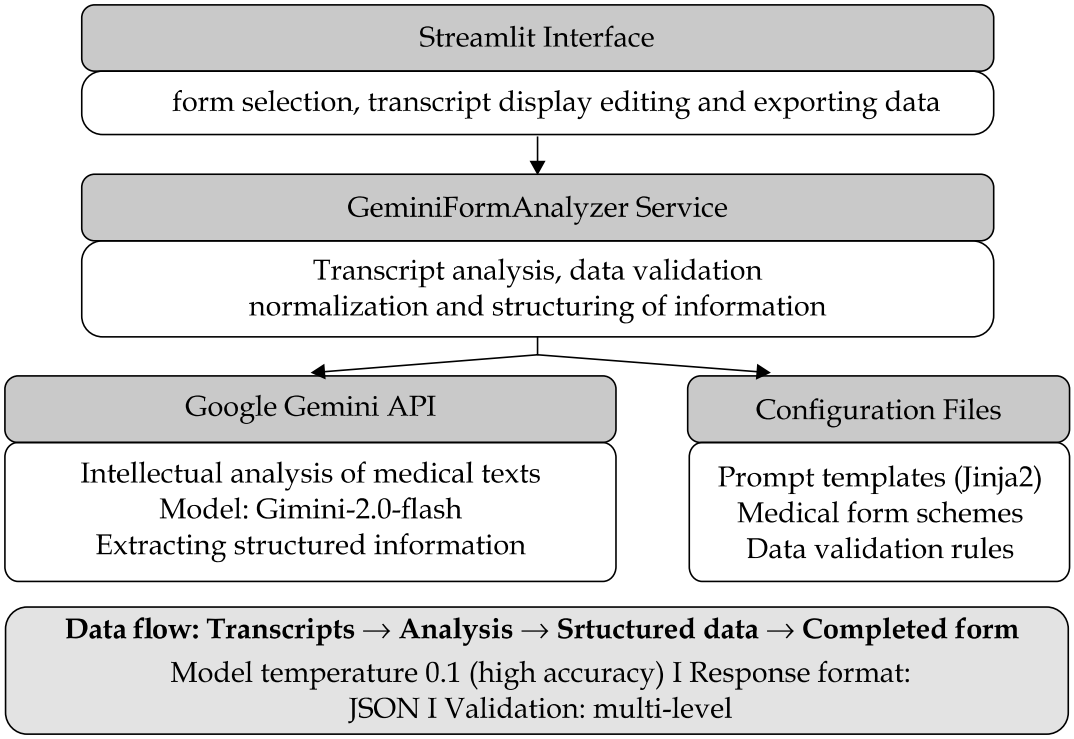

Fig 3. Architecture of the automatic analysis subsystem

API client for text analysis, the validator of the received data, and the generator of source documents in DOCX and JSON formats (Fig. 3).

An example of a module for selecting a medical form is shown on Figure 4.

An example of a generated document in DOCX format with fully populated medical form fields, including all mandatory and optional sections according to the standards of the Ministry of Health of Ukraine [18] (Fig. 5).

The JSON file format is created simultaneously with DOCX and contains identical data, but is presented in a machine-readable format with a clear hierarchical structure, where each form field has a corresponding key and a typed value. It is particularly useful for automated integration

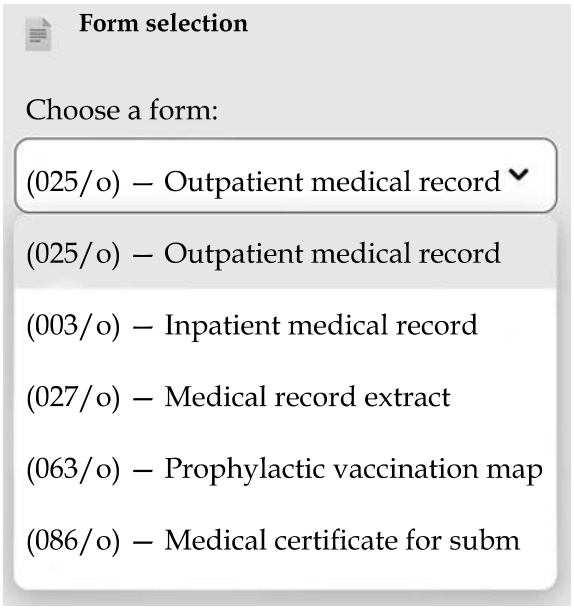

Fig. 4. Example of a module for selecting a medical form

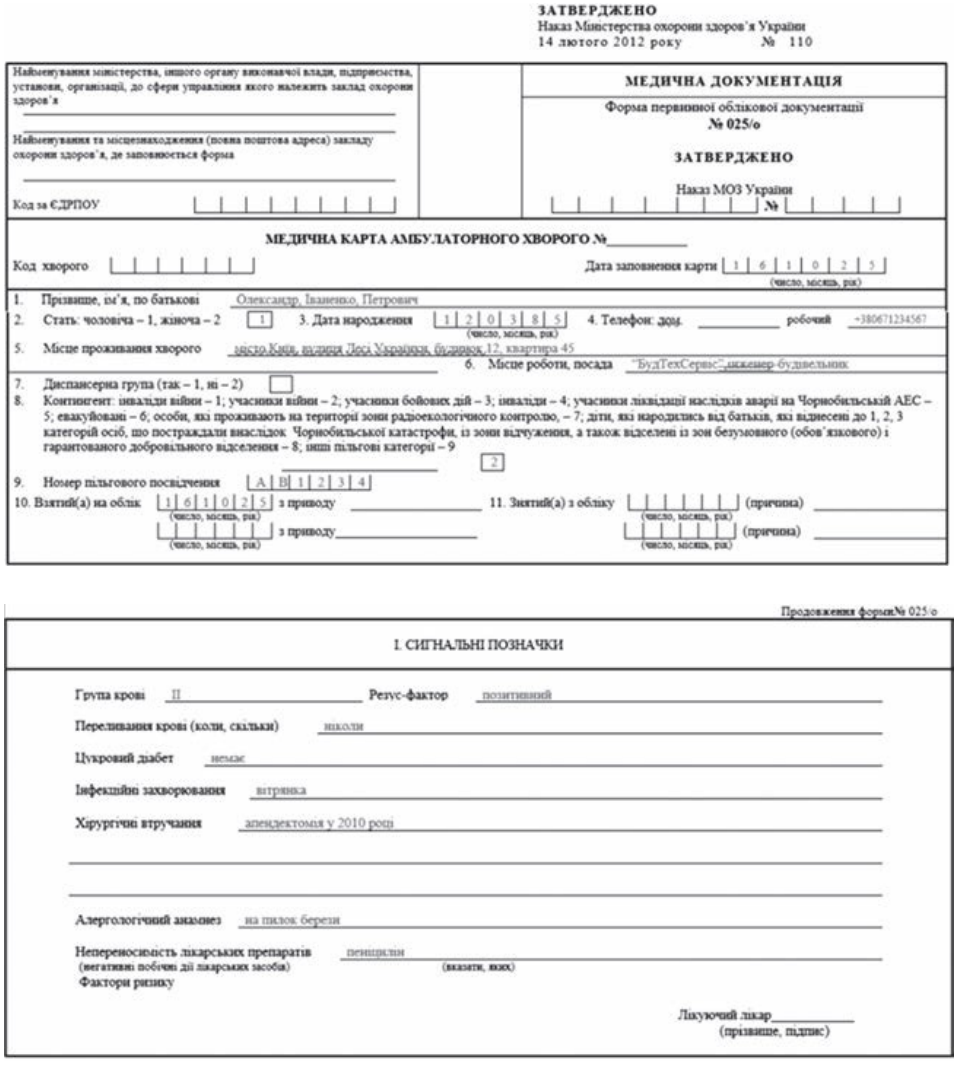

Fig. 5. Example of saving in DOCX format

with electronic medical records, electronic document management systems of medical institutions, regional and national patient databases, and for transferring data to the Unified Healthcare System (EHS) via standardized APIs without the need for manual data transfer.

Conclusions

This study aims to address the pressing issue of excessive administrative burden on medical personnel in Ukraine, which leads to significant time expenditures and the risk of errors in manual documentation. The main goal was to create an intelligent software system capable of automatically converting audio recordings of medical consultations into structured reporting adapted to national standards.

This goal was achieved through the development of a modular architecture combining a browser-based client and a high-performance server processor. A key technical feature is the implementation of a confidential and continuous audio stream capture mechanism thanks to an unique "ping-pong" architecture. High-precision transcription of Ukrainian medical terminology is ensured through the use of a locally deployed Whisper Large v3 Turbo model, a fundamental advantage over cloud-based foreign counterparts and guaranteeing 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. This enables reliable extraction of key entities (diagnosis, complaints, anamnesis) from the transcript and the automatic generation of validated documents in DOCX format for official use, as well as JSON. Computational experiments fully confirmed both the high quality of recognition and the successful fulfillment of all requirements defined in the technical specifications. The developed system has clear competitive advantages: it is adapted to the linguistic and regulatory environment of Ukraine, ensures an unprecedented level of data confidentiality by localizing the transcription process and directly generates reports in accordance with national standards.

DECLARATION

Declaration of Competing Interest. *No potential conflict of interest is reported by the authors.*

Funding. *The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.*

Use of AI. *The authors declare that Artificial intelligence tools were not used in writing the paper.*

Contribution of the authors. *Zatulovskyi H.A. is responsible for the software implementation, preparation of drawings, diagrams, and their description. Pidnebesna H.A. is responsible for the main idea of the article and its methodological part.*

REFERENCES

1. Arndt B.G., Beasley J.W., Watkinson M.D., et al. Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Mo-

tion Observations. *Annals of Family Medicine,* 2017, Vol. 15 (5), 419–426. https://doi.org/10.1370/afm.2121
2. Mamykina L., Vawdrey D.K., Hripcsak G. How Do Residents Spend Their Shift Time? A Time and Motion Study With a Particular Focus on the Use of Computers. *Academic Medicine,* 2016, Vol. 91 (6), 827–832. https://doi.org/10.1097/ACM.0000000000001148
3. Shanafelt T.D., West C.P., Sinsky C., et al. Changes in Burnout and Satisfaction With Work-Life Integration in Physicians and the General US Working Population Between 2011 and 2020. *Mayo Clinic Proceedings,* 2022, Vol. 97 (3), 491–506. https://doi.org/10.1016/j.mayocp.2021.11.021
4. LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature,* 2015, Vol. 521, 436–444. https://doi.org/10.1038/nature14539
5. Radford A., Kim J. W., Xu T., et al. Robust Speech Recognition via Large-Scale Weak Supervision. *arXiv,* 2022, Article 2212.04356. https://doi.org/10.48550/arXiv.2212.04356
6. OpenAI Whisper Model Card. URL: https://github.com/openai/whisper/blob/main/model-card.md [Accessed 07 Nov. 2025]
7. Google Cloud Speech-to-Text Documentation. URL: https://cloud.google.com/speech-to-text/docs [Accessed 07 Nov. 2025]
8. Vosk Offline Speech Recognition API. URL: https://alphacephei.com/vosk/ [Accessed 07 Nov. 2025]
9. Microsoft Azure Speech Services Documentation. URL: https://azure.microsoft.com/en-us/products/ai-foundry/tools/speech/ [Accessed 07 Nov. 2025]
10. Jensen P. B., Jensen L. J., Brunak S. Mining electronic health records… *Nature Reviews Genetics,* 2012, Vol. 13, 395–405. https://doi.org/10.1038/nrg3208
11. Yadav V., Bethard S. A Survey on Recent Advances in Named Entity Recognition… *The 27th International Conference on Computational Linguistics,* COLING, 2018, 2145–2158.
12. Touvron H., Lavril T., Izacard G., et al. LLaMA: Open and Efficient Foundation Language Models. *ArXiv*, 2023, Article 2302.13971.
13. International Statistical Classification of Diseases… ICD-10, WHO, 2019. URL: https://icd.who.int/ [Accessed 14 Nov. 2025]
14. Suki AI Platform Overview. URL: https://www.suki.ai/ [Accessed 15 Nov. 2025]
15. Nuance Dragon Medical One Documentation. https://dragon.nuance.com/en-us/user-documentation [Accessed 16 Nov. 2025]
16. DeepScribe: Ambient AI Scribe for Healthcare. URL: https://www.deepscribe.ai/ [Accessed 16 Nov. 2025]
17. Electronic health care system in Ukraine. URL: https://ehealth.gov.ua/ [Accessed 16 Nov. 2025]
18. Order Of The Ministry Of Health Of Ukraine 14 Feb. 2012 No. 110 On approval of forms of primary accounting documentation and Instructions for their completion, used in healthcare institutions regardless of the form of ownership and subordination. URL: https://zakon.rada.gov.ua/laws/show/z0661-12 [Accessed 16 Nov. 2025]

Received 13.11.2025
Accepted 25.03.2026.
Published 30.04.2026

Г.А. ЗАТУЛОВСЬКИЙ, магістрант,
Національний технічний університет України
«Київський політехнічний інститут ім. Ігоря Сікорського»,
просп. Берестейський, 37, м. Київ, 03056, Україна
https://orcid.org/0009-0005-1579-5461
joranosorog@gmail.com

Г.А. ПІДНЕБЕСНА, канд. техн. наук, старш. викладач,
Національний технічний університет України
«Київський політехнічний інститут ім. Ігоря Сікорського»,
просп. Берестейський, 37, м. Київ, 03056, Україна;
старш. наук. співроб.,
Інститут інформаційних технологій та систем НАН України,
просп. Акад. Глушкова, 40, м. Київ, 03187, Україна
https://orcid.org/0000-0002-5735-9861
pidnebesna@ukr.net

СИСТЕМА ЗАПОВНЕННЯ МЕДИЧНОЇ ДОКУМЕНТАЦІЇ НА ОСНОВІ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ АУДІОЗАПИСІВ

Вступ. 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 невпинне зростання загального навантаження на медичну систему та одночасне посилення регуляторних вимог до якості, повноти та деталізації медичних записів. Вирішенням цієї проблеми є запровадження інтелектуальних систем, здатних автоматизувати перетворення усного мовлення лікаря та пацієнта в структуровані медичні записи. Поєднання технологій автоматичного розпізнавання мовлення та генеративних мовних моделей дозволяє створити інструмент, який забезпечує транскрибацію, витяг ключових клінічних сутностей та заповнення стандартизованих форм у реальному часі. Така система є необхідним кроком для цифрової трансформації медичної галузі.

Метою статті є описати проблему і запропонувати рішення проблеми створення інтелектуальної програмної системи, здатної автоматично перетворювати аудіозаписи лікарських консультацій на структуровану звітність, адаптовану до національних стандартів, що підвищить ефективність документообігу в медичних закладах шляхом розробки та впровадження інформаційної системи для автоматизованої обробки аудіоданих консультацій та інтелектуального заповнення медичної звітності.

Методи. В роботі поєднано технології автоматичного розпізнавання мовлення (*ASR*) та генеративні мовні моделі (*LLM*), що 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.

Результати. Розроблено модульну архітектуру, 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 високоточну транскрибацію української медичної термінології через використання локально розгорнутої моделі *Whisper Large v*3 *Turbo*, що є фундаментальною перевагою над хмарними іноземними аналогами та гарантує повну конфіденційність даних пацієнтів.

Висновок. 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.

***Ключові слова**: інтелектуальна програмна система, автоматичне розпізнавання аудіозаписів, адміністрування медичної документації, автоматичне розпізнавання мовлення, генеративні мовні моделі, структурована звітність.*