

SEMMELWEIS EGYETEM
DOKTORI ISKOLA

Ph.D. értekezések

3449.

ANGYAL VIOLA ZSUZSANNA

Nép- és közegészségtudományok
című program

Programvezető: Dr. Ács Nándor, egyetemi tanár

Témavezető: Dr. Dinya Elek, professor emeritus

ScreenGPT – Developing and Evaluating a Custom Large Language Model to Support Cervical Cancer Prevention

PhD thesis

Angyal Viola

Semmelweis University Doctoral School

Health Sciences Division

Digital Health Sciences



Supervisor: Prof. Dr. Dinya Elek, PhD

Official reviewer: Dr. Tóth Zoltán, PhD Dr. Molnár László, PhD

Head of the Complex Examination Committee: Prof. Dr. Maurovich Horvat Pál, MD, PhD, DSc

Members of the Complex Examination Committee:

Dr. Kaposi András, PhD

Dr. Tamus Zoltán Ádám, PhD

Budapest
2026

Table of Contents

1. INTRODUCTION	5
1.1. Natural Language Processing	5
1.2. Generative Artificial Intelligence	6
1.3. Large Language Models in Healthcare	7
1.3. Custom Generative Pre-trained Transformers	8
1.4. Prompt engineering	10
1.5. Streamlit framework	11
1.6. Cervical cancer screening	12
2. OBJECTIVES	14
3. METHODS	15
3.1. ScreenGPT system architecture	15
3.1.1. Backend	15
3.1.2. Frontend	17
3.1.3. Prompts	17
3.2. User experience measurement	19
3.3. Evaluation of the system's responses by two independent medical doctors	20
3.4. Statistical analysis	20
4.1. ScreenGPT architecture	22
4.2. ScreenGPT web application user pathway	24
4.2. Characteristic of the system tester population	27
4.3. System Performance Metrics	31
4.4. Results of the user experience questionnaire	32
4.5. Results of the system's response evaluation by medical doctors	38
5. DISCUSSION	40
5.1. ScreenGPT web application development	40
5.2. User experience questionnaire	41
5.3. ScreenGPT's answer evaluation by two medical doctors	42
5.4. Comparison with other AI-based tools	43
5.5. Limitations	44
5.6. Relevance and Future Utility	45
5.7. Summary of the main results of the dissertation	46
6. CONCLUSION	47
7. SUMMARY	49
8. REFERENCES	50

9. BIBLIOGRAPHY OF THE CANDIDATE’S PUBLICATIONS	63
9.1. Publications/abstracts related to PhD dissertation.....	63
9.2. Additional papers/abstracts not related to the PhD thesis	65
9.3. Awards.....	66
10. ACKNOWLEDGEMENTS	67

LIST OF ABBREVIATIONS

AI:	Artificial Intelligence
API:	Application Programming Interface
CUQ:	Chatbot Usability Questionnaire
FAQ:	Frequently Asked Questions
GPT:	Generative Pre-Trained Transformer
HITL:	Human-in-the-Loop
HPV:	Human Papillomavirus
LLM:	Large Language Model
LMIC:	Low- and middle-income countries
NER:	Named Entity Recognition
NLP:	Natural Language Processing
POS:	Part-of-speech
PSSUQ:	Post-Study System Usability Questionnaire
QA:	Question Answering
RLHF:	Reinforcement Learning with Human Feedback
RLHF:	Reinforcement Learning with Human Feedback
SIT:	Sexually Transmitted Infection
SLM:	Small Language Model
SUS:	System Usability Scale
UEQ:	User Experience Questionnaire
UI:	User Interface
UX:	User Experience
WHO:	World Health Organization

1. INTRODUCTION

1.1. Natural Language Processing

In the field of artificial intelligence (AI), natural language processing (NLP) is a key subdomain that focuses on enabling computers to understand, interpret, generate, and interact using human languages [1]. The primary goals of NLP are to enable machines to comprehend the meaning and context of human language, extract useful information from text data, and facilitate communication in a manner that feels natural to humans. NLP encompasses a broad range of computational techniques designed to enable machines to analyze, understand, and generate human language in a meaningful way [2]. The five main core tasks of NLP are the following:

1. Text preprocessing

Tokenization: Tokenization is the process of segmenting continuous text into smaller units known as tokens, which may correspond to words, subwords, or sentences. This step is essential as it forms the basis for subsequent linguistic analysis and downstream processing [3].

2. Syntactic Analysis

Part-of-Speech tagging: Part-of-speech (POS) tagging involves assigning grammatical categories, such as noun, verb, adjective, or adverb, to each token in a text. The objective is to identify the syntactic and morphological function of each word within its local and global context [4].

Parsing: Parsing refers to the process of constructing a syntactic structure (e.g., constituency trees or dependency graphs) that represents the grammatical relationships among the elements of a sentence. This structural representation facilitates further semantic and pragmatic analysis [5].

3. Semantic Analysis

Named entity recognition: Named Entity Recognition (NER) focuses on detecting and classifying proper names and other specific entities in text, including people,

organizations, locations, dates, and numerical expressions. NER is a critical component for information extraction and knowledge base population [6].

4. Text Representation & Transformation

Vectorization: Vectorization is the process of converting text into numerical representations that machines can understand and process. By mapping words, sentences, or documents to vectors in a high-dimensional space, it captures semantic relationships such that similar meanings are placed closer together.

5. Higher-Level Tasks

Machine translation: Machine translation is the automatic conversion of text or speech from one natural language to another. Modern approaches predominantly rely on neural network architectures, particularly sequence-to-sequence models and transformer-based systems, which have demonstrated significant improvements in translation quality [7].

Question answering: Question Answering (QA) systems are designed to automatically provide accurate and contextually appropriate answers to questions posed in natural language. QA integrates techniques from information retrieval, semantic parsing, and reading comprehension, and is often evaluated on benchmarks [8].

Text summarization: Text summarization aims to produce a condensed version of a document that preserves its most important content. Approaches can be broadly categorized into extractive methods, which select key spans from the source text, and abstractive methods, which generate novel sentences that paraphrase the original [9].

1.2. Generative Artificial Intelligence

Within NLP, generative artificial intelligence (GenAI) refers to models that can generate new and coherent language outputs, such as text or speech, based on patterns and information learned from existing data. GenAI systems can be further categorized into small language models (SLMs) and large language models (LLMs). Both SLMs and LLMs belong to the broader class of GenAI however, SLMs typically have fewer

parameters, lower computational requirements, and reduced operational costs [10-13]. In contrast, LLMs, due to their greater number of parameters and higher model complexity, generally achieve superior performance on complex language understanding and generation tasks [14, 15] (Table 1.).

	SLMs	LLMs
Size, Number of Parameters	110 million – 1.5 billion	several billions
Data Processing Capacity (context window)	512–1024 tokens	4096–32,000+ tokens (depending on model generation)
Resource Requirements	lower computational capacity	higher computational capacity
Customizability	a few thousand examples	broad and diverse datasets
Costs	10–100x lower	higher
Response Time, Speed	~1 second	2–3 seconds depending on complexity
Errors	fewer, harder to eliminate	higher susceptibility to bias

Table 1. Parameters of SLMs and LLMs

1.3. Large Language Models in Healthcare

LLMs represent a significant advancement in natural language processing, offering powerful capabilities for understanding and generating human-like text. In healthcare, their adoption is accelerating due to their ability to analyze large volumes of unstructured clinical data, assist medical professionals in decision-making, and improve patient engagement. Unlike traditional rule-based or task-specific models, LLMs are trained on massive datasets containing diverse linguistic patterns, which enables them to generalize across multiple healthcare applications.

LLMs also play an increasing role in patient-facing applications. They can power conversational agents that answer health-related questions, provide medication instructions, and offer lifestyle guidance. Importantly, these systems have the potential to increase accessibility to healthcare information for populations with limited medical literacy or access to specialists. In addition, multilingual LLMs can break down language barriers by providing real-time translation of medical advice, enhancing equity in global healthcare delivery. They offer substantial potential to strengthen public health efforts by improving the accessibility of healthcare information for diverse populations [16-18].

These technologies can enhance the efficiency of information dissemination, streamline routine communication tasks, and provide individuals with timely, understandable, and relevant health guidance [19-23]. By alleviating informational and administrative burdens, LLMs enable healthcare professionals to focus more effectively on strategic decision-making, health promotion, and the delivery of equitable, person-centered care [24-32].

Despite their promise, LLMs present notable challenges and risks in healthcare as well. Their training data may embed biases that could lead to inequitable recommendations across demographic groups. Moreover, the probabilistic nature of LLMs means they may generate hallucinations, plausible-sounding but factually incorrect statements. While LLMs are unlikely to fully replace healthcare professionals, they hold significant potential as an adjunct that enhances the efficiency and effectiveness of healthcare delivery [33-39].

1.3. Custom Generative Pre-trained Transformers

In the field of GenAI, GPTs (Generative Pre-trained Transformers) represent a class of neural network architectures specifically designed to process, understand, and generate human language with remarkable fluency and coherence [40-42]. These models, trained on vast corpora of general-domain text, have demonstrated impressive capabilities across a wide range of language tasks, including text generation, summarisation, translation, and question answering. Their versatility has enabled their application in diverse domains, from creative writing to technical support [19, 24, 28, 43-45].

While GPTs were originally conceived as general-purpose models capable of addressing a broad spectrum of linguistic tasks, there has been increasing demand for versions of these models that are optimised for specific domains or use cases [25, 26]. A custom GPT can be defined as a variant of a general GPT model that has been adapted, configured, or fine-tuned to meet the specific linguistic, informational, or functional requirements of a particular domain, organisation, or task [40, 42, 46, 47]. Creating a custom GPT involves modifying a base language model so that it can better understand and generate text that aligns with the specialised vocabulary, conventions, and knowledge structures of the target domain [48-51]. This adaptation process may include supervised

fine-tuning on domain-specific data, reinforcement learning with human feedback (RLHF), prompt engineering, or configuring system instructions to shape model behaviour without altering its underlying parameters.

Furthermore, creating a custom GPT often includes defining tailored instructions, system prompts, and guardrails to ensure appropriate, accurate, and ethical outputs [52-56]. Depending on the platform, developers and researchers may also incorporate structured knowledge bases, application programming interface (API) integrations, and role-specific behaviour through configuration tools without the need for additional model training [57, 58]. This multidisciplinary process requires expertise in machine learning, domain knowledge, and human-computer interaction to ensure that the resulting model is both effective and responsible in its deployment.

In the healthcare domain, this approach facilitates the creation of AI tools that are better tailored to the specific context, offering greater precision and responsiveness to the intricate nature of medical information and patient communication [59].

Custom GPTs, which differ from general-purpose models through domain-specific prompt design, fine-tuning, or deployment methodologies, offer distinct advantages [60, 61]. These include superior task performance by minimizing the processing of extraneous data, enhanced accuracy in specialized applications, and outputs that are more closely aligned with the professional standards, regulatory requirements, and communication conventions of the healthcare sector.

In the context of healthcare prevention, custom GPTs offer significant potential [62, 63]. These models can contribute to patient education, enhance access to pertinent health information, and promote adherence to guidelines by providing accurate, empathetic, and personalized guidance [64-68]. This dissertation positions itself within this emerging technological landscape by examining both the opportunities and challenges associated with designing a custom GPT-based system to support cervical cancer prevention initiatives. Specifically, the study seeks to explore how a custom GPT model can function as a patient education tool to improve the accessibility, clarity, and uptake of screening and vaccination recommendations.

1.4. Prompt engineering

Prompt engineering provides a practical and resource-efficient methodology for customising LLMs, including GPTs, to address domain-specific tasks without altering the model's parameters through fine-tuning [69, 70]. Rather than retraining or adapting the model weights, prompt engineering involves crafting carefully designed inputs, often including instructions, context, examples, or constraints, to guide the model toward producing outputs that align with the desired domain knowledge, style, or purpose [71, 72].

This approach capitalises on the in-context learning capabilities of GPT models, wherein the model uses information provided in the prompt to condition its response generation dynamically [73]. By embedding domain-specific terminology, procedural instructions, or representative examples within the prompt, practitioners can elicit responses that conform to the conventions and informational requirements of the target field [74, 75]. For instance, in healthcare applications, prompts can include explicit instructions to use formal clinical language or adhere to ethical guidelines [76]. Several strategies are commonly employed in prompt engineering for custom GPTs [77]:

1. **Instruction-based prompting:** The prompt specifies, in natural language, the desired behaviour of the model (e.g., “Summarise the following clinical note using formal medical terminology.”)
2. **Few-shot prompting:** The prompt includes a small number of input-output examples to illustrate the task [78, 79].
3. **Chain-of-thought prompting:** The prompt encourages the model to reason through a problem step-by-step to improve accuracy and reliability [80, 81].
4. **Role conditioning:** The prompt assigns the model a role (e.g., “You are an expert medical advisor. Provide a response that complies with clinical best practices.”).

An advantage of prompt engineering is that it requires minimal computational overhead compared to supervised fine-tuning. This makes it particularly suitable for rapid

prototyping, low-resource settings, and applications where model updates are impractical. However, the design of effective prompts is often an iterative process, requiring systematic testing, evaluation, and refinement. Through successive rounds of prompt adjustment, practitioners can optimise outputs for quality, factual accuracy, and domain relevance [82]. While prompt engineering does not modify the underlying model, it can serve as a powerful tool for creating custom GPTs in practice, enabling general-purpose models to function effectively within specialised contexts. Nevertheless, careful evaluation and validation are essential to ensure that the responses generated via prompt engineering meet professional and ethical standards, particularly in high-stakes domains such as healthcare.

1.5. Streamlit framework

Streamlit is an open-source Python framework designed to facilitate the rapid development and deployment of data-driven applications, particularly in the context of machine learning and artificial intelligence. Its declarative syntax and high-level abstractions allow developers to create interactive web interfaces with minimal code. When integrated with a cloud-based infrastructure, Streamlit enables scalable, real-time deployment of applications, allowing users to interact with custom GPT models directly through a web browser without the need for specialised software or local computational resources. As a cloud-based service, Streamlit supports seamless application hosting, continuous deployment, and access from any location, making it particularly well-suited for collaborative and remote environments. This architecture ensures high availability, facilitates version control, and enables applications to scale according to demand.

In the context of custom GPT implementations, this means users can access tailored, domain-specific language model outputs in real time, regardless of their technical background or physical location. Moreover, Streamlit provides a rich set of interactive features, including dynamic widgets and data visualisations, that enhance user engagement and facilitate more meaningful interaction with the underlying model. These capabilities are especially valuable in domains such as healthcare, education, and research, where iterative feedback, user adaptability, and interpretability are critical. Thus, by leveraging Streamlit as a cloud-based platform, developers can transform static

language models into dynamic, accessible tools for a broad range of end users, supporting both scalability and practical utility in real-world settings.

1.6. Cervical cancer screening

Cervical cancer screening is classified as a form of secondary prevention and is among the earliest screening protocols recommended by public health authorities. Secondary prevention refers to the strategies used to detect precancerous changes or early-stage cervical cancer before symptoms develop, with the goal of treating them promptly and preventing progression to invasive disease. It is endorsed by major health organizations, including the American Cancer Society and the World Health Organization (WHO) [83-85]. This type of screening enables the early identification of precancerous lesions and facilitates timely medical intervention, substantially lowering both the incidence and mortality rates associated with cervical cancer [86]. It has been widely recognized as one of the most effective and broadly implemented preventive health measures worldwide [87, 88].

Despite its demonstrated effectiveness, access to cervical cancer screening remains unequal, with pronounced disparities observed across different countries and population groups [89]. Low- and middle-income countries (LMICs) often face critical barriers due to limited healthcare infrastructure and resources [90]. Simultaneously, high-income countries continue to encounter inequities in screening access among marginalized populations, such as immigrants, individuals of low socioeconomic status, and minority communities. Various factors contribute to these disparities, including language barriers, financial limitations, lack of trust in healthcare systems, and inefficiencies in referral and follow-up mechanisms [91]. Additionally, cultural beliefs, stigma surrounding gynecological care, and geographic isolation further impede access to timely screening and diagnosis. Addressing these gaps requires innovative, multi-dimensional strategies, including the incorporation of digital health technologies [92, 93].

Many individuals lack adequate knowledge regarding cervical cancer and the importance of regular preventive screening, highlighting the urgent need for enhanced patient education and accessible, evidence-based information [94-96]. Strengthening awareness empowers people to recognize the value of early detection and adopt preventive health behaviors, thereby reducing the risk of disease progression.

Furthermore, offering platforms that allow individuals to engage in open, anonymous dialogue with digital systems may help overcome barriers of stigma, embarrassment, or discomfort that often arise in face-to-face interactions with healthcare providers, ultimately encouraging greater participation in screening programs [97-100].

One promising solution lies in the deployment of custom GPT-based conversational agents to support personalized patient education and engagement [101-103]. These AI-driven tools can mitigate both informational and emotional barriers by delivering accessible, context-specific content about cervical cancer screening and related risk factors. They can also play a crucial role in educating men about human papillomavirus (HPV) transmission and the public health benefits of HPV vaccination, thereby supporting broader prevention efforts [104-106]. HPV vaccination is a cornerstone of prevention, offering protection against high-risk HPV strains responsible for the majority of cervical cancer cases. Its implementation has significantly contributed to the decline in disease incidence. Custom GPT tools can enhance public understanding of HPV and vaccination, improve health literacy, and promote participation in screening programs, thereby contributing to more equitable outcomes in cervical cancer prevention [107, 108].

A further advantage of cervical cancer screening programs is their relevance to younger populations, who are generally more engaged with digital technologies, including internet-based platforms and AI applications. Screening is typically recommended to begin at age 30, though some guidelines advocate for earlier initiation at 25 [109]. The technological literacy of this demographic may support the adoption and effectiveness of AI-enhanced educational tools in healthcare.

2. OBJECTIVES

The primary objective of this study was to investigate whether a customised GPT-based system could improve access to relevant cervical cancer prevention information by providing personalised advice tailored to individual user needs:

- Develop a custom GPT specialised for cervical cancer screening to improve patient education and preventive behavior, by delivering reliable, publicly available information via a user-friendly interface by employing prompt engineering techniques.
- Gather feedback via a user experience questionnaire.
- Evaluating the system's responses by two independent medical doctors.

The underlying hypothesis posited that such a model would enhance the accessibility, comprehensibility, and adherence to cervical cancer screening guidelines, thereby contributing to improved patient education and supporting patients in locating and understanding specific information.

3. METHODS

3.1. ScreenGPT system architecture

3.1.1. Backend

As part of the implementation process, we registered for the Streamlit base package and utilized its framework to develop our web application. Streamlit was chosen due to its reactive programming model and seamless integration with Python, allowing for rapid prototyping and interactive data-driven interfaces. To build the application, we imported the Streamlit library, which provides a high-level API for creating the user interface. For conversational capabilities, we integrated the OpenAI API, enabling natural language interaction through GPT-based models [110].

To ensure the secure management of sensitive information such as API keys, we utilized the `st.secrets` functionality provided by the Streamlit framework. This feature enables the storage of confidential credentials in a secure and encrypted environment, preventing the exposure of keys within the source code. By separating secrets from the main application logic, this approach reduces the risk of unauthorized access and enhances the overall security posture of the system. Additionally, the execution environment in Streamlit Cloud is designed so that the application code is only retained during active runtime. Once execution is complete, the code and associated secrets are no longer stored on the platform. This temporary storage model contributes to the protection of intellectual property and sensitive components, ensuring that proprietary logic and confidential data remain secure.

The web application is capable of handling concurrent interactions from over 100 users, facilitated by Streamlit's session management capabilities through the `st.session_state` API. This mechanism ensures that each user's input, state, and progress within the application are maintained independently, allowing for personalized and isolated user experiences during simultaneous usage. As a result, the application can scale effectively to support multiple sessions without conflicts or data overlap between users.

To structure and manage the communication between user queries and model responses, we employed the JSON library. This allowed for efficient serialization and deserialization of data exchanged between the frontend and backend. For handling HTTP requests, particularly for database interactions such as querying and storing conversation history, we used the requests library. For tasks requiring structured output (e.g., JSON-formatted Python objects such as dictionaries, with predefined keys and values extracted from the user's input), we employ the GPT-4 model. In cases where a structured response is unnecessary, we utilize the GPT-4o model, which offers a more cost-effective solution while maintaining a high quality of response generation. It is important to note that the response latency between the two models differs slightly.

ScreenGPT has been implemented in two languages: English and Hungarian. For language management, we designed a language data loading function and defined the relevant text elements (e.g., greetings, disclaimers, etc.) in JSON files, where keys correspond to specific content elements, each with translations in both English and Hungarian.

To facilitate interaction, we defined thematic topics through which users can engage with the chatbot. At present, the system supports cervical cancer screening; however, the framework is extensible, and our future development plans include expanding the platform to cover breast cancer screening and colorectal cancer screening.

At the end of the workflow, the web application displays a chat input field that enables users to continue the conversation. While the system processes the user's query and the assistant generates a response, the interface presents an informational block indicating that the application is working. This feature prevents the impression of inactivity by displaying a "please wait" message, thereby maintaining a smooth and transparent user experience even during short response delays.

ScreenGPT's response time was recorded using the Python (`datetime.now`) module across 11 trials. For each trial, the duration was measured from the moment OpenAI's response was received until it was displayed on the Streamlit interface.

3.1.2. Frontend

Page-level configuration, including layout and display parameters, was accomplished using the `st.set_page_config` function provided by the Streamlit framework. This allowed us to define the application's title, icon, layout orientation, and initial state. Furthermore, we implemented a flexible column-based layout with dynamically adjustable widths, enabling multiple interface components to be presented side by side. This layout is responsive, automatically adapting to different screen resolutions and device types, thereby ensuring a consistent and user-friendly experience across both desktop and mobile platforms.

A landing page was designed to serve as the entry point of the application. This page includes the application name, key descriptive information, and a logo in PNG format, which was rendered using the `st.image` function. To enhance navigation and usability, we incorporated a sidebar menu, positioned on the left-hand side of the interface. The sidebar was divided into three distinct sections to logically separate elements and center the logo, thereby improving the visual hierarchy and aesthetics.

In addition, we utilized the `st.markdown` function, which supports rendering of custom HTML and Markdown content. This was particularly useful in cases where the default `st.title` or `st.header` functions did not support the desired formatting or styling requirements. By leveraging Markdown and embedded HTML, we were able to maintain a consistent design language throughout the application while also extending the presentation capabilities of the Streamlit framework.

3.1.3. Prompts

The application also provides differentiated pathways depending on the user profile, which are male users, female users under 25 years, and female users over 25 years. These pathways are implemented through separate JSON files, which define distinct system role prompts. This design ensures that the conversational flow is both personalized and context-sensitive, thereby enhancing user experience. The prompts were developed in accordance with international

and national guidelines, including the recommendations of the American Cancer Society and the World Health Organization for cervical cancer screening and the management of precancerous lesions, as well as evidence drawn from relevant peer-reviewed scientific literature.

In addition, two Hungarian guidelines were incorporated, namely the Professional Protocol of the Hungarian Ministry of Health and the recommendations of the Hungarian National Public Health and Pharmaceutical Center. The used prompts are included in the Supplementary 1 file.

Within the application, three fundamental roles can be distinguished, each contributing to the structuring of the interaction process. The system role establishes the framework of the interaction and provides the contextual and background information required to guide the model's behavior. The user role conveys the input submitted by the end-user, while the assistant role is responsible for generating, recording, and communicating the chatbot's responses. This structure enables information persistence throughout the dialogue. Once a user provides input, the system retains it, and all subsequent interactions automatically re-include the relevant prior data. As a result, the assistant can generate reliable, coherent, and tailored responses that reflect the full conversational history. Based on these defined keys, the conversation progresses dynamically.

Furthermore, we also defined interactive buttons that allow users to select from predefined options. This feature not only simplifies navigation but also ensures that users can engage with the system efficiently and consistently, even when facing complex decision pathways. Prompts were developed using Basic, Structured, Reasoning & Thought Process, Knowledge & Context Enhancement, and Creative & Language Manipulation techniques (Table 2.).

Category	Technique	Description
Basic Prompting Techniques	Zero-Shot Prompting	No examples provided; the model infers from context.
	Few-Shot Prompting	A few examples are given to guide the model's response.
	One-Shot Prompting	Only one example is included before asking for a response.
Structured Prompting Techniques	Instruction-Based Prompting	Explicit instructions direct the model on how to respond.
	Template-Based Prompting	Predefined templates structure the input for consistency.
Reasoning & Thought Process Techniques	Chain-of-Thought (CoT) Prompting	Encourages step-by-step reasoning to improve logical accuracy.
Knowledge & Context Enhancement Techniques	Persona-Based Prompting	Assigns the model a specific role to generate domain-specific responses (e.g., "You are a medical researcher").
Creative & Language Manipulation Techniques	Style Transfer Prompting	Modifies the tone, style, or complexity of the response.

Table 2. Prompt engineering techniques

3.2. User experience measurement

Participants evaluated the web application by interacting with the custom GPT model for cervical cancer screening, followed by completing an anonymous 10-question online survey. The user experience questionnaire has been put in the menu options, and those who tried the ScreenGPT web application were able to fill out the questionnaire after that. The questionnaire was specifically developed for this study, drawing on established user experience (UX) tools such as the Chatbot Usability Questionnaire (CUQ) and the User Experience Questionnaire (UEQ), but tailored to meet the specific needs of the application [75]. A pre-test with four study team researchers assessed clarity, relevance, and effectiveness, with feedback used to refine the questions.

Responses were captured using a four-point scale: 1 (totally agree), 2 (mostly agree), 3 (marginally agree), and 4 (do not agree), offering clear, interpretable results. To support demographic analysis, participants were grouped into three categories: females under 25 (Female <25), females 25 and older (Female >25), and all males (Male). Data collection took place over three weeks, focusing on relevance, customization, usability, and accuracy as part of the human-in-the-loop (HITL) evaluation.

3.3. Evaluation of the system's responses by two independent medical doctors

To comprehensively assess the performance of the ScreenGPT model in addressing inquiries related to cervical cancer, an independent evaluation was conducted by two medical doctors. A set of thirty frequently asked questions concerning cervical cancer screening was curated from reputable sources, including the Hungarian National Public Health Center, PennMedicine, and the Mayo Clinic (Supplementary 2). The model was prompted to generate responses to each of these questions. Subsequently, both the questions and the corresponding model-generated answers were provided to the two independent medical experts for evaluation.

The evaluators were instructed to assess each response based on three key criteria:

- Accuracy: The extent to which the answer aligns with current medical knowledge and established clinical guidelines.
- Clarity: The degree to which the response is comprehensible, well-structured, and logically presented.
- Usefulness: The practical value of the response for clinical application or patient education.

Each response was rated on a 4-point Likert scale: 1 = Strongly Disagree / Not Accurate, 2 = Disagree, 3 = Agree, and 4 = Strongly Agree / Very Accurate. Evaluators were also encouraged to provide brief written comments to justify or expand upon their ratings where appropriate.

3.4. Statistical analysis

Descriptive statistics were used to summarize the tester demographics. Continuous variables, such as age, were summarized using the mean and standard deviation, while testers were subsequently categorized into age groups, with percentage distributions calculated to establish a comprehensive demographic profile. To examine associations between age groups and user experience variables, the Chi-Square test was applied, appropriate for categorical data. A significance level of $p < 0.05$ was used to identify meaningful differences. Gender-related differences in user experience were assessed using the Mann–Whitney U test, suitable for comparing two independent groups with non-normally distributed or ordinal data.

To analyze the evaluations provided by the two medical doctors (referred to as MD1 and MD2), both descriptive statistics and the Wilcoxon Signed-Rank Test were employed to determine inter-rater consistency and assess any significant differences in their scoring.

All the statistical analyses were performed via the SPSS 29 (IBM®) software package.

4. RESULTS

4.1. ScreenGPT architecture

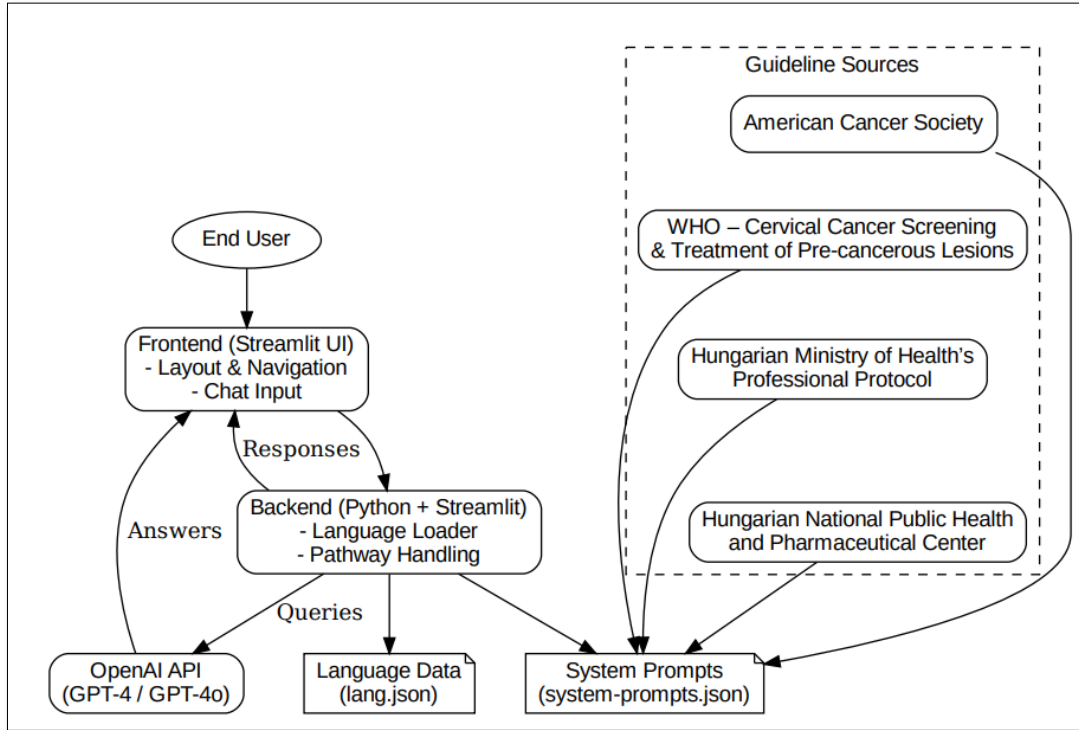


Figure 1. ScreenGPT system architecture

The architecture of ScreenGPT was designed to integrate a user-friendly web interface with a flexible backend and evidence-based knowledge sources, ensuring reliable and personalized conversational support for cancer screening.

At the frontend, the system employs the Streamlit framework, which provides a responsive and interactive user interface. This layer handles page configuration, sidebar navigation, and chat input, and it displays real-time feedback such as an informational “please wait” message during processing.

The backend, implemented in Python with Streamlit, manages the application logic. Its responsibilities include loading bilingual resources from JSON files (English and Hungarian), handling user pathways (male, female under 25, female 25 and older), and mapping predefined prompts to interactive conversation flows. The backend also processes free-text input and prepares structured queries for the models when required.

The external services include the OpenAI API, which powers conversational responses. Two models are utilized: GPT-4, for tasks requiring structured JSON-formatted outputs, and GPT-4o, for cost-effective generation of unstructured conversational responses. Additional resources include lang.json (language data) and system_prompts.json (pathway-specific prompts). Importantly, the predefined prompts were developed in alignment with international and national guidelines, including the American Cancer Society, the WHO recommendations for cervical cancer screening and the treatment of pre-cancerous lesions, the Hungarian Ministry of Health's Professional Protocol, and the Hungarian National Public Health and Pharmaceutical Center (NNK) (Figure 1.).

4.2. ScreenGPT web application user pathway

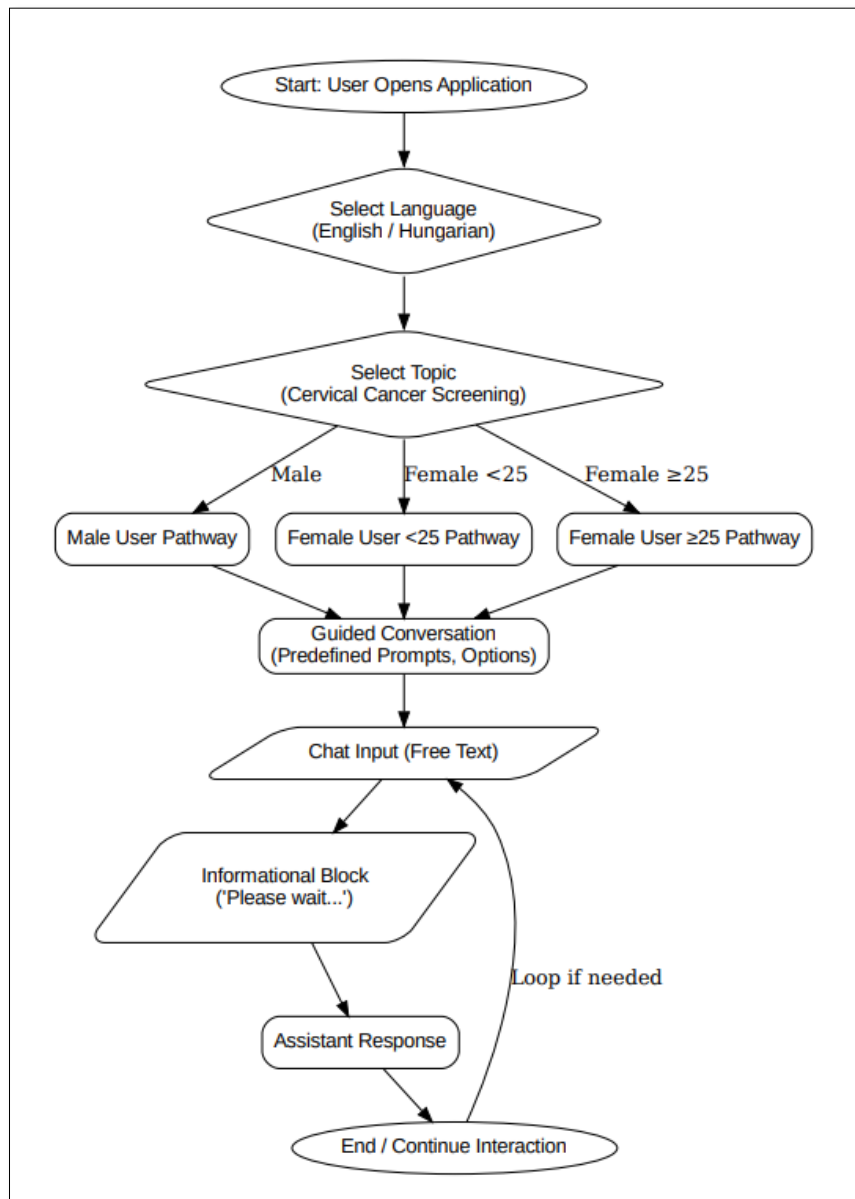


Figure 2. ScreenGPT user pathway

ScreenGPT web application's full user pathway is presented in Figure 2. When entering the system, the user selects the preferred language (English or Hungarian) and then chooses a thematic topic (currently cervical cancer screening). Based on the user's profile (male, female under 25, or female 25 and older), the application directs the user to a personalized pathway defined in JSON files.

The conversation progresses through guided options presented as buttons, followed by a free-text chat input when needed. While the assistant generates a response,

the interface displays an informational block (“please wait”) to prevent the perception of inactivity. The assistant’s response is then shown, and the interaction may loop until completion.

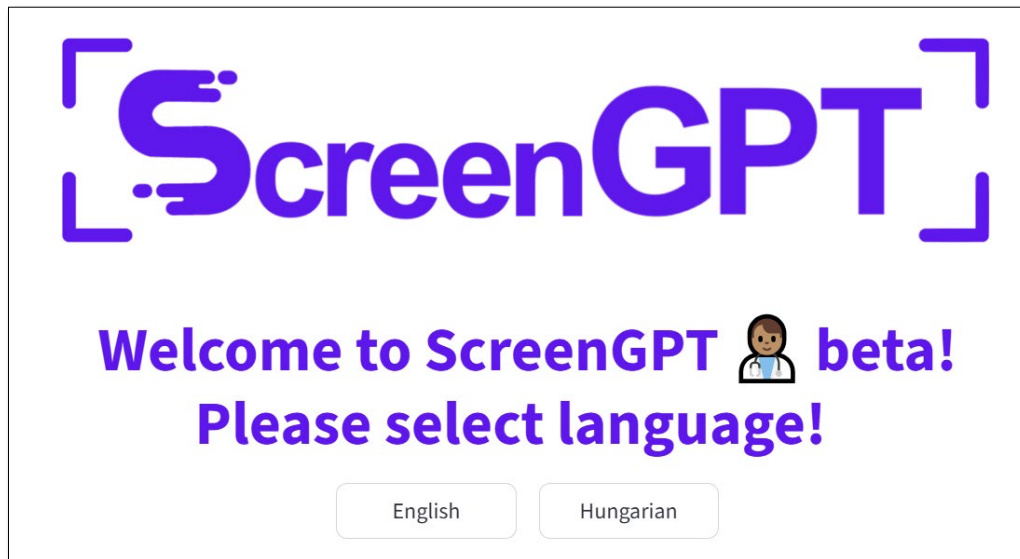


Figure 3. ScreenGPT landing page

Upon accessing the application, users are first directed to the landing page, which serves as the initial interface for interaction. A key feature of this landing page is the option for language selection, enabling users to tailor the interface to their preferred linguistic context. At present, the system supports two languages: English and Hungarian (Figure 3.). This functionality ensures accessibility for a broader user base and reflects an emphasis on usability and inclusivity. The multilingual capability also demonstrates the system’s scalability, as the underlying architecture allows for the straightforward integration of additional languages in future iterations. By providing a clear entry point and user-friendly customization through language preferences, the landing page establishes an intuitive and accessible environment, thereby enhancing the overall user experience from the outset.

Hello! I'm ScreenGPT, your professional healthcare assistant. My mission is to help you lead a healthier lifestyle and provide you with information for prevention. Please choose the topic you are interested in!

The topic I would like to talk about is HPV infection and cervical cancer prevention

Before we proceed, could you please tell me your gender and age? This information will help me provide the most relevant information for you.

Biological gender: Female (dropdown menu)
 Age (years): 18 (input field with +/- buttons)

OK (button)

Figure 4. ScreenGPT question flow

To provide you with the most accurate information, could you please tell me your gender and age?

I'm a 30 years old woman.

At 30 years old, you are at the age when regular cervical cancer screening is highly recommended according to professional guidelines. Have you ever had a cervical cancer screening, such as a Pap test or an HPV test?

No, I didn't

Yes, within 3 years

Yes, more than 3 years ago

Figure 5. Predefined questions in the ScreenGPT web application

After completing the language selection process, users are directed to the question flow of the ScreenGPT web application. This stage is designed to collect essential background information in a structured and user-friendly manner. The interface presents users with a series of predefined response options, which can be selected by clicking,

thereby minimizing input effort and reducing the likelihood of typographical errors (Figure 4. and Figure 5.).

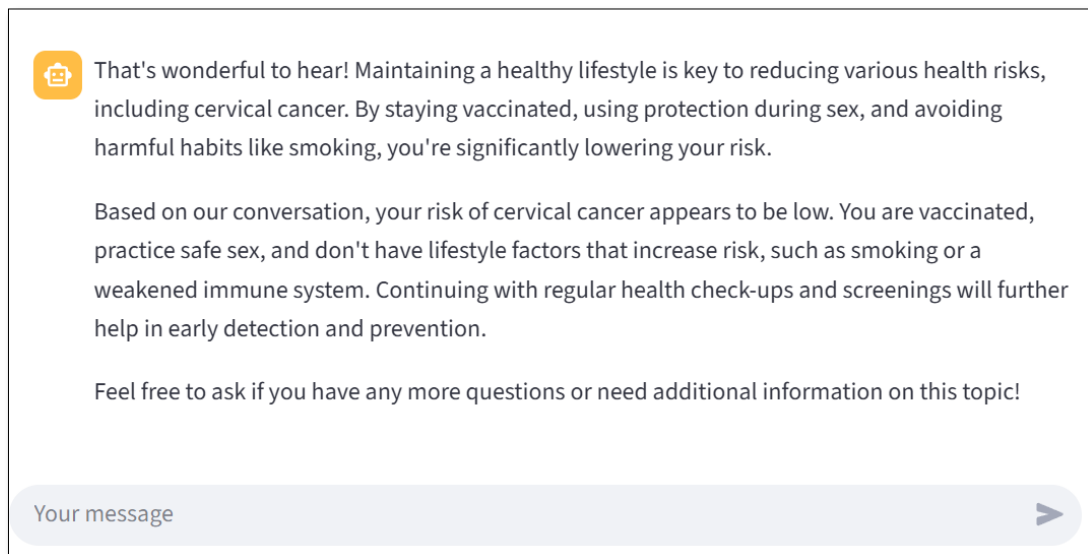


Figure 6. Free text questions in the ScreenGPT web application

Once users have navigated through and completed all of the required questions, the application automatically generates a basic anamnesis summary. This summary serves as an initial health profile, consolidating the collected data into a coherent and clinically relevant format. The automated generation of such a summary facilitates both efficiency and consistency, ensuring that essential information is captured systematically. In addition to this initial summary, the application also provides users with the opportunity to pose further questions or elaborate on their health concerns.

This feature is intended to extend beyond the predefined options and support a more personalized interaction, thereby balancing standardization with flexibility. In doing so, ScreenGPT aims to combine the advantages of structured data collection with the responsiveness of an intelligent conversational system (Figure 6.).

4.2. Characteristic of the system tester population

Data were collected over a three-week period, during which a total of 115 individuals tested the application, comprising 87 females (76%) and 28 males (24%). The mean age of participants was 41.2 years (SD = 13.5), with ages ranging from 18 to 74 years. The sample was predominantly female. Table 3. illustrates the age distribution across the two biological sexes. Among the 87 female participants, 16 individuals (18%)

were younger than 25 years, while the remaining 71 (82%) were aged 25 years or older. Based on age and sex, participants were categorized into three subgroups.

The first subgroup consisted of female testers under 25 years old (Female <25), the second subgroup included female testers over 25 years old (Female >25), and the third subgroup comprised male testers (Male).

Variables	Mean or N	SD or %
Biological gender		
Female	87	76%
Male	28	24%
Age Male (years)	38	±13.5
18-24	3	11%
25-34	10	36%
35-44	8	29%
45-54	3	11%
55+	4	14%
Age Female (years)	42	±12.7
18-24	13	15%
25-34	10	11%
35-44	17	20%
45-54	29	33%
55+	18	21%

Table 3. Age distribution of the system tester population

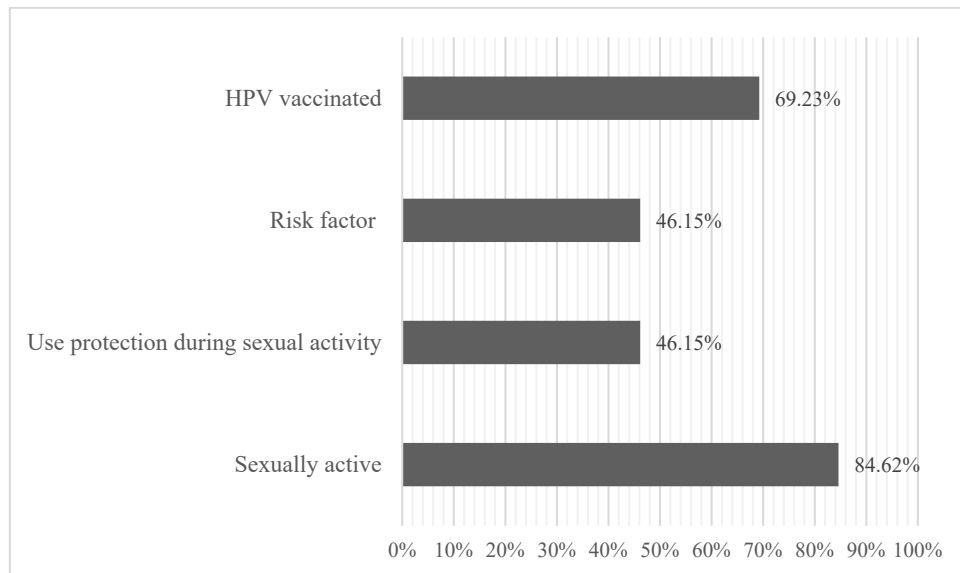


Figure 7. Female <25 (N=13)

The majority of participants (84.62%) reported being sexually active; however, only 46.15% of women under the age of 25 (Female <25) reported consistent use of protection during sexual activity. These findings underscore the presence of behavioral risk factors associated with sexually transmitted infections (STIs), particularly human papillomavirus (HPV), which is a well-established etiological agent in the development of cervical cancer. HPV vaccination coverage varied substantially across subgroups. Among females under 25, 69.23% reported being vaccinated against HPV.

This comparatively high rate is likely attributable to the introduction of publicly funded HPV vaccination programs in Hungary, which have offered free immunization to 12-year-old girls in the 7th grade since 2014. The impact of such governmental policy changes is evident in the elevated vaccination rates observed among younger female participants. In terms of cervical cancer risk factors and screening history, 46.15% of women under 25 reported at least one recognized risk factor, such as smoking, associated with increased cervical cancer susceptibility (Figure 7.).

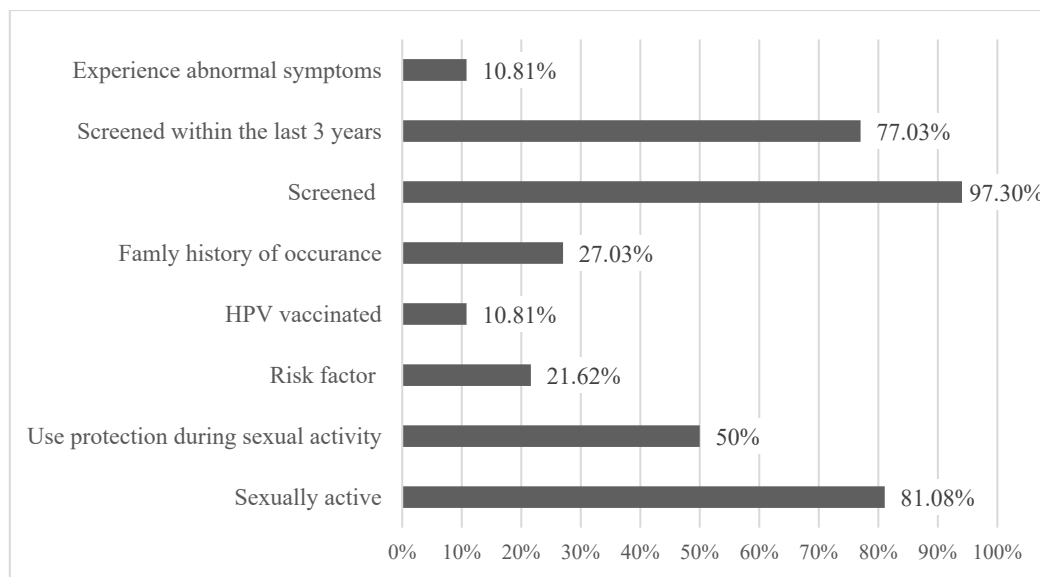


Figure 8. Female >25 (N=74)

Among females aged 25 and older, only 10.81% reported having received the HPV vaccine. In terms of cervical cancer risk factors and screening history, 21.62% reported at least one recognized risk factor, such as smoking, associated with increased cervical cancer susceptibility. Among women in the ≥ 25 age group, 97.3% indicated they

had undergone cervical cancer screening at least once in their lifetime, and 77.03% had been screened within the previous three years.

These findings suggest a relatively high level of adherence to national cervical cancer screening recommendations, although gaps in coverage remain. Furthermore, within the same age group, 27.03% of respondents reported a known family history of cervical cancer, and 10.81% acknowledged experiencing recent abnormal symptoms. These self-reported indicators may warrant further investigation in the context of early detection and risk stratification strategies (Figure 8).

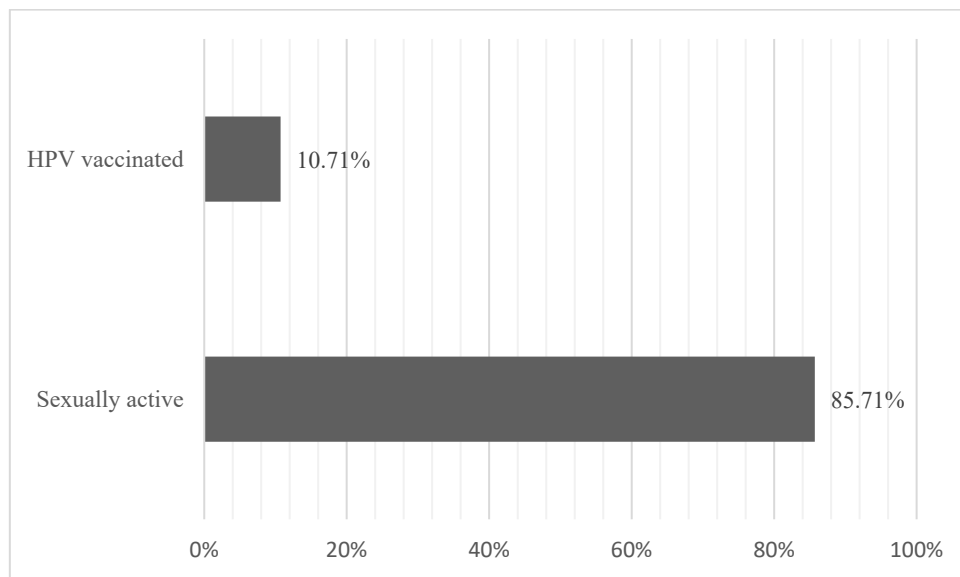


Figure 9. Male (N=28)

Among males, 85.7% said that they are sexually active, but only 10.71% reported having received the HPV vaccine. These low uptake levels highlight a significant public health concern and the need for more targeted and inclusive vaccination initiatives, especially for males, who are frequently underrepresented in HPV prevention efforts. Notably, in Hungary, boys have also been included in the national HPV vaccination program in recent years, typically receiving the vaccine at the age of 12, which may contribute to increased HPV immunisation rates in the population in the near future (Figure 9).

4.3. System Performance Metrics

Response time was assessed using Python's `datetime.now` function across 11 test iterations. Each measurement captured the time interval between receiving a response from the OpenAI API and rendering it on the Streamlit user interface.

The application was deployed via Streamlit's cloud-based service, so it is important to note that the recorded response times were influenced not only by the performance of the underlying model but also by factors such as internet connectivity and browser efficiency.

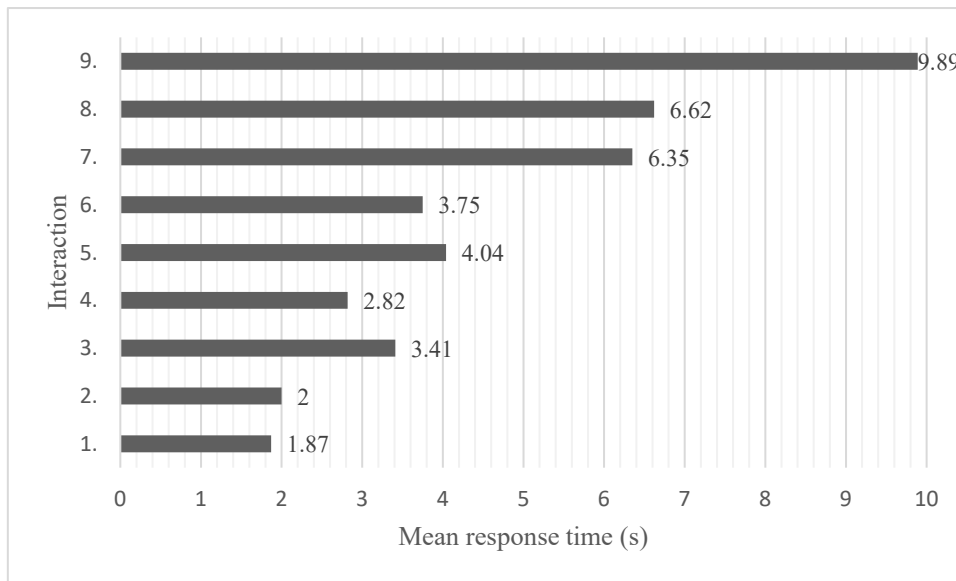


Figure 10. System response time per interaction

OpenAI's API generally responded within a range of 500 milliseconds to 3 seconds. However, this latency varied depending on factors such as query complexity, network bandwidth, and the user's browser environment. Notably, a gradual increase in average response time was observed across sequential interactions, which may be attributed to the system's need to reprocess the full conversational context at each step (Figure 10).

Throughout the testing process, no API failures, timeouts, or user input errors were encountered, indicating a high level of system stability. In terms of scalability, the combined use of Streamlit and OpenAI's API provides a robust framework capable of accommodating increased user demand, contingent upon effective management of API rate limits and allocation of sufficient cloud resources.

4.4. Results of the user experience questionnaire

Out of 115 individuals who participated in the testing of the web application, 112 completed the feedback form, offering essential data for assessing the application's performance and overall user experience (Table 4.).

Overall, responses to items related to design attributes were notably positive. Participants evaluated aspects such as initial impression, interface clarity, simplicity of use, and system interactivity. A substantial proportion (81.25%) reported a generally positive overall experience, while 15.18% indicated a highly favorable perception. Conversely, 3.57% reported a less favorable impression. The clarity of the user interface received particularly strong ratings. More than half of the respondents (51.79%) strongly agreed that the interface was clear and required no prior training, with an additional 38.39% mostly agreeing. A minority (8.93%) marginally agreed, and only 0.89% disagreed.

These results underscore the overall clarity of the interface, while also suggesting areas where improvements in accessibility and user guidance may be beneficial. In relation to system usability, the majority of participants affirmed the ease of use, 40.18% strongly agreed, and 54.45% mostly agreed with the statement that the system was easy to use. Only 4.46% marginally agreed, and 0.89% disagreed. These findings reinforce the importance of intuitive design in supporting efficient and satisfying user interaction, particularly across diverse user profiles. Regarding system interactivity, 12.5% of participants perceived the system as highly engaging, while 70.54% mostly agreed with this. A smaller proportion (16.07%) marginally agreed, and 0.89% disagreed. Although the majority of users responded favorably, these results indicate potential value in enhancing interactive elements to further support user engagement and satisfaction.

Indicators related to the Tailored User Experience, including system customization, relevance of responses, personalization of advice, and adaptability to individual needs, were evaluated positively by the majority of participants. System-generated responses were generally perceived as relevant, with 22.9% of respondents rating them as entirely relevant, 74.11% as mostly relevant, and 3.57% as marginally relevant. Notably, no participants deemed the responses irrelevant. These findings underscore the critical role of maintaining a high degree of response relevance to support meaningful and context-aware user interactions. Concerning personalization, 12.4% fully

agreed that the system's responses were tailored to their specific queries, while 62.5% mostly agreed. However, 13.39% marginally agreed, and 11.62% felt the responses were overly generic. Regarding adaptability, 22.32% believed the system completely adapted to their needs, 63.39% mostly agreed, 11.61% marginally agreed, and 2.68% felt it did not respond effectively.

These findings indicate a need for further refinement in the system's personalization and adaptability features to improve user satisfaction and inclusivity. In addition, participants broadly acknowledged the system's intuitive interface and empathetic response style as key contributors to a positive and engaging user experience. These features are particularly important in the context of digital health interventions, where user trust, comfort, and engagement are closely linked to proactive health behaviors such as adherence to screening recommendations and vaccination uptake.

User Loyalty Indicators, specifically, participants' willingness to reuse the application in the future and recommend it to others, were less favorable compared to the other two evaluated dimensions (Perceived Usability and Tailored User Experience). Only 6.25% of participants indicated that they would definitely use the system again, while 36.61% reported that they would most likely do so. Notably, 15.18% stated that they would not use the application again.

In terms of recommendations, the overall responses were more encouraging. A total of 20.54% of participants strongly agreed that they would recommend the tool to others, and 63.39% mostly agreed. However, 12.50% marginally agreed, and four participants (3.57%) explicitly indicated that they would not recommend the system. These findings suggest that while initial perceptions and personalized features were positively received, there may be limitations in sustained engagement and endorsement. This highlights the need for further refinement in areas that influence long-term user satisfaction and advocacy.

		1		2		3		4	
		Totally agree		Mostly agree		Marginally agree		Do not agree	
		<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Design Attributes	Positive first impression	17	15.18%	91	81.25%	4	3.57%	0	0%
	Clarity of the interface usability	58	51.79%	43	38.39%	10	8.93%	1	0.89%
	Simplicity of system usage	45	40.18%	61	54.46%	5	4.46%	1	0.89%
	System interactivity level	14	12.50%	79	70.54%	18	16.07%	1	0.89%
Tailored User Experience	Customization of the system	9	8.04%	62	55.36%	38	33.93%	3	2.68%
	Relevance of responses	25	22.32%	83	74.11%	4	3.57%	0	0%
	Personalization of advice	14	12.50%	70	62.50%	15	13.39%	13	11.61%
	Adaptability to personal needs	25	22.32%	71	63.39%	13	11.61%	3	2.68%
User Loyalty Indicators	Future reuse	7	6.25%	41	36.61%	47	41.96%	17	15.18%
	Recommend to others	23	20.54%	71	63.39%	14	12.50%	4	3.57%

Table 4. Results of the user experience questionnaire

The mean age of participants who completed the user experience questionnaire was 39.69 years. Among those who rated the user interface as highly interactive, the average age was 37.64 years; those who found it moderately interactive had a mean age of 39.22 years. Conversely, participants who considered the interface less interactive had a mean age of 42.06 years, while the average age among those who rated it as not interactive at all was 63 years. Despite these age-related trends, the relationship between age and perceived interactivity was not statistically significant when tested using the Chi-Square Test ($p = 0.911$).

A statistically significant association was observed between age group and perceived personalization of system responses ($p = 0.047$) (Table 5.). No significant associations were identified for the remaining nine variables. This suggests that while age may influence subjective perceptions of interactivity, it is not a decisive factor across age groups. These findings underscore the potential need for adaptive interface designs to ensure usability across a broad demographic spectrum, particularly for digital health tools such as ScreenGPT. It is also important to recognize that age is not the only determinant influencing perceptions of interactivity. Other sociocultural and contextual variables, such as geographic location, education level, digital literacy, infrastructure, and income,

may play a significant role. However, these factors were not captured by the current user experience questionnaire, indicating a potential direction for future research.

	Age groups	N	%	Mean	SD	p-values
Positive first impression	18-24	22	19.64%	1.86	0.35	0.881
	25-34	21	18.75%	1.86	0.36	
	35-44	25	22.32%	1.92	0.49	
	45-54	28	25.00%	1.86	0.45	
	55+	16	14.29%	1.94	0.44	
Clarity of the interface usability	18-24	22	19.64%	1.55	0.51	0.175
	25-34	21	18.75%	1.57	0.81	
	35-44	25	22.32%	1.52	0.71	
	45-54	28	25.00%	1.61	0.69	
	55+	16	14.29%	1.75	0.77	
Simplicity of system usage	18-24	22	19.64%	1.55	0.51	0.631
	25-34	21	18.75%	1.62	0.80	
	35-44	25	22.32%	1.72	0.61	
	45-54	28	25.00%	1.64	0.56	
	55+	16	14.29%	1.81	0.54	
System interactivity level	18-24	22	19.64%	2.00	0.55	0.914
	25-34	21	18.75%	2.05	0.50	
	35-44	25	22.32%	1.96	0.54	
	45-54	28	25.00%	2.14	0.71	
	55+	16	14.29%	2.19	0.66	
Customization of the system	18-24	22	19.64%	2.27	0.70	0.612
	25-34	21	18.75%	2.29	0.56	
	35-44	25	22.32%	2.28	0.79	
	45-54	28	25.00%	2.25	0.65	
	55+	16	14.29%	2.56	0.51	
Relevance of responses	18-24	22	19.64%	1.77	0.53	0.877
	25-34	21	18.75%	1.86	0.48	
	35-44	25	22.32%	1.76	0.44	
	45-54	28	25.00%	1.89	0.42	
	55+	16	14.29%	1.75	0.58	
Personalization of advice	18-24	22	19.64%	2.18	0.64	0.047
	25-34	21	18.75%	2.25	1	
	35-44	25	22.32%	2.2	0.83	
	45-54	28	25.00%	2.17	0.83	
	55+	16	14.29%	2.55	0.93	
Adaptability to personal needs	18-24	22	19.64%	1.82	0.50	0.172
	25-34	21	18.75%	2.05	0.74	
	35-44	25	22.32%	1.92	0.70	
	45-54	28	25.00%	1.86	0.52	
	55+	16	14.29%	2.19	0.91	
Future reuse	18-24	22	19.64%	2.55	0.80	0.188
	25-34	21	18.75%	2.52	0.75	
	35-44	25	22.32%	2.52	0.77	
	45-54	28	25.00%	2.68	0.82	
	55+	16	14.29%	3.19	0.83	
Recommend to others	18-24	22	19.64%	1.86	0.71	0.853

25-34	21	18.75%	1.95	0.59
35-44	25	22.32%	2.00	0.71
45-54	28	25.00%	2.07	0.77
55+	16	14.29%	2.06	0.68

Table 5. Age-Based Comparison Using the Chi-Square Test

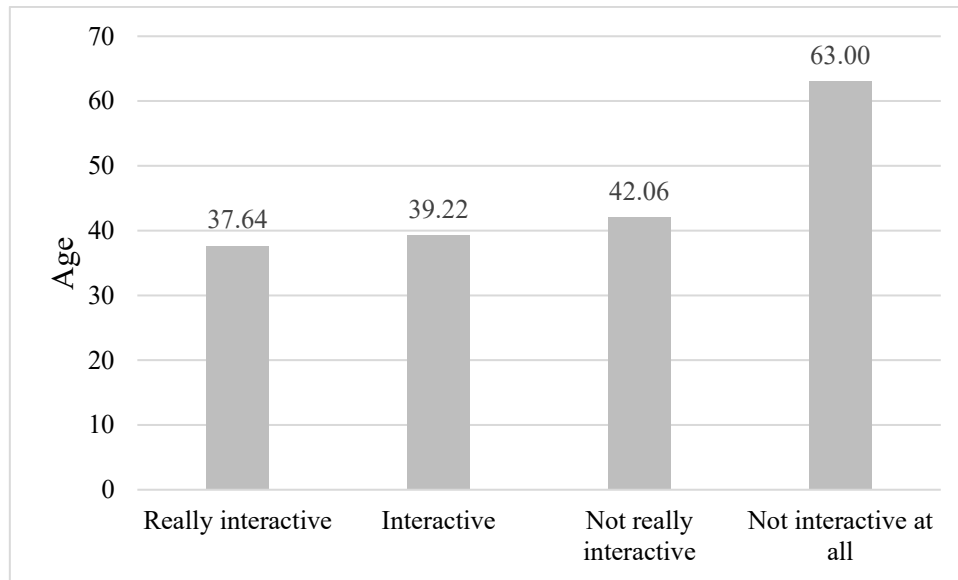


Figure 11. Web Application Interactivity Perceived by Different Age Groups

Further analysis of the data revealed a trend indicating that younger participants (mean age below 40 years) generally reported a higher level of engagement with the system interface. They described the platform as more interactive and user-friendly when compared to participants aged over 40 (Figure 11.). This pattern may reflect younger users' greater familiarity and comfort with digital technologies and AI-driven interfaces. In contrast, although older participants were able to navigate the application, they were more likely to perceive the interface as less intuitive.

Regarding gender, analysis using the Mann–Whitney U test revealed a significant association between gender and perceived customization of the system ($p = 0.037$). Additionally, an association between gender and the variable intention for future reuse approached statistical significance ($p = 0.069$), although it did not meet the conventional threshold (Table 4). These findings suggest that both age and gender may influence specific aspects of user experience, particularly personalization and system

customization. They should be considered in the future development and evaluation of similar applications (Table 6.)

	Gender	N	Mean	SD	p-values
Positive first impression	Male	35	1.97	0.43	0.137
	Female	77	1.84	0.38	
Clarity of the interface usability	Male	35	1.46	0.66	0.143
	Female	77	1.65	0.7	
Simplicity of system usage	Male	35	1.69	0.58	0.671
	Female	77	1.65	0.62	
System interactivity level	Male	35	2.03	0.51	0.793
	Female	77	2.08	0.62	
Customization of the system	Male	35	2.51	0.66	0.037
	Female	77	2.22	0.64	
Relevance of responses	Male	35	1.86	0.55	0.587
	Female	77	1.79	0.44	
Personalization of advice	Male	35	2.37	0.84	0.227
	Female	77	2.18	0.81	
Adaptability to personal needs	Male	35	2.03	0.66	0.342
	Female	77	1.91	0.67	
Future reuse	Male	35	2.86	0.73	0.069
	Female	77	2.57	0.83	
Recommend to others	Male	35	1.97	0.71	0.695
	Female	77	2	0.69	

Table 6. Gender-Based Comparison Using the Mann–Whitney U Test

Gender-based differences in perceived informational value were also observed. While the majority of participants across both genders acknowledged the relevance of the information provided by ScreenGPT, subtle variations emerged.

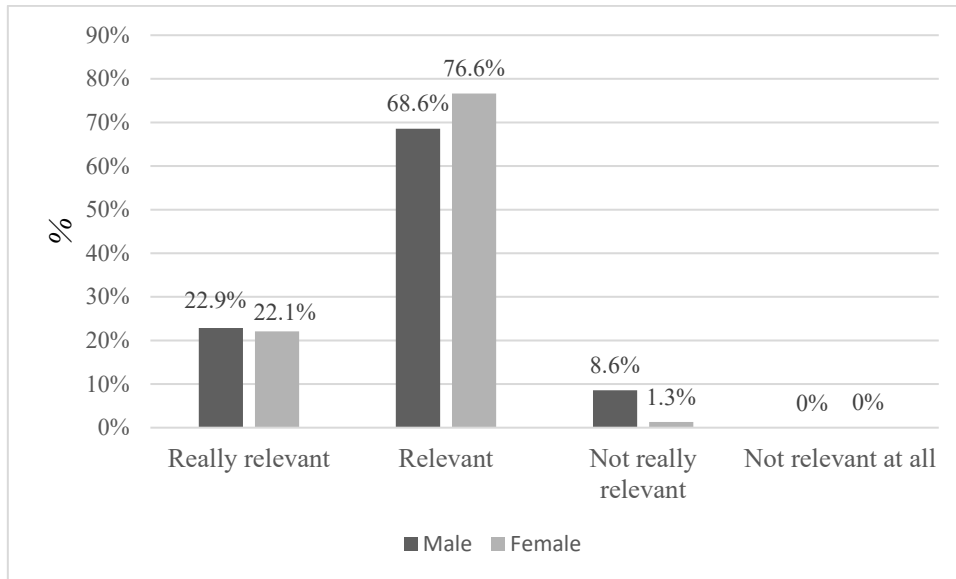


Figure 12. Gender-Based Perceptions of Information Relevance Provided by ScreenGPT

Among female respondents, 22.1% rated the information as entirely relevant and 76.6% as mostly relevant. Similarly, 22.9% of male respondents rated it as entirely relevant, with 68.6% considering it mostly relevant. Despite this overall agreement, differing content priorities were noted. Female participants emphasized the importance of information related to screening intervals and early detection, whereas male participants highlighted the usefulness of content concerning HPV vaccination and its role in cancer prevention.

This distinction is particularly significant given the historically lower engagement of men in HPV-related public health initiatives. The findings suggest that gender-sensitive customization of health communication may enhance the relevance and impact of digital health tools such as ScreenGPT, thereby supporting more equitable health engagement across populations (Figure 12.).

4.5. Results of the system’s response evaluation by medical doctors

For each evaluation criterion—Accuracy, Clarity, and Usefulness—the mean and standard deviation of the scores were calculated across all items. Both medical doctor 1. (MD1) and medical doctor 2. (MD2) yielded high mean scores (all exceeding 3.5), with particularly low standard deviations observed in the Clarity and Usefulness dimensions.

This low variability suggests a high level of agreement among raters. In several instances, the standard deviation was minimal or even zero, rendering the computation of a reliable correlation coefficient infeasible. As a result, the Wilcoxon Signed-Rank Test was utilized to evaluate whether the median differences between paired ratings from MD1 and MD2 were statistically distinguishable from zero.

For the Accuracy dimension, the p-value was 0.059, which approaches but does not meet the conventional threshold for statistical significance ($p < 0.05$), suggesting that there is insufficient evidence to assert a meaningful difference between MD1 and MD2. Similarly, non-significant differences were observed for Clarity ($p = 0.063$) and Usefulness ($p = 0.317$). These results indicate that the ratings provided by MD1 and MD2 do not significantly differ across any of the evaluation dimensions.

This convergence in scoring supports the conclusion that the evaluations are consistent and comparable, demonstrating a high degree of reliability in the assessment process across independent raters (Table 7.).

	MD1		MD2		p-values
	Mean	SD	Mean	SD	
Accuracy	3.83	0.38	3.57	0.50	0.059
Clarity	3.80	0.55	4.00	0	0.063
Usefulness	3.97	0.18	4	0	0.317

Table 7. Wilcoxon Signed-Rank Test Results for Independent Evaluations by Medical Doctors on Accuracy, Clarity, and Usefulness

5. DISCUSSION

As outlined in the objectives, this study aimed to develop and evaluate a custom GPT-based web application specialized in cervical cancer, with the goal of enhancing patient education and supporting public health screening efforts. The objectives were threefold: to design a model capable of delivering reliable, publicly accessible information through a user-friendly interface by applying prompt engineering techniques; to assess usability through a structured user experience questionnaire; and to evaluate the accuracy and appropriateness of the system's responses through independent review by two medical doctors.

The central hypothesis was that a custom GPT model, specialized for a specific topic, could enhance the accessibility and comprehensibility of guidelines, thereby supporting improved patient education and facilitating user engagement with personalized, guideline-based health information.

Our evaluation encompassed technical validation, structured beta testing with potential users, and external review by two independent medical doctors. Collectively, these assessments highlight both the promise and the limitations of the current prototype.

5.1. ScreenGPT web application development

The first step of this research involved the successful development of ScreenGPT, a custom GPT-based web application tailored specifically for cervical cancer screening and prevention. The system was built on the GPT-4 and GPT-4o models and deployed through Streamlit. It was designed to deliver reliable, publicly accessible information in an intuitive and user-friendly format. To achieve this, prompt engineering techniques were employed to optimize the accuracy, clarity, and contextual relevance of the model's responses.

The application was developed not merely as a generic conversational AI, but as a specialized tool to enhance patient education, support comprehension of cervical cancer screening guidelines, and contribute to broader public health efforts.

5.2. User experience questionnaire

Following its technical implementation, ScreenGPT was subjected to preliminary validation through structured beta testing with potential users. Findings from the user experience questionnaire indicated that participants across genders perceived the information provided by the system as both relevant and informative.

Statistical analysis revealed a significant association between age groups and the perceived personalization of system responses, as well as between gender and perceptions of customization. The significant association between age and the perceived personalization of advice may be partly explained by younger individual's greater familiarity with digital environments and AI-mediated communication, which likely enhances their ability to recognize and appreciate personalized system responses. The significant association between gender and perceived system customization may be attributed to the system's design, which guides female users through a more detailed and branched question flow. This allows the system to collect richer contextual information and generate more tailored responses, thereby enhancing the perception of personalization compared to the more streamlined interaction pathway provided for male users.

While a trend was observed between age and the intention to reuse the system in the future, the corresponding p-value did not reach the threshold for statistical significance. No other associations among the remaining variables were statistically significant. Notably, younger participants reported higher levels of engagement, describing the interface as more interactive and accessible relative to the experience reported by users over the age of 40. Although this age-related difference did not attain statistical significance, it nonetheless underscores the importance of incorporating age-sensitive design considerations in the development of digital health technologies.

The sample population primarily comprised individuals aged 25–54, which is congruent with the WHO most recent guidance on cervical cancer prevention, emphasizing routine screening for women aged 30–49 and continuing relevance up to age 65. While screening remains pertinent for older age groups, the development of a digital educational tool is particularly justified for the 25–49 demographic, which constituted the majority of the study sample, as they are typically more receptive to digital health tools.

This demographic alignment also informed the decision to focus the GPT model on cervical cancer screening, as opposed to other preventive measures, which are generally targeted toward older populations. It is important to acknowledge that the study was conducted in Hungary. Consequently, the generalizability of the findings to other populations, cultural contexts, and healthcare infrastructures warrants further investigation in subsequent research.

5.3. ScreenGPT's answer evaluation by two medical doctors

An important aspect of this study was the independent evaluation of ScreenGPT's responses by two medical doctors. The findings showed that the ratings provided by MD1 and MD2 did not significantly differ across any of the evaluation dimensions. This convergence in scoring indicates that both evaluators similarly interpreted the system's outputs, which suggests that ScreenGPT's responses were sufficiently clear, consistent, and aligned with established screening guidelines to allow for comparable judgments.

Furthermore, the consistency of ratings suggests that ScreenGPT could potentially serve as a reliable tool for patient education and public health communication, since its outputs withstand scrutiny across different expert perspectives. While broader validation involving a larger pool of medical professionals is necessary to confirm generalizability, the strong agreement between MD1 and MD2 provides an encouraging indication of the system's potential for trustworthy deployment in preventive health contexts. The broader implications of this project extend to the role of AI in public health education and behavioral support.

By leveraging conversational AI, tools like ScreenGPT can complement existing cervical cancer screening programs, potentially reducing barriers such as misinformation, stigma, and lack of access to timely health education. Future development should focus on integrating multilingual support, embedding evidence-based behavioral nudges, and exploring interoperability with electronic health records (EHRs) or national screening registries.

5.4. Comparison with other AI-based tools

As per our knowledge, while various AI-based tools have been developed to support cervical cancer screening, the majority of these have predominantly focused on image-based diagnostic applications, such as cytology and colposcopy image analysis [111-114]. In contrast, our work introduces a novel LLM-driven application specifically designed as a patient education tool. Its primary aim is to disseminate accurate, guideline-based information on cervical cancer screening in a publicly accessible, user-centered format. This approach seeks to address a critical gap in healthcare delivery, particularly the persistent issue of inadequate patient awareness, which contributes to the underutilization of life-saving screening services.

In comparison with other health-oriented chatbot tools, existing literature generally supports the effectiveness of AI-driven conversational agents in promoting health behavior change across diverse populations. This finding is consistent with the results observed during our work. Prior research has demonstrated that LLMs are capable of generating accurate, readable, and contextually appropriate responses to user inquiries, making them effective tools for enhancing health literacy and facilitating patient-provider communication. Scoping reviews have further highlighted the utility of LLMs in simplifying complex medical content, generating patient education materials, and improving information accessibility, particularly for populations with limited health literacy [115-117]. Specifically in the context of cervical cancer screening, recent studies suggest that LLM-generated responses tend to align closely with established clinical guidelines, underscoring their potential to support patient understanding of screening procedures, associated risks, and follow-up protocols [118, 119].

However, the current body of evidence remains limited, and there is a pressing need for more rigorous methodological approaches, particularly randomized controlled trials, to substantiate these preliminary findings and establish the clinical utility of LLM-based tools with greater confidence [120, 121].

While most studies report generally positive or mixed outcomes regarding the usability, effectiveness, and user satisfaction associated with conversational agents, qualitative user feedback remains variable, reflecting the diversity of user expectations and needs. This observation echoed in our findings. Furthermore, the existing literature often reveals methodological limitations, including small sample sizes, inadequate

controls, and inconsistent reporting standards. These shortcomings highlight the necessity for improved study design and transparent reporting practices to enable more accurate assessments of LLM-based agents in healthcare contexts and to guide iterative system improvements.

5.5. Limitations

Nevertheless, several limitations emerged during the evaluation phase. While the model performed well in delivering general preventive information, its responses occasionally lacked the specificity required for complex clinical scenarios. This underscores the importance of framing ScreenGPT as an adjunct tool rather than a replacement for professional medical advice. Furthermore, despite the engagement of independent physicians in the validation process, broader clinical trials and large-scale usability studies are necessary to establish reliability, trustworthiness, and safety across diverse populations. Issues of ethical deployment, including data privacy, potential misinformation, and the boundaries of AI authority in health communication, also require ongoing scrutiny.

A key limitation of ScreenGPT, as with other applications built on large language models, is the risk of AI hallucination. This refers to instances where the model generates information that is plausible-sounding but factually incorrect or unsupported by evidence. While prompt engineering and expert validation helped to reduce such occurrences, they cannot be fully eliminated. In the context of cervical cancer screening, even minor inaccuracies, such as misstating eligibility criteria, misrepresenting test procedures, or oversimplifying risk factors, may have significant implications for patient understanding and decision-making. Therefore, continuous monitoring, systematic updates in line with evolving clinical guidelines, and clear disclaimers emphasizing that ScreenGPT is not a substitute for professional medical advice are essential to mitigate the potential harms of hallucinated outputs.

One notable limitation of this study is the use of a custom, non-validated questionnaire to assess user experience. The primary motivation for developing a custom questionnaire was to ensure that the questionnaire items were specifically tailored to the unique context and objectives of ScreenGPT, however, a validated questionnaire would significantly strengthen the reliability and credibility of the findings. Employing an

established, validated questionnaire would have enhanced the evaluation's robustness and provided stronger empirical support for the study's conclusions.

Overall, ScreenGPT demonstrates the potential of LLM-powered applications to enhance cervical cancer prevention efforts at scale. While the current prototype provides a proof of concept validated through technical testing, user feedback, and expert review, further research is required to rigorously assess its clinical utility, ethical robustness, and long-term public health impact.

Future research should expand to include comprehensive evaluations of cost-effectiveness, data privacy, and security, areas that remain underexplored but are essential for the broader adoption of such technologies [122]. As the integration of LLMs into clinical and public health practice accelerates, ethical, regulatory, and patient safety considerations will assume increasing importance and must be proactively addressed to ensure responsible and equitable implementation [123].

5.6. Relevance and Future Utility

Hungary has maintained an organized cervical cancer screening program since 2003, inviting women aged 25–65 for cytology-based testing every three years [124]. Despite this, participation remains persistently low, with recent figures indicating that fewer than one-third of eligible women attend organized screening, even though private healthcare partly compensates for this shortfall [125, 126]. These trends reveal structural and informational barriers that limit the program's effectiveness, including low health literacy, uncertainty about screening recommendations, and stigma associated with gynecological examinations.

In this context, ScreenGPT offers a timely and innovative opportunity to strengthen public health communication and outreach. By combining the contextual reasoning power of large language models with a user-friendly web interface, the tool can make evidence-based screening information both accessible and comprehensible to a wide audience. Its conversational design enables personalized explanations that adapt to users' literacy levels, thereby reducing misunderstandings that often discourage participation. Moreover, ScreenGPT can act as a trusted first point of contact for individuals hesitant to seek professional advice, offering clear guidance on eligibility, procedures, and the importance of timely screening.

Beyond individual education, ScreenGPT could be deployed at scale as a complementary public health instrument. Integrated into awareness campaigns or linked with primary care services, it could support general practitioners in counseling patients, help national programs reach underserved populations, and provide continuous, on-demand access to reliable information. ScreenGPT has the potential to improve adherence, reduce disparities, and ultimately contribute to lowering cervical cancer incidence and mortality. Furthermore, embedding ScreenGPT into national awareness campaigns or primary care workflows could extend its reach to underserved or rural populations who may otherwise face significant obstacles to accessing preventive services. Overall, ScreenGPT has the potential to function as more than an educational resource, it could become a scalable, complementary public health intervention. By reducing informational and psychological barriers, supporting informed decision-making, and encouraging adherence to national screening recommendations, ScreenGPT may help reverse declining participation rates in Hungary and ultimately contribute to reducing the burden of cervical cancer.

5.7. Summary of the main results of the dissertation

1. Demonstrated that a customized LLM can be applied to health prevention purposes via the Streamlit platform, and live test results confirmed this from users.
2. Demonstrated the system's ability to conduct empathetic and motivating dialogue contributes to user engagement.
3. Demonstrated that by providing personalized, interactive healthcare information, AI-based conversational agent systems have the potential to improve public health knowledge and promote preventive activities.
4. The research has shown that although AI does not replace healthcare professionals, it can be a valuable complement.
5. The developed system can serve as a model for other healthcare areas to support and improve existing screening and prevention efforts.

6. CONCLUSION

The development of ScreenGPT, an AI-driven web application designed to support and strengthen public health efforts in cervical cancer prevention, represents an important step toward the integration of LLMs into digital health interventions. Building upon the capabilities of GPT-4 and GPT-4o, the system was deployed through Streamlit, ensuring accessibility and ease of use for both healthcare providers and the general public. Our evaluation encompassed technical validation, structured beta testing with potential users, and external review by two independent medical doctors. Collectively, these assessments highlight both the promise and the limitations of the current prototype. A key contribution of this work lies in demonstrating the feasibility of adapting cutting-edge LLMs to a focused public health challenge. Unlike many generic AI health chatbots, ScreenGPT was purpose-built to promote cervical cancer awareness, risk assessment, and preventive behaviors.

From a technical standpoint, the system demonstrated stable functionality and consistent response generation across a range of test queries. The deployment on Streamlit proved to be a practical choice, offering rapid prototyping capabilities and straightforward accessibility without the need for complex installations. Importantly, the system's reliance on GPT-4 and GPT-4o models endowed ScreenGPT with contextual reasoning and natural language generation abilities that far exceed traditional rule-based or static FAQ-style systems.

This study demonstrated that a custom GPT model, tailored specifically to cervical cancer screening, can enhance the accessibility and comprehensibility of clinical guidelines, thereby supporting patient education and encouraging user engagement with personalized, guideline-based health information. Findings from the user experience questionnaire provided empirical support for this hypothesis, with participants across genders perceiving the information as relevant and informative. Age and gender were significantly associated with perceptions of personalization and customization, while younger participants, though not at a statistically significant level, tended to report greater interactivity and accessibility. These insights highlight the importance of age-sensitive design considerations in the development of digital health interventions.

The study sample, predominantly aged 25–54, closely aligns with the WHO-recommended target group for cervical cancer screening, further justifying the choice to

focus the system on this domain. By addressing a critical gap in patient awareness and health literacy, ScreenGPT distinguishes itself from prior AI-based approaches, which have largely concentrated on image-based diagnostics, by instead offering a patient-centered educational tool.

The findings are consistent with the broader literature, which supports the effectiveness of conversational AI agents in simplifying complex medical content, enhancing health literacy, and promoting health behavior change. Nonetheless, the evidence base remains limited, with many studies, like the present one, constrained by sample size, methodological variation, and limited generalizability. These challenges underscore the need for rigorous future research, particularly randomized controlled trials, to establish the clinical utility of LLM-based applications with greater confidence.

Custom GPTs, such as ScreenGPT, may serve as valuable entry points for enhancing awareness of cervical cancer screening and supporting women in preparing for informed decision-making. However, these tools should not be regarded as standalone or authoritative sources of information, as they may omit critical content or contextual nuances. Instead, ScreenGPT and similar applications are best positioned as adjuncts to, rather than replacements for, evidence-based guidance and counseling provided by qualified health care professionals.

Future evaluations should expand to encompass cost-effectiveness, privacy, security, and ethical considerations, which will be critical to the responsible integration of LLM-driven tools into healthcare systems. As such, technologies become increasingly embedded in both clinical practice and public health, ensuring equitable, transparent, and safe implementation will be essential for maximizing their potential to improve patient outcomes and reduce barriers to life-saving preventive care.

7. SUMMARY

As a summary, this study presents the successful development of a custom GPT-based model designed to support public health efforts in cervical cancer screening. The model was built by processing established medical guidelines and applying prompt engineering strategies using the OpenAI GPT-4 and GPT-4o within the Streamlit framework, utilizing Python for implementation. The system was deployed through the Streamlit platform and underwent a three-week beta testing phase involving 115 participants. Of these, 112 completed a user experience questionnaire, offering valuable insights into various aspects of system performance.

Furthermore, two independent medical practitioners assessed the system's responses to the 30 most frequently asked questions for clinical accuracy. The findings underscore the potential of AI-driven tools to enhance cervical cancer screening efforts and address healthcare workforce limitations, especially in resource-constrained environments. By providing tailored, interactive health information, such systems can improve public health literacy and foster greater engagement in preventive care. A key strength of the approach lies in its capacity to distill complex clinical information into accessible, actionable guidance for diverse user populations. Moreover, the model's ability to deliver empathetic and motivational dialogue, emulating interactions with a knowledgeable, supportive interlocutor, significantly contributes to user engagement and trust.

In line with existing literature, the results suggest that while AI is not a substitute for healthcare professionals, it serves as a valuable complement. These systems can optimize healthcare delivery by streamlining routine tasks and expediting access to reliable information, thereby enabling clinicians to dedicate more time to individualized, high-priority care.

Sustained user engagement is vital for the long-term success of digital health interventions. Future research should explore strategies to improve personalization, user-centered design, and the integration of interactive educational components aimed at encouraging continued usage and positive user referrals. Subsequent development phases can also aim to expand the model's conversational scope to encompass other critical preventive screenings, including those for breast and colorectal cancers.

8. REFERENCES

1. Li, J., et al., *ChatGPT in healthcare: A taxonomy and systematic review*. *Comput Methods Programs Biomed*, 2024. **245**: p. 108013. <https://doi.org/10.1016/j.cmpb.2024.108013>
2. Jurafsky, D., & Martin, J. H., *Speech and language processing (3rd ed. draft)*. 2021: Stanford University.
3. Manning, C.D., Schütze, H., & Schütze, H., *Foundations of statistical natural language processing*. MIT press., 1999.
4. Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. *Feature-rich part-of-speech tagging with a cyclic dependency network*. in *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*. 2003.
5. Kübler, S., McDonald, R., & Nivre, J. *Dependency parsing*. in *Morgan & Claypool*. 2009.
6. Tjong Kim Sang, E.F., & De Meulder, F. *Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition*. in *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*. 2003.
7. Vaswani, A., et al., *Attention Is All You Need*. 2017. <https://doi.org/10.48550/arXiv.1706.03762>
8. Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. , *SQuAD: 100,000+ questions for machine comprehension of text*. 2016, Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing: Association for Computational Linguistics.
9. Nenkova, A. and K. McKeown, *A Survey of Text Summarization Techniques*. 2012.
10. Kim, H., Hwang, H., Lee, J., Park, S., Kim, D., Lee, T., Yoon, C., Sohn, J., Park, J., Reykhart, O., Fetherston, T., Choi, D., Kwak, S.H., Chen, Q., Kang, J., *Small language models learn enhanced reasoning skills from medical textbooks*. *npj Digital Medicine* 8., 2025. <https://doi.org/> <https://doi.org/10.1038/s41746-025-01653-8>

11. Magnini, M., G. Aguzzi, and S. Montagna, *Open-source small language models for personal medical assistant chatbots*. *Intelligence-Based Medicine*, 2025. **11**: p. 100197. <https://doi.org/https://doi.org/10.1016/j.ibmed.2024.100197>
12. Griewing, S., et al., *Proof-of-concept study of a small language model chatbot for breast cancer decision support – a transparent, source-controlled, explainable and data-secure approach*. *Journal of Cancer Research and Clinical Oncology*, 2024. **150**(10). <https://doi.org/10.1007/s00432-024-05964-3>
13. Wang, F., et al., *A Comprehensive Survey of Small Language Models in the Era of Large Language Models: Techniques, Enhancements, Applications, Collaboration with LLMs, and Trustworthiness*. 2024.
14. B, M., *The Impact of Multimodal Large Language Models on Health Care's Future*. *J Med Internet Res* 2023;25:e52865, 2023. <https://doi.org/doi:10.2196/52865>
15. Xiao, H., et al., *A comprehensive survey of large language models and multimodal large language models in medicine*. *Information Fusion*, 2025. **117**: p. 102888. <https://doi.org/10.1016/j.inffus.2024.102888>
16. Biswas, S.S., *Role of Chat GPT in Public Health*. *Ann Biomed Eng*, 2023. **51**(5): p. 868-869. <https://doi.org/10.1007/s10439-023-03172-7>
17. Hu, S., et al., *Natural Language Processing Technologies for Public Health in Africa: Scoping Review*. *J Med Internet Res*, 2025. **27**: p. e68720. <https://doi.org/10.2196/68720>
18. Jerfy, A., O. Selden, and R. Balkrishnan, *The Growing Impact of Natural Language Processing in Healthcare and Public Health*. *Inquiry*, 2024. **61**: p. 469580241290095. <https://doi.org/10.1177/00469580241290095>
19. Cascella, M., et al., *Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios*. *J Med Syst*, 2023. **47**(1): p. 33. <https://doi.org/10.1007/s10916-023-01925-4>
20. Cheng, S.W., et al., *The now and future of ChatGPT and GPT in psychiatry*. *Psychiatry Clin Neurosci*, 2023. **77**(11): p. 592-596. <https://doi.org/10.1111/pcn.13588>
21. Bazzari, F.H. and A.H. Bazzari, *Utilizing ChatGPT in Telepharmacy*. *Cureus*, 2024. **16**(1): p. e52365. <https://doi.org/10.7759/cureus.52365>

22. Lecler, A., L. Duron, and P. Soyer, *Revolutionizing radiology with GPT-based models: Current applications, future possibilities and limitations of ChatGPT*. *Diagn Interv Imaging*, 2023. **104**(6): p. 269-274. <https://doi.org/10.1016/j.diii.2023.02.003>
23. Jeyaraman, M., et al., *ChatGPT in action: Harnessing artificial intelligence potential and addressing ethical challenges in medicine, education, and scientific research*. *World J Methodol*, 2023. **13**(4): p. 170-178. <https://doi.org/10.5662/wjm.v13.i4.170>
24. Mondal, H., et al., *ChatGPT in Answering Queries Related to Lifestyle-Related Diseases and Disorders*. *Cureus*, 2023. **15**(11): p. e48296. <https://doi.org/10.7759/cureus.48296>
25. Sun, H., et al., *An AI Dietitian for Type 2 Diabetes Mellitus Management Based on Large Language and Image Recognition Models: Preclinical Concept Validation Study*. *J Med Internet Res*, 2023. **25**: p. e51300. <https://doi.org/10.2196/51300>
26. Spallek, S., et al., *Can we use ChatGPT for Mental Health and Substance Use Education? Examining Its Quality and Potential Harms*. *JMIR Med Educ*, 2023. **9**: p. e51243. <https://doi.org/10.2196/51243>
27. Cakir, H., et al., *Evaluating the performance of ChatGPT in answering questions related to urolithiasis*. *Int Urol Nephrol*, 2023. <https://doi.org/10.1007/s11255-023-03773-0>
28. Zheng, Y., et al., *Enhancing Diabetes Self-management and Education: A Critical Analysis of ChatGPT's Role*. *Ann Biomed Eng*, 2023. <https://doi.org/10.1007/s10439-023-03317-8>
29. Rajjoub, R., et al., *ChatGPT and its Role in the Decision-Making for the Diagnosis and Treatment of Lumbar Spinal Stenosis: A Comparative Analysis and Narrative Review*. *Global Spine J*, 2023: p. 21925682231195783. <https://doi.org/10.1177/21925682231195783>
30. Withers, C.A., et al., *Natural language processing in drug discovery: bridging the gap between text and therapeutics with artificial intelligence*. *Expert Opin Drug Discov*, 2025. **20**(6): p. 765-783. <https://doi.org/10.1080/17460441.2025.2490835>

31. Eguia, H., et al., *Clinical Decision Support and Natural Language Processing in Medicine: Systematic Literature Review*. J Med Internet Res, 2024. **26**: p. e55315. <https://doi.org/10.2196/55315>
32. Roberts, L.J., et al., *Challenges for implementing generative artificial intelligence (GenAI) into clinical healthcare*. Intern Med J, 2025. **55**(7): p. 1063-1069. <https://doi.org/10.1111/imj.70035>
33. Mese, I., C.A. Taslicay, and A.K. Sivrioglu, *Improving radiology workflow using ChatGPT and artificial intelligence*. Clin Imaging, 2023. **103**: p. 109993. <https://doi.org/10.1016/j.clinimag.2023.109993>
34. Liu, J., C. Wang, and S. Liu, *Utility of ChatGPT in Clinical Practice*. J Med Internet Res, 2023. **25**: p. e48568. <https://doi.org/10.2196/48568>
35. Dave, T., S.A. Athaluri, and S. Singh, *ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations*. Front Artif Intell, 2023. **6**: p. 1169595. <https://doi.org/10.3389/frai.2023.1169595>
36. Au Yeung, J., et al., *Natural language processing data services for healthcare providers*. BMC Med Inform Decis Mak, 2024. **24**(1): p. 356. <https://doi.org/10.1186/s12911-024-02713-x>
37. Paganelli, A., et al., *Natural language processing in dermatology: A systematic literature review and state of the art*. J Eur Acad Dermatol Venereol, 2024. **38**(12): p. 2225-2234. <https://doi.org/10.1111/jdv.20286>
38. Scharp, D., et al., *Natural Language Processing Applied to Clinical Documentation in Post-acute Care Settings: A Scoping Review*. J Am Med Dir Assoc, 2024. **25**(1): p. 69-83. <https://doi.org/10.1016/j.jamda.2023.09.006>
39. Meskó, B., G. Hetényi, and Z. Györffy, *Will artificial intelligence solve the human resource crisis in healthcare?* BMC Health Services Research, 2018. **18**(1). <https://doi.org/10.1186/s12913-018-3359-4>
40. Liu, C.L., C.T. Ho, and T.C. Wu, *Custom GPTs Enhancing Performance and Evidence Compared with GPT-3.5, GPT-4, and GPT-4o? A Study on the Emergency Medicine Specialist Examination*. Healthcare (Basel), 2024. **12**(17). <https://doi.org/10.3390/healthcare12171726>

41. Fisher, A.D. and G. Fisher, *Evaluating performance of custom GPT in anesthesia practice*. J Clin Anesth, 2023. **93**: p. 111371. <https://doi.org/10.1016/j.jclinane.2023.111371>
42. Sathe, T.S., et al., *How I GPT It: Development of Custom Artificial Intelligence (AI) Chatbots for Surgical Education*. J Surg Educ, 2024. **81**(6): p. 772-775. <https://doi.org/10.1016/j.jsurg.2024.03.004>
43. Darkhabani, M., et al., *ChatGPT and autoimmunity - A new weapon in the battlefield of knowledge*. Autoimmun Rev, 2023. **22**(8): p. 103360. <https://doi.org/10.1016/j.autrev.2023.103360>
44. Haver, H.L., et al., *Appropriateness of Breast Cancer Prevention and Screening Recommendations Provided by ChatGPT*. Radiology, 2023. **307**(4): p. e230424. <https://doi.org/10.1148/radiol.230424>
45. Rao, A., et al., *Evaluating ChatGPT as an Adjunct for Radiologic Decision-Making*. 2023, Cold Spring Harbor Laboratory.
46. Kambire, J., S.G. Barro, and P. Staccini, *Designing a Custom GPT Model for Medical Dialogue Simulation*. Stud Health Technol Inform, 2025. **328**: p. 11-15. <https://doi.org/10.3233/shti250663>
47. Alsabawi, Y., P.R. Quesada, and D.T. Rouse, *Readability of custom chatbot vs. GPT-4 responses to otolaryngology-related patient questions*. Am J Otolaryngol, 2025. **46**(5): p. 104717. <https://doi.org/10.1016/j.amjoto.2025.104717>
48. Holderried, F., et al., *A Generative Pretrained Transformer (GPT)-Powered Chatbot as a Simulated Patient to Practice History Taking: Prospective, Mixed Methods Study*. JMIR Med Educ, 2024. **10**: p. e53961. <https://doi.org/10.2196/53961>
49. Aykut, A. and A.S. Sezenoz, *Exploring the Potential of Code-Free Custom GPTs in Ophthalmology: An Early Analysis of GPT Store and User-Creator Guidance*. Ophthalmol Ther, 2024. **13**(10): p. 2697-2713. <https://doi.org/10.1007/s40123-024-01014-w>
50. Shalong, W., et al., *Enhancing self-directed learning with custom GPT AI facilitation among medical students: A randomized controlled trial*. Med Teach, 2025. **47**(7): p. 1126-1133. <https://doi.org/10.1080/0142159x.2024.2413023>

51. Pu, J., et al., *Accuracy, satisfaction, and impact of custom GPT in acquiring clinical knowledge: Potential for AI-assisted medical education*. *Med Teach*, 2025: p. 1-7. <https://doi.org/10.1080/0142159x.2025.2458808>
52. Woo, J.J., et al., *Custom Large Language Models Improve Accuracy: Comparing Retrieval Augmented Generation and Artificial Intelligence Agents to Noncustom Models for Evidence-Based Medicine*. *Arthroscopy*, 2025. **41**(3): p. 565-573.e6. <https://doi.org/10.1016/j.arthro.2024.10.042>
53. Chiang, K.L., et al., *Customized GPT model largely increases surgery decision accuracy for pharmaco-resistant epilepsy*. *J Clin Neurosci*, 2024. **130**: p. 110918. <https://doi.org/10.1016/j.jocn.2024.110918>
54. Mehta, N., et al., *Creating custom GPTs for faculty development: An example using the Johari Window and Crucial Conversation frameworks for providing feedback to struggling students*. *Med Teach*, 2025. **47**(8): p. 1266-1268. <https://doi.org/10.1080/0142159x.2024.2445043>
55. Young, C.C., et al., *Diagnostic Accuracy of a Custom Large Language Model on Rare Pediatric Disease Case Reports*. *Am J Med Genet A*, 2025. **197**(2): p. e63878. <https://doi.org/10.1002/ajmg.a.63878>
56. Collins, B.R., E.W. Black, and K.E. Rarey, *Introducing AnatomyGPT: A customized artificial intelligence application for anatomical sciences education*. *Clin Anat*, 2024. **37**(6): p. 661-669. <https://doi.org/10.1002/ca.24178>
57. Kiyak, Y.S. and A.A. Kononowicz, *Case-based MCQ generator: A custom ChatGPT based on published prompts in the literature for automatic item generation*. *Med Teach*, 2024. **46**(8): p. 1018-1020. <https://doi.org/10.1080/0142159x.2024.2314723>
58. Miao, J., et al., *How to improve ChatGPT performance for nephrologists: a technique guide*. *J Nephrol*, 2024. **37**(5): p. 1397-1403. <https://doi.org/10.1007/s40620-024-01974-z>
59. Bicknell, B.T., et al., *Domain-Specific Customization for Language Models in Otolaryngology: The ENT GPT Assistant*. *OTO Open*, 2025. **9**(2): p. e70125. <https://doi.org/10.1002/oto2.70125>
60. Daungsupawong, H. and V. Wiwanitkit, *Comment on "Using a customized GPT to provide guideline-based recommendations for management of pancreatic*

- cystic lesions*". *Endosc Int Open*, 2024. **12**(6): p. E810. <https://doi.org/10.1055/a-2335-8369>
61. Sengul, I. and D. Sengul, *Exegesis on using a customized GPT to provide guideline-based recommendations for the management of pancreatic cystic lesions*. *Endosc Int Open*, 2025. **13**: p. a26051215. <https://doi.org/10.1055/a-2605-1215>
 62. Cevik, A.A. and F.M. Abu-Zidan, *Utilizing AI-Powered Thematic Analysis: Methodology, Implementation, and Lessons Learned*. *Cureus*, 2025. **17**(6): p. e85338. <https://doi.org/10.7759/cureus.85338>
 63. Gorelik, Y., *Reply to: Exegesis on using a customized GPT to provide guideline-based recommendations for management of pancreatic cystic lesions*. *Endosc Int Open*, 2025. **13**: p. a26051278. <https://doi.org/10.1055/a-2605-1278>
 64. McMurry, A.J., et al., *Large Language Model Symptom Identification From Clinical Text: Multicenter Study*. *J Med Internet Res*, 2025. **27**: p. e72984. <https://doi.org/10.2196/72984>
 65. Gong, E.J., et al., *The Potential Clinical Utility of the Customized Large Language Model in Gastroenterology: A Pilot Study*. *Bioengineering (Basel)*, 2024. **12**(1). <https://doi.org/10.3390/bioengineering12010001>
 66. Flynn, J.C., et al., *The Performance of a Customized Generative Pre-trained Transformer on the American Society for Surgery of the Hand Self-Assessment Examination*. *Cureus*, 2024. **16**(9): p. e70205. <https://doi.org/10.7759/cureus.70205>
 67. Owens, O.L. and M.S. Leonard, *Evaluating an AI Chatbot "Prostate Cancer Info" for Providing Quality Prostate Cancer Screening Information: Cross-Sectional Study*. *JMIR Cancer*, 2025. **11**: p. e72522. <https://doi.org/10.2196/72522>
 68. Schumacher, I., et al., *Performance analysis of an emergency triage system in ophthalmology using a customized CHATBOT*. *Digit Health*, 2025. **11**: p. 20552076251320298. <https://doi.org/10.1177/20552076251320298>
 69. Meskó, B., *Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial*. *J Med Internet Res*, 2023. **25**: p. e50638. <https://doi.org/10.2196/50638>

70. Zaghir, J., et al., *Prompt Engineering Paradigms for Medical Applications: Scoping Review*. J Med Internet Res, 2024. **26**: p. e60501. <https://doi.org/10.2196/60501>
71. Shrestha, P., et al., *Post-translational modification prediction via prompt-based fine-tuning of a GPT-2 model*. Nat Commun, 2024. **15**(1): p. 6699. <https://doi.org/10.1038/s41467-024-51071-9>
72. Majdik, Z.P., et al., *Sample Size Considerations for Fine-Tuning Large Language Models for Named Entity Recognition Tasks: Methodological Study*. Jmir ai, 2024. **3**: p. e52095. <https://doi.org/10.2196/52095>
73. Liu, P., et al., *Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing*. ACM Computing Surveys, 2022. **55**. <https://doi.org/10.1145/3560815>
74. Giray, L., *Prompt Engineering with ChatGPT: A Guide for Academic Writers*. Ann Biomed Eng, 2023. **51**(12): p. 2629-2633. <https://doi.org/10.1007/s10439-023-03272-4>
75. Larbi, D., K. Denecke, and E. Gabarron, *Usability Testing of a Social Media Chatbot for Increasing Physical Activity Behavior*. Journal of Personalized Medicine, 2022. **12**: p. 828. <https://doi.org/10.3390/jpm12050828>
76. Singhal, K., et al., *Large language models encode clinical knowledge*. Nature, 2023. **620**(7972): p. 172-180. <https://doi.org/10.1038/s41586-023-06291-2>
77. Li, Y., *A PRACTICAL SURVEY ON ZERO-SHOT PROMPT DESIGN FOR IN-CONTEXT LEARNING*. 2023.
78. Brown, T., et al., *Language Models are Few-Shot Learners*. 2020.
79. Attal, K., et al., *Using Deep Learning With Few-Shot Learning to Improve Data Capture in Total Hip Arthroplasty Operative Notes*. J Arthroplasty, 2025. <https://doi.org/10.1016/j.arth.2025.06.008>
80. Wei, J., *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. 2023.
81. Tassone, D.M., et al., *Evaluating chain-of-thought prompting in a GPT chatbot for BCID2 interpretation and stewardship: how does AI compare to human experts?* Antimicrob Steward Healthc Epidemiol, 2025. **5**(1): p. e154. <https://doi.org/10.1017/ash.2025.10059>

82. Zhou, D., et al., *Least-to-Most Prompting Enables Complex Reasoning in Large Language Models*. 2022.
83. Fontham, E.T.H., et al., *Cervical cancer screening for individuals at average risk: 2020 guideline update from the American Cancer Society*. *CA Cancer J Clin*, 2020. **70**(5): p. 321-346. <https://doi.org/10.3322/caac.21628>
84. *WHO Guidelines Approved by the Guidelines Review Committee*, in *WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention: Use of mRNA tests for human papillomavirus (HPV)*. 2021, World Health Organization
85. Sultanov, M., et al., *Investigating feasibility of 2021 WHO protocol for cervical cancer screening in underscreened populations: PREvention and SCReening Innovation Project Toward Elimination of Cervical Cancer (PRESCRIP-TEC)*. *BMC Public Health*, 2022. **22**(1). <https://doi.org/10.1186/s12889-022-13488-z>
86. *WHO Guidelines Approved by the Guidelines Review Committee*, in *WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention: Use of dual-stain cytology to triage women after a positive test for human papillomavirus (HPV)*. 2024, World Health Organization
87. Gavinski, K. and D. DiNardo, *Cervical Cancer Screening*. *Med Clin North Am*, 2023. **107**(2): p. 259-269. <https://doi.org/10.1016/j.mcna.2022.10.006>
88. Voelker, R.A., *Cervical Cancer Screening*. *Jama*, 2023. **330**(20): p. 2030. <https://doi.org/10.1001/jama.2023.21987>
89. Tao, L., et al., *Prevalence and risk factors for cervical neoplasia: a cervical cancer screening program in Beijing*. *BMC Public Health*, 2014. **14**(1): p. 1185. <https://doi.org/10.1186/1471-2458-14-1185>
90. Varatharajan, S., et al., *Cervical cancer in Malaysia: can we improve our screening and preventive practice?* *BMC Public Health*, 2012. **12**(S2): p. A17. <https://doi.org/10.1186/1471-2458-12-s2-a17>
91. Vidrine, J.I., et al., *Enhancing long-term smoking abstinence among individuals with a history of cervical intraepithelial neoplasia or cervical cancer (Project ACCESS): protocol for a randomized clinical trial*. *BMC Public Health*, 2023. **23**(1). <https://doi.org/10.1186/s12889-023-16189-3>

92. Popalis, M.L., et al., *Improving cervical cancer screening rates: a scoping review of resources and interventions*. *Cancer Causes Control*, 2022. **33**(11): p. 1325-1333. <https://doi.org/10.1007/s10552-022-01618-2>
93. Tao, Y., et al., *Cultural bias and cultural alignment of large language models*. *PNAS Nexus*, 2024. **3**(9): p. pgae346. <https://doi.org/10.1093/pnasnexus/pgae346>
94. Hall, B., K. Howard, and K. McCaffery, *Do cervical cancer screening patient information leaflets meet the HPV information needs of women?* *Patient Educ Couns*, 2008. **72**(1): p. 78-87. <https://doi.org/10.1016/j.pec.2008.01.024>
95. Mohamed Amin, S., et al., *Unlocking prevention: the role of health literacy in cervical cancer screening: community nursing perspective*. *BMC Nursing*, 2025. **24**(1). <https://doi.org/10.1186/s12912-025-02797-4>
96. Musa, J., et al., *Effect of cervical cancer education and provider recommendation for screening on screening rates: A systematic review and meta-analysis*. *PLOS ONE*, 2017. **12**(9): p. e0183924. <https://doi.org/10.1371/journal.pone.0183924>
97. Thiel De Bocanegra, H., et al., *Impact of an educational tool on young women's knowledge of cervical cancer screening recommendations*. *Cancer Causes & Control*, 2022. **33**(6): p. 813-821. <https://doi.org/10.1007/s10552-022-01569-8>
98. Lee, Y.-C., et al. *Exploring Effects of Chatbot-based Social Contact on Reducing Mental Illness Stigma*. *ACM*.
99. Miles, O., R. West, and T. Nadarzynski, *Health chatbots acceptability moderated by perceived stigma and severity: A cross-sectional survey*. *DIGITAL HEALTH*, 2021. **7**: p. 205520762110630. <https://doi.org/10.1177/20552076211063012>
100. Rains, S.A., *The implications of stigma and anonymity for self-disclosure in health blogs*. *Health Commun*, 2014. **29**(1): p. 23-31. <https://doi.org/10.1080/10410236.2012.714861>
101. Chen, P., et al., *The acceptability and effectiveness of artificial intelligence-based chatbot for hypertensive patients in community: protocol for a mixed-methods study*. *BMC Public Health*, 2024. **24**(1). <https://doi.org/10.1186/s12889-024-19667-4>
102. Morgenstern, J.D., et al., *"AI's gonna have an impact on everything in society, so it has to have an impact on public health": a fundamental qualitative descriptive*

- study of the implications of artificial intelligence for public health*. BMC Public Health, 2021. **21**(1). <https://doi.org/10.1186/s12889-020-10030-x>
103. Fisher, S. and L.C. Rosella, *Priorities for successful use of artificial intelligence by public health organizations: a literature review*. BMC Public Health, 2022. **22**(1). <https://doi.org/10.1186/s12889-022-14422-z>
 104. Bhatla, N. and S. Singhal, *Primary HPV screening for cervical cancer*. Best Pract Res Clin Obstet Gynaecol, 2020. **65**: p. 98-108. <https://doi.org/10.1016/j.bpobgyn.2020.02.008>
 105. Nishimura, Y., et al., *Mailing human papillomavirus self-sampling kits to women under-screened for cervical cancer improved detection in cervical cancer screening in a general population study in Japan*. BMC Public Health, 2023. **23**(1). <https://doi.org/10.1186/s12889-023-15402-7>
 106. Hong, Y., et al., *HPV and cervical cancer related knowledge, awareness and testing behaviors in a community sample of female sex workers in China*. BMC Public Health, 2013. **13**(1): p. 696. <https://doi.org/10.1186/1471-2458-13-696>
 107. Scott Duncan, T., et al., *Empowered patients and informal care-givers as partners?-a survey study of healthcare professionals' perceptions*. BMC Health Serv Res, 2023. **23**(1): p. 404. <https://doi.org/10.1186/s12913-023-09386-8>
 108. Girasek, E., et al., *E-páciensek Magyarországon: Digitális egészséggel kapcsolatos ismeretek, szokások egy országos reprezentatív felmérés tükrében*. Orvosi Hetilap, 2022. **163**(29): p. 1159-1165. <https://doi.org/10.1556/650.2022.32512>
 109. BÔSZE PÉTER DR., G.P.D., HERNÁDI ZOLTÁN DR., PAP KÁROLY DR., UNGÁR LÁSZLÓ DR, *A méhnyakrák szűrésének szempontjai: hazai irányelvek*. Nőgyógyászati Onkológia, 2009. **14**: p. 11-17.
 110. OpenAI, *OpenAI API Reference*. 2024.
 111. Viñals, R., et al., *Artificial Intelligence-Based Cervical Cancer Screening on Images Taken during Visual Inspection with Acetic Acid: A Systematic Review*. Diagnostics (Basel), 2023. **13**(5). <https://doi.org/10.3390/diagnostics13050836>
 112. Kanavati, F., et al., *A Deep Learning Model for Cervical Cancer Screening on Liquid-Based Cytology Specimens in Whole Slide Images*. Cancers (Basel), 2022. **14**(5). <https://doi.org/10.3390/cancers14051159>

113. Mathivanan, S.K., et al., *Enhancing cervical cancer detection and robust classification through a fusion of deep learning models*. *Sci Rep*, 2024. **14**(1): p. 10812. <https://doi.org/10.1038/s41598-024-61063-w>
114. Hou, X., et al., *Artificial Intelligence in Cervical Cancer Screening and Diagnosis*. *Front Oncol*, 2022. **12**: p. 851367. <https://doi.org/10.3389/fonc.2022.851367>
115. AlSammarrarie, A. and M. Househ, *The Use of Large Language Models in Generating Patient Education Materials: a Scoping Review*. *Acta Inform Med*, 2025. **33**(1): p. 4-10. <https://doi.org/10.5455/aim.2024.33.4-10>
116. Aydin, S., et al., *Large language models in patient education: a scoping review of applications in medicine*. *Front Med (Lausanne)*, 2024. **11**: p. 1477898. <https://doi.org/10.3389/fmed.2024.1477898>
117. Lin, C. and C.F. Kuo, *Roles and Potential of Large Language Models in Healthcare: A Comprehensive Review*. *Biomed J*, 2025: p. 100868. <https://doi.org/10.1016/j.bj.2025.100868>
118. Reicher, L., et al., *Exploring the role of artificial intelligence, large language models: Comparing patient-focused information and clinical decision support capabilities to the gynecologic oncology guidelines*. *Int J Gynaecol Obstet*, 2025. **168**(2): p. 419-427. <https://doi.org/10.1002/ijgo.15869>
119. Kuerbanjiang, W., et al., *Performance Evaluation of Large Language Models in Cervical Cancer Management Based on a Standardized Questionnaire: Comparative Study*. *J Med Internet Res*, 2025. **27**: p. e63626. <https://doi.org/10.2196/63626>
120. Aggarwal, A., et al., *Artificial Intelligence-Based Chatbots for Promoting Health Behavioral Changes: Systematic Review*. *J Med Internet Res*, 2023. **25**: p. e40789. <https://doi.org/10.2196/40789>
121. Xu, L., et al., *Chatbot for Health Care and Oncology Applications Using Artificial Intelligence and Machine Learning: Systematic Review*. *JMIR Cancer*, 2021. **7**(4): p. e27850. <https://doi.org/10.2196/27850>
122. Milne-Ives, M., et al., *The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review*. *J Med Internet Res*, 2020. **22**(10): p. e20346. <https://doi.org/10.2196/20346>

123. Huo, B., et al., *Large Language Models for Chatbot Health Advice Studies: A Systematic Review*. JAMA Netw Open, 2025. **8**(2): p. e2457879. <https://doi.org/10.1001/jamanetworkopen.2024.57879>
124. Boncz, I., et al., [*The coverage of cervical screening in Hungary*]. Orv Hetil, 2007. **148**(46): p. 2177-82. <https://doi.org/10.1556/oh.2007.28956>
125. Gyulai, A., et al., *General practitioners can increase participation in cervical cancer screening - a model program in Hungary*. BMC Fam Pract, 2018. **19**(1): p. 67. <https://doi.org/10.1186/s12875-018-0755-0>
126. Vajda, R., K. Juhász, and D. Endrei, *Clinical and health policy experiences with HPV prevalence within the Hungarian organized cervical cancer screening program*. J Gynecol Oncol, 2018. **29**(3): p. e45. <https://doi.org/10.3802/jgo.2018.29.e45>

9. BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS

9.1. Publications/abstracts related to PhD dissertation

Papers

Angyal Viola, Bertalan Ádám, Domján Péter, Feith Helga Judit, Dinya Elek
Development of a Questionnaire for Assessing the Use of ChatGPT in Primary and Secondary

Disease Prevention

FRONTIERS IN PUBLIC HEALTH 13 Paper: 1709611, 10 p. (2026)

Közlemény: 36831463 | Szakcikk (Folyóiratcikk) | Tudományos

Scopus - Public Health, Environmental and Occupational Health SJR indikátor: Q1

IF: 3,2

Angyal Viola, Bertalan Ádám, Domján Péter, Dinya Elek

Exploring the possibilities and limitations of customized large language model to support and

improve cervical cancer screening

BMC MEDICAL INFORMATICS AND DECISION MAKING 25: 1 Paper: 242, 16 p. (2025)

Közlemény: 36226867 | Szakcikk (Folyóiratcikk) | Tudományos

Scopus - Computer Science Applications SJR indikátor: Q1

Scopus - Health Policy SJR indikátor: Q1

Scopus - Health Informatics SJR indikátor: Q2

IF: 3,8

Bertalan Ádám, **Angyal Viola**, Domján Péter, Sædder Eva Aggerholm, Király Gyula, Erdélyi Lóránd, Gyimesi Nóra, Dinya Elek

Development of an Artificial Intelligence Powered Medication Risk Score Calculator Application

BASIC & CLINICAL PHARMACOLOGY & TOXICOLOGY 137: 4 Paper: e70109, 15 p. (2025)

Közlemény: 36346440 | Szakcikk (Folyóiratcikk) | Tudományos

Scopus - Medicine (miscellaneous) SJR indikátor: Q2

Scopus - Pharmacology SJR indikátor: Q2

Scopus - Toxicology SJR indikátor: Q2

IF: 3,3

Angyal Viola, Bertalan Ádám, Dinya Elek

A ScreenGPT rendszer megbízhatóságának vizsgálata méhnyakrák prevencióval kapcsolatos információátadás támogatásában

In: V. Életmóddorvosi és II. Mozgás Receptre Konferencia Magyar Életmód

Orvostani Társaság (2025) 102 p. pp. 6-7.

Közlemény: 36381127 | Absztrakt | Kivonat (Egyéb konferenciaközlemény) |

Tudományos

Kőrös-Angyal Viola, Bertalan Ádám, Domján Péter, Dinya Elek
CN10 Evaluating prompt engineering techniques for artificial intelligence-generated cervical cancer screening information: A comparative content analysis
ANNALS OF ONCOLOGY 36: Suppl. 2 pp. S1518-S1519. (2025)
Közlemény: 36704634 | Absztrakt | Kivonat (Folyóiratcikk) | Tudományos

Angyal Viola, Bertalan Ádám, Domján Péter, Dinya Elek
ScreenGPT – A mesterséges intelligencia alkalmazásának lehetőségei és korlátai a primer, szekunder és terciér prevencióban [ScreenGPT-The opportunities and limitations of artificial intelligence in primary, secondary and tertiary prevention]
ORVOSI HETILAP 165: 16 pp. 629-635. (2024)
Közlemény: 34817394 | Szakcikk (Folyóiratcikk) | Tudományos
Scopus - Medicine (miscellaneous) SJR indikátor: Q4 IF: 0,9

Angyal Viola, Dinya Elek
Mesterséges intelligencia alapú GPT-modellek felhasználásának lehetőségei és korlátai egészségügyi alkalmazások fejlesztésében
IME 23: KSZ 1. pp. 23-29. (2024)
Közlemény: 34854971 | Szakcikk (Folyóiratcikk) | Tudományos

Summarized IF: 11,2

Conference abstracts

2025 – Semmelweis Symposium 2025

2025 - European Society for Medical Oncology

2025 – XIX. Hungarian Health Economics Society

2025 - V. Hungarian Lifestyle Medicine Conference

2025 - Semmelweis PhD Scientific Days 2025

2025 - Selye János Scientific Forum

2025 - Association of Hungarian Biotechnology Students – Biotechnology Days

2024 - Neumann Colloquium 2024

2024 - IV. Hungarian Lifestyle Medicine Conference

2024 - Semmelweis Symposium 2024

2024 - Semmelweis PhD Scientific Days 2024

2024 - Association of Hungarian Biotechnology Students – Biotechnology Days

2023 - Semmelweis Symposium 2023

2023 - Semmelweis PhD Scientific Days 2023

2023 - „Day of the Heart" Health Development Offices Meeting 2023

9.2. Additional papers/abstracts not related to the PhD thesis

Papers

Domján Péter, **Angyal Viola**, Vingender István

Concentration and geospatial modelling of Health Development Offices' accessibility for the total and elderly populations in Hungary BMC PUBLIC HEALTH 25 Paper:

1466, 16 p. (2025)

Közlemény: 36115753 | Szakcikk (Folyóiratcikk) | Tudományos

Scopus - Public Health, Environmental and Occupational Health SJR indikátor: Q1

IF: 3,6

Domján P, **Angyal VZs**, Vingender I

The role and potential of Health Development Offices (HDOs) in maintaining the health of older people in Hungary

DEVELOPMENTS IN HEALTH SCIENCES 7: 2 pp. 38-45. (2024) Közlemény:

35623540 | Szakcikk (Folyóiratcikk) | Tudományos

Angyal Viola Zsuzsanna

Homoktövis (Hippophae rhamnoides L.) antioxidáns kapacitásának és antibakteriális hatásának vizsgálata

In: Áldorfai György; Tóth Tamás (szerk.) "SZIE kiváló tehetségei" konferencia előadásainak összefoglaló kiadványa

Gödöllő, Magyarország: Szent István Egyetemi Kiadó (2018) 157 p. pp. 9-9.

Közlemény: 3344984 | Absztrakt | Kivonat (Könyvrészlet) | Tudományos

Angyal Viola Zsuzsanna, Kovács Mónika, Stefanovitsné Bányai Éva

Homoktövis (Hippophae rhamnoides L.) antioxidáns kapacitásának és antibakteriális hatásának vizsgálata

In: Fodor Marietta; Bodor Péter (szerk.) ITT - Ifjú Tehetségek Találkozója

Budapest, Magyarország: Szent István Egyetem, Budai Karok (volt KÉE) (2017) 162 p. pp. 72-75.

Közlemény: 3303936 | Konferenciaközlemény (Könyvrészlet) | Tudományos

Summarized IF: 3,6

Overall IF value: 14,8

Conference abstracts

2020 - 18th International Congress of the Hungarian Society for Microbiology

2020 - Abstract for the 4th National Conference of Young Biotechnologists

XXXIV. OTDK Conference abstract (ISBN 978-963-490-078-8)

2019 - Annual Meeting 2019 of the Institution of Enzymology, Laboratory of Genome Metabolism and Repair"

2018 - Meeting of Young Talents conference abstract (ISBN 978-963-269-114-5)

2018 - Abstract for the International Symposium on Analytical and Environmental Problems (ISBN 978-963-306-623-2)

2018 - Conference of Outstanding Talents abstract (ISBN: 978-963-269-732-1)

2018 - Hungarian National MÉTE TDK conference abstract

2018 - Hungarian Academy of Sciences, Committee of Horticulture and Food Science, Subcommittee on Food Science workshop abstract (ISBN: 978-963-508-900-0)

9.3. Awards

- Semmelweis Symposium 2025 – Poster Presentation Award 1st place
- Semmelweis University PhD Scientific Days 2024 – 1st place
- Future Health Development Conference - Article Writing Challenge – 3rd place
- Hungarian Ministry of Interior – Special Award (ScreenGPT)

10. ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to **Prof. Dr. Dinya Elek**, to my supervisor, whose continuous support, insightful guidance, and thoughtful advice have accompanied me throughout these past years. His exceptional expertise and encouragement have greatly enriched both my professional development and the completion of this dissertation.

I also gratefully acknowledge **Dr. Tóth János**, who guided me with his expertise as an industry professional and continuously supported me with practical insights throughout this work.

I am also sincerely thankful to **Dr. Bertalan Ádám**, who not only made the ScreenGPT web application available but also implemented the GPT-4,4o API and provided essential support in preparing the user experience questionnaire.

My thanks extend to **Péter Domján**, who provided valuable assistance with the statistical analysis of the study.

I would like to express my deep appreciation to **Dr. Feith Helga** for sharing her valuable knowledge in questionnaire design.

I wish to highlight the dedicated contributions of the BSc students from the Digital Health Sciences Institute at Semmelweis University, who actively assisted in distributing the user experience questionnaire and supported the development process.

I am equally grateful to our colleagues from the Health Sciences Division for their valuable advice, encouragement, and constant support during the development phase and data evaluation.

Finally, I owe my deepest appreciation to my family, whose patience and unwavering support have sustained me throughout this journey. I would especially like to highlight my sister, who not only encouraged me every step of the way but also shared the unique experience of pursuing a PhD in parallel with me.

Funding

SUPPORTED BY THE 2024-2.1.2-EKÖP-KDP-2024-00002 UNIVERSITY RESEARCH SCHOLARSHIP PROGRAMME OF THE MINISTRY FOR CULTURE AND INNOVATION FROM THE SOURCES OF THE NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION FUND