

SEMMELWEIS EGYETEM
DOKTORI ISKOLA

Ph.D. értekezések

3524.

BERTALAN ÁDÁM

Nép-és közegészségtudományok
című program

Programvezető: Dr. Ács Nándor, egyetemi tanár

Témavezető: Dr. Dinya Elek, professzor emeritus

Enhancing Patient Safety Through Electronic Health Record System Integrated Decision Support for Medication Risk Assessment

PhD thesis

Bertalan Ádám, Pharm.D.

Semmelweis University Doctoral School
Mental Health Sciences Division



Supervisor: Elek Dinya, PhD.

Official reviewers: Balázs Balogh, PhD.
László Molnár, MD., PhD

Head of the Complex Examination Committee:
Pál Maurovich-Horvat, MD., PhD., MPH.

Members of the Complex Examination Committee:
András Kaposi, PhD.
Zoltán Ádám Tamus, PhD.

Budapest
2026

Table of Contents

List of Abbreviations	5
1 Introduction	6
1.1 Background and Context of AI in Clinical Pharmacy	6
1.1.1 Evolution of Clinical Pharmacy Practice	6
1.1.1.1 Historical Development	6
1.1.1.2 Shift Towards Patient-Centered Care	6
1.1.2 Medication Safety and Risk Assessment	7
1.1.2.1 Definition and Importance	7
1.1.2.2 Common Challenges in Medication Safety	8
1.1.3 Artificial Intelligence in Healthcare	8
1.1.3.1 Historical and Technological Shifts	8
1.1.3.2 Overview of AI Technologies	9
1.1.3.3 Applications of AI in Clinical Decision Support	10
1.2 The AUGMERIS Approach to Unstructured Clinical Data	11
1.2.1 The Medication Risk Score	11
1.2.1.1 Augmented Medication Risk Score	12
1.2.2 The Streamlit Framework	13
2 Objectives	15
3 Methods	17
3.1 Software Infrastructure and Development Frameworks	17
3.2 Development Environment and Interactive Computing	17
3.3 Evolution of Applied Language Models	19
3.4 Automation of MERIS Assessment	19
3.5 Development of Text Processing Modules	21
3.5.1 The AUGMERIS Hybrid Architecture	21
3.5.2 The Four-Step Extraction Pipeline	22
3.6 In Silico Validation on Synthetic Data	25
3.7 Internal Validation	26

3.7.1	Setting & Data Collection	26
3.7.2	Evaluation Protocol	26
3.7.3	Data Analysis	26
3.8	Prospective Pilot Testing (Web Application)	27
3.8.1	Deployment and Tracking	27
3.8.2	Data Analysis	27
3.9	Impact Analysis (Clinical Implementation)	29
3.9.1	Architecture and Deployment	29
3.9.2	Workflow Integration	30
3.9.3	Data Aggregation	31
4	Results	33
4.1	In Silico Performance Comparison of LLM Generations	33
4.2	Internal Validation of the Text-Processing Module	35
4.3	Prospective Pilot Testing: Global Usage Data	35
4.3.1	Risk Distribution and Performance	36
4.3.2	AI Recognition Rates	36
4.4	Impact Analysis: PiGuard Clinical Implementation	36
4.4.1	Risk Distribution and Pharmacist Focus	36
4.4.2	Intervention Characteristics	38
4.4.3	Delta MERIS (Risk Reduction)	38
5	Discussion	40
5.1	Performance of the Hybrid Text-Processing Module	40
5.2	Insights from Global Pilot Testing	41
5.3	Clinical Impact and the PiGuard System	41
5.3.1	User Trust, Transparency, and Clinical Impact	41
5.3.2	Interdisciplinary Integration and Technical Robustness	42
5.3.3	PiGuard Deployment and Efficacy	42
5.4	Strengths and Limitations	43
5.4.1	Architectural Advantages and Structural Strengths	43
5.4.2	Methodological and Operational Limitations	43
5.4.3	Navigating Rapid Technological Evolution in Medical AI	44
5.5	International Overview and Parallel Experiences	44
5.5.1	Comparing AUGMERIS with the ScreenGPT Project	45
5.6	Professional Recognition and Industry Feedback	47
5.7	Future Directions	47
5.7.1	Algorithmic and Data Expansion	47

5.7.2	System Integration and Scalability	48
5.7.3	Commercialization and Technology Transfer	48
6	Conclusions	49
7	Summary	50
8	References	51
9	Bibliography of the Candidate's Publications	62
9.1	Publications Related to the Dissertation	62
9.2	Other Publications	63
9.3	Book	64
9.4	Conferences	64
10	Acknowledgements	66

List of Abbreviations

ADR	Adverse Drug Reaction	HTML	HyperText Markup Language
AI	Artificial Intelligence		
API	Application Programming Interface	IDE	Integrated Development Environment
APP	Application	INN	International Nonproprietary Name
ATC	Anatomical Therapeutic Chemical	JSON	JavaScript Object Notation
AUGMERIS	Augmented Medication Risk Score	LAN	Local Area Network
B2B	Business-to-Business	LLM	Large Language Model
B2C	Business-to-Consumer	MAE	Mean Absolute Error
CDS	Clinical Decision Support	MERIS	Medication Risk Score
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration	ML	Machine Learning
		NEAK	National Health Insurance Fund
		NLP	Natural Language Processing
CDSS	Clinical Decision Support System	NLTK	Natural Language Toolkit
CSS	Cascading Style Sheets	NNGYK	National Centre for Public Health and Pharmacy
ECG	Electrocardiogram		
eGFR	Estimated Glomerular Filtration Rate	PUPHA	Public Pharmaceutical Database
EHR	Electronic Health Record	re	Regular Expression
EIR	Enterprise Information Repository	RMSE	Root Mean Square Error
		SaaS	Software as a Service
GCE	Google Cloud Engine	SLM	Small Language Model
GDPR	General Data Protection Regulation	UI	User Interface
		VM	Virtual Machine
GPT	Generative Pre-trained Transformer	VS Code	Visual Studio Code
HIS	Hospital Information System	WHO	World Health Organization

1. Introduction

1.1. Background and Context of AI in Clinical Pharmacy

1.1.1. Evolution of Clinical Pharmacy Practice

1.1.1.1 Historical Development

Clinical pharmacy has evolved from a product-focused discipline into a patient-oriented field in which pharmacists contribute directly to therapeutic decision-making and medication safety [1, 2]. The spread of electronic health records (EHRs) widened this clinical role, but it also exposed a fundamental limitation of earlier decision support: much of the information relevant to medication risk is documented in heterogeneous free text rather than in structured variables [3]. As a result, pharmacists increasingly need tools that can interpret both coded data and narrative clinical documentation.

The first generations of clinical decision support systems (CDSS) were largely rule-based and derived from guidelines, pharmacological knowledge, and structured inputs such as laboratory values or diagnosis codes [4]. These systems offered consistency and traceability, but they were poorly suited to fragmented notes, local abbreviations, and cross-institutional variation in terminology. Later machine learning and deep learning methods expanded predictive capability and enabled the analysis of more complex clinical patterns, including temporal prescription signals and high-dimensional data streams, yet this gain often came at the cost of transparency and clinician trust [3, 5]. In this context, hybrid models emerged as a pragmatic compromise: they preserve deterministic safeguards while adding the flexibility needed to interpret unstructured records and infer clinically relevant relationships [6]. This historical trajectory is directly relevant to our research, which reflects the broader movement in clinical pharmacy away from purely static rule engines and away from unconstrained probabilistic systems toward architectures that combine both. Generative models can now process narrative EHR text with far greater contextual sensitivity than earlier approaches, but safe implementation still depends on lifecycle evaluation, workflow fit, and regulatory oversight [4, 5]. The lesson from previous CDSS deployments is therefore not that automation should replace pharmacists, but that effective systems must strengthen professional judgment while reducing manual review burden.

1.1.1.2 Shift Towards Patient-Centered Care

The transition toward patient-centered care has further changed what is expected from clinical decision support. Instead of optimizing only generalized population-level

rules, modern pharmacy practice requires tools that account for individual physiology, adherence, preferences, and social context [7]. Many of these determinants are documented unevenly across EHR narratives, so systems limited to structured inputs may miss factors that materially affect medication safety. Hybrid artificial intelligence (AI) approaches are therefore attractive because they can combine laboratory and prescribing data with narrative evidence about tolerance, lifestyle, or barriers to adherence [3].

Patient-centered care also requires that recommendations remain explainable in clinician - patient communication. Real-time monitoring, updated progress notes, and longitudinal EHR review can support earlier recognition of changing risk, but those benefits are limited if clinicians cannot explain why a recommendation was generated [4, 8]. By grounding AI interpretation in validated rule structures, systems can support personalized risk assessment while preserving trust, privacy awareness, and the evidential clarity needed for shared decision-making [5, 9].

1.1.2. Medication Safety and Risk Assessment

1.1.2.1 Definition and Importance

Medication safety aims to prevent harm arising from medication use through accurate prescribing, careful monitoring, and timely intervention. Within clinical pharmacy, risk assessment estimates the likelihood of medication-related problems by combining patient characteristics, therapeutic complexity, and contextual information from health records [10]. Validated scoring systems are important because they convert this complexity into reproducible priorities, helping pharmacists identify high-risk patients before adverse events occur [11–15].

The clinical value of such assessment is both patient-centered and organizational. Preventing medication-related harm reduces avoidable hospitalizations and supports more efficient allocation of limited professional attention [5, 16–19]. AI-enhanced CDSS can strengthen this process by screening large volumes of data, including narrative documentation that may contain clues about adherence problems, side effects, or contraindications not visible in structured fields [3, 20]. At the same time, the use of AI raises the risk of automation bias if outputs are accepted uncritically. Hybrid systems help counter this by keeping risk predictions anchored to interpretable rules and by preserving the pharmacist’s active role in verification and intervention.

Medication safety also carries ethical, regulatory, and usability implications. Risk scores influence communication with patients, quality improvement programs, and, increasingly, the compliance pathways required for clinical software deployment [5, 9]. For these reasons, a useful medication safety tool must do more than detect risk accurately: it

must present evidence clearly, fit into practice workflows, and remain dependable even when source documentation is incomplete or inconsistent [10, 21].

1.1.2.2 Common Challenges in Medication Safety

The most persistent challenge is data quality. Unstructured EHR entries often contain missing details, ambiguous phrasing, and inconsistent terminology, all of which can weaken medication risk assessment [3]. Deterministic systems may ignore such data altogether, whereas generative models may over-interpret weak context and introduce speculative conclusions [5]. A related problem is algorithmic bias: models trained on non-representative datasets may perform unevenly across patient groups, making local validation and fairness-oriented evaluation essential.

Equally important are trust, workflow integration, and regulation. Pharmacists are less likely to adopt opaque systems that cannot justify their outputs, especially when alerts affect high-stakes decisions [4]. Even accurate systems may fail in practice if they interrupt workflow, communicate uncertainty poorly, or prove difficult to adapt across institutions with different documentation habits [22]. Regulatory expectations further raise the bar by requiring privacy protection, evidence of safe operation, and revalidation when models or guidelines change [3, 5].

These challenges are amplified in longitudinal care, where patient risk changes over time. Static scoring systems can become outdated, but highly adaptive models may also lose interpretability if their reasoning shifts unpredictably. Hybrid architectures provide a useful middle ground by combining stable rule-based foundations with adaptable language processing, provided they are monitored continuously and refined through pharmacist feedback [4].

1.1.3. Artificial Intelligence in Healthcare

1.1.3.1 Historical and Technological Shifts

To synthesize the historical and technological shifts described above, Figure 1 provides a chronological overview of the evolution of clinical pharmacy practice. This timeline illustrates the transition from early, purely rule-based systems applied to structured data, toward the current paradigm where hybrid AI models facilitate personalized, patient-centered care by safely interpreting unstructured EHR narratives.

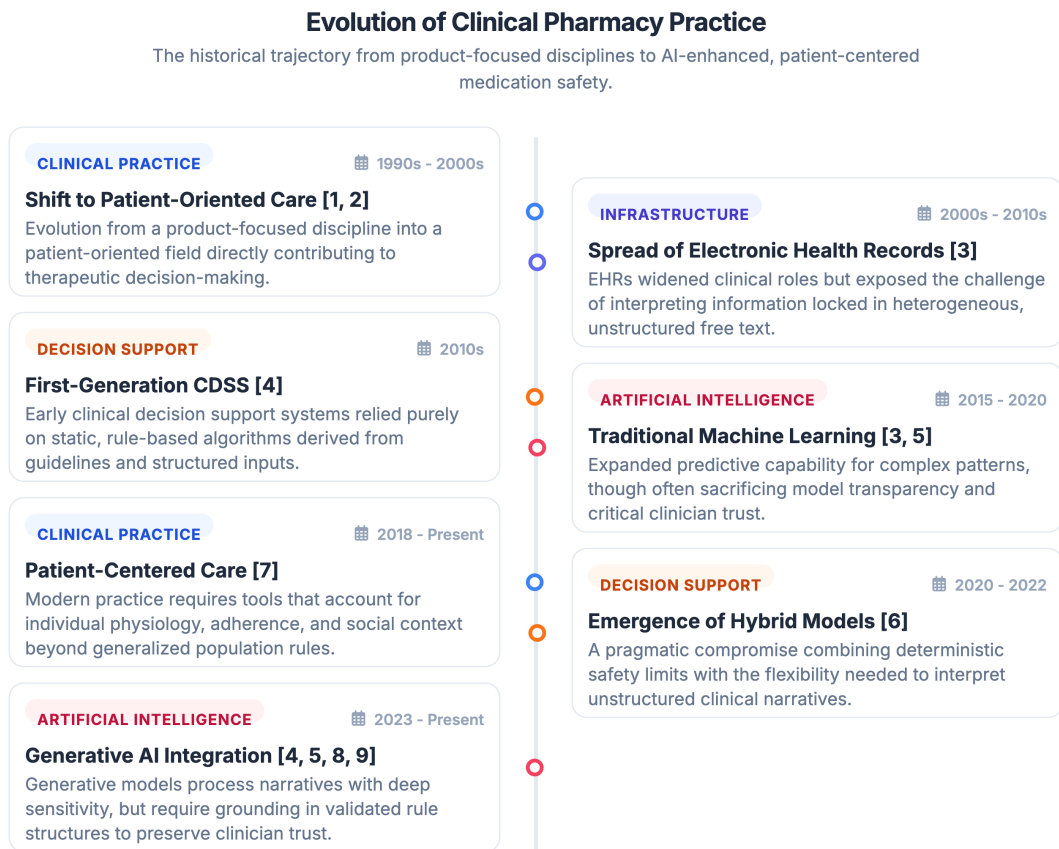


Figure 1. Timeline of clinical pharmacy practice evolution and AI integration

1.1.3.2 Overview of AI Technologies

AI in healthcare includes several overlapping technological families. Rule-based systems remain valuable when decisions can be tied to explicit guidelines and well-structured variables, but they are limited when clinically meaningful information is buried in narrative text [23]. Machine learning methods broadened the scope of prediction by learning non-linear relationships from data, while deep learning further improved performance in image analysis, temporal signal processing, and other high-dimensional tasks [10, 20]. However, as model complexity increases, so do concerns about interpretability, bias, and computational demands.

The evolution of Natural Language Processing (NLP) has been fundamentally reshaped by the emergence of Large Language Models (LLMs), particularly through the introduction of the Transformer architecture and its self-attention mechanisms [24]. Unlike earlier statistical methods [25] or traditional sequence tagging approaches [26], modern LLMs are trained on extensive text corpora. This allows them to capture complex linguistic patterns,

achieve high-level machine comprehension [27], and generate human-like text with deep contextual awareness [28].

In the healthcare domain, this capability is transformative. A substantial portion of clinically relevant data—including medication histories, physician observations, and treatment plans—is siloed in unstructured clinical notes rather than standardized fields. LLMs and their multimodal counterparts [29, 30] offer a systematic way to categorize and review these complex datasets [31].

Furthermore, the recent shift toward Small Language Models (SLMs) has demonstrated that models with fewer parameters, when trained on high-quality medical textbooks and domain-specific data, can achieve enhanced reasoning skills and provide reliable decision support for specific conditions like breast cancer [32, 33]. These efficient models provide a transparent and data-secure alternative for tasks such as medication risk assessment, clinical documentation analysis, and the development of personal medical assistants [34, 35].

NLP has become particularly important for clinical pharmacy because medication risk often depends on details embedded in notes rather than in coded fields. Earlier NLP pipelines extracted entities and keywords from text, but large language models can interpret longer contexts and infer clinically relevant meaning more flexibly [36–38]. This makes them attractive for EHR-based medication risk assessment. Yet their probabilistic nature also creates safety concerns, especially when hallucinated or guideline-inconsistent statements are presented with high fluency [6, 23]. For this reason, domain adaptation, explainability, and deterministic guardrails are central to safe deployment rather than secondary technical refinements.

Across all AI categories, governance remains essential. Bias auditing, transparent evaluation, and privacy-preserving methods such as federated learning are increasingly important for healthcare implementation [3, 5]. In practice, the most robust systems are often those that integrate complementary methods rather than relying on a single model family, which explains the growing interest in hybrid architectures [6, 36].

1.1.3.3 Applications of AI in Clinical Decision Support

AI is already being used across several medication-related clinical decision support (CDS) tasks. In pharmacovigilance, models combine laboratory values, medication histories, and symptom documentation to identify patients at risk of adverse drug events, while hybrid pipelines can extract additional warning signs from nursing or physician notes [4, 20]. Similar principles apply to medication reconciliation, where AI can compare structured medication lists with discharge summaries or referrals to identify omissions, duplications, and unsafe continuations across care transitions [39]. In resource-limited

settings, AI-assisted triage and dispensing support can also help pharmacies prioritize high-risk cases more efficiently [40].

Clinical applications extend into more specialized domains as well [41–43]. Hybrid models can support cardiology-related medication management by linking Electrocardiogram (ECG)-derived risk with narrative evidence about recent drug changes or comorbidities [44]. At the institutional level, AI contributes to laboratory quality control and to forms of personalized medicine such as genetic risk interpretation [8, 20]. Across these use cases, the same adoption principle recurs: systems are most useful when they are explainable, minimally disruptive, locally adaptable, and designed to support – rather than replace – clinical accountability [4, 5, 9].

1.2. The AUGMERIS Approach to Unstructured Clinical Data

1.2.1. The Medication Risk Score

Medication-related harm represents a substantial challenge in contemporary healthcare systems, particularly in populations exposed to polypharmacy and complex pharmacotherapy. Adverse drug reactions (ADRs), drug–drug interactions, and medication-related complications are major contributors to hospitalizations and healthcare burden [45, 46]. Consequently, the development of reliable clinical decision-support tools capable of identifying high-risk medication regimens has become a key objective in clinical pharmacology and pharmacy practice [47].

One approach to addressing this challenge is the use of structured medication risk assessment models. Among these, the Medication Risk Score (MERIS), originally developed in Denmark [48, 49], provides a systematic framework for evaluating the potential risks associated with pharmacotherapy. The conceptual logic and decision flow of the original algorithm is illustrated in Figure 2 [50]. The MERIS methodology assesses medication-related risks by considering factors such as the number of simultaneous drugs and their chance for interaction and making a harmful effect, as well as estimated glomerular filtration rate (eGFR), which refers to renal function. The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation provides the eGFR value, which more accurately represents the patient’s kidney function than other estimation methods. [51, 52]. It is the most important parameter from the aspect of medication because impaired renal function can lead to decreased drug clearance, and it’s the most common reason for increased accumulation and toxicity, particularly for renally excreted medications and those with a narrow therapeutic index. As a result, dose adjustments or drug avoidance may be necessary based

Medication Risk Score (MERIS)		
Variable	Interval	Points
 Impaired renal function	eGFR > 60	0
	60 > eGFR > 30	5
	eGFR < 30	10.6
 Total number of drugs	0–5	0
	6–11	5
	>12	10.6
 Number of drugs with	Low risk of harm (Max. number of drugs = 3)	0.25
	Medium risk of harm (Max. number of drugs = 8)	0.5
	High risk of harm (Max. number of drugs = 7)	1
	Low + medium risk of interaction (Max. number of drugs = 12)	0.25
	High risk of interaction (Max. number of drugs = 2)	0.5

Figure 2. Logic of the original MERIS algorithm [50].

on eGFR levels. These components work together to help clinicians, especially clinical pharmacists, identify patients who might benefit from targeted medication review and therapy optimization [53].

1.2.1.1 Augmented Medication Risk Score

MERIS approach allows clinicians and clinical pharmacists to identify patients who may benefit from targeted medication review and optimization of therapy. The automation of clinical decision support for medication safety presents an intriguing intersection between data science and pharmacy practice.

This concept can expand through the integration of artificial intelligence. The Augmented Medication Risk Score (AUGMERIS) represents an AI-supported implementation of the MERIS methodology designed to automate medication risk assessment using EHR

data [54]. By applying natural language processing techniques and generative language models to extract medication information from unstructured clinical text, the system can automatically calculate medication risk scores. Such AI-assisted tools have the potential to improve the efficiency and scalability of medication risk evaluation and support safer pharmacotherapy management in clinical practice.

This dual approach can directly process unstructured data stored in EHRs, where conventional deterministic systems alone often fail due to the variability and incompleteness of patient documentation. Such unstructured datasets – characterized by missing values, ambiguous phrasing, or inconsistent terminologies – pose significant challenges for model interpretability and reliability [55]. Historically, CDSS have relied on structured inputs from in vitro diagnostics and laboratory data. However, in contexts such as MERIS scoring, crucial information is often buried within unstandardized, free-text EHR notes, where traditional machine learning struggles with semantic nuances and context-specific interpretations.

1.2.2. The Streamlit Framework

Streamlit is an open-source Python framework that enables the rapid creation and deployment of data-driven applications, particularly those involving machine learning and artificial intelligence. Through its intuitive, declarative programming style and high-level abstractions, developers can build interactive web interfaces with relatively little code. When combined with cloud infrastructure, Streamlit allows applications to be deployed in a scalable and real-time environment, enabling users to interact with custom Generative Pre-trained Transformer (GPT) models directly through a standard web browser without requiring specialized software or local computing resources.

As a cloud-based platform, Streamlit supports straightforward hosting, continuous updates, and remote access, making it well suited for collaborative workflows and distributed teams. This infrastructure enhances system reliability, simplifies version management, and enables applications to scale dynamically according to user demand [56]. In the case of custom GPT solutions, this architecture allows users to access specialized, domain-specific language model outputs instantly, regardless of their technical expertise or geographic location.

Additionally, Streamlit offers a wide range of interactive components, including widgets and visualization tools, which improve usability and encourage more effective interaction with the underlying models. These features are particularly valuable in fields such as healthcare, education, and scientific research, where interpretability, user feedback, and adaptability are essential. By utilizing Streamlit as a cloud-based deployment platform, developers can convert static language models into interactive and accessible applications,

thereby enhancing their practical usability and scalability in real-world contexts [57].

2. Objectives

The primary objective of this research is to develop, validate, and operationalize the AUGMERIS system as a hybrid medication risk assessment tool capable of functioning across heterogeneous EHR environments, which ensures that outputs remain anchored in authoritative pharmacological evidence while the sources are augmented with semi-structured or free-text inputs. The ambition is to combine deterministic rule-based algorithms with GPT modules in a manner that maximizes both reproducibility and interpretive flexibility for patient safety applications. This dual-layer design aimed to address persistent gaps in existing approaches, including limitations of traditional models in handling the diversity of drug nomenclatures and the incomplete coverage obtained from structured field-only parsing.

Furthermore, by automatically prioritizing patients according to empirically validated MERIS categories [58], the system aims to focus pharmacists' attention toward those most probable to benefit from immediate intervention. Our vision is that a risk map of AUGMERIS ranking, integrated directly into routine workflows, reduces time spent finding the high-need patients, leads to measurable gain in intervention efficiency, decreases preventable adverse events, and optimizes resource usage without compromising safety or equity in patient care.

The research encompasses multiple phases, from algorithmic architecture refinement through multicenter validation with their own partial purposes:

- **Automation of MERIS assessment:** Develop a rule-based engine that codifies the MERIS algorithm, ensuring that it can process structured data inputs with high accuracy.
- **Development of text processing module:** Design GPT-supported NLP models to extract relevant medication risk factors from unstructured EHR entries, including handling of local abbreviations and variable documentation styles.
- **In Silico Validation of the Text-Processing Module:** To assess the isolated extraction accuracy, reliability, and error-tolerance of the AI-powered parser using a controlled, synthetic dataset containing simulated complex therapies and typographical errors.
- **Internal Validation:** Evaluate the system's performance on retrospective EHR datasets from multiple institutions, comparing AUGMERIS outputs against manual pharmacist assessments and existing risk scores.

- **Prospective Pilot Testing:** Deploy the AUGMERIS web app via a Streamlit server to reach live clinical environments to assess real-time performance, worldwide user interest, and workflow integration.
- **Impact Analysis:** Measure the effect of AUGMERIS on pharmacist efficiency, intervention rates, and patient outcomes related to medication safety.

In summary, our main hypothesis is that the hybrid risk evaluation model, which combines GPT-based natural language processing with rule-based algorithms, will improve the clinical usability and accuracy of MERIS assessments across diverse EHR environments, and this will help clinical pharmacists make better decisions to reduce preventable adverse drug events and improve the efficacy of interventions.

3. Methods

3.1. Software Infrastructure and Development Frameworks

The methodology of this research was structured to encompass the full lifecycle of software development for clinical decision support, moving from algorithmic design to real-world clinical validation. The development environment evolved during the research period; while initial experiments were conducted using Python 3.8.10, the final validation and system deployment utilized version 3.11.6 alongside client-side web technologies (HyperText Markup Language 5 (HTML5), Cascading Style Sheets 3 (CSS3), JavaScript ES6) for user interfaces. Key Python libraries included the Natural Language Toolkit (NLTK) for initial text processing [59], Pandas for data manipulation [60], Streamlit for rapid web application deployment [61], and the OpenAI Application Programming Interface (API) for accessing generative models [62]. The full list of tools and frameworks used in this research is shown in Table 1, and the methodology is detailed below according to the specific phases outlined in the research objectives.

3.2. Development Environment and Interactive Computing

The computational experiments and data analysis pipelines were primarily developed using interactive computing environments. The core of the development process relied on **IPython** (Interactive Python) notebooks, which provide an architecture for interactive computing by decoupling the code execution engine (the kernel) from the user interface [85].

To manage these notebooks, the **Jupyter** ecosystem was utilized. Jupyter Notebook is an open-source web application that allows for the creation and sharing of documents containing live code, equations, visualizations, and narrative text [75]. This environment proved essential for the iterative nature of LLM prompt engineering and clinical data preprocessing, as it allows for immediate feedback and state preservation between execution cells.

The primary Integrated Development Environment (IDE) used for the project was **Visual Studio Code** (VS Code). VS Code is a highly extensible source-code editor that supports Jupyter workflows natively through specialized extensions [84]. To ensure the reproducibility of the development process and maintain rigorous version control of the source code, the Git distributed version control system was employed throughout the

Table 1. Summary of information technology tools and frameworks

Technology / Tool	Version	Primary Function	Reference
Apache	2.4.41	Web server (PiGuard)	[63]
CSS3	–	Client-side styling	[64]
Folium	0.19.4	Interactive maps	[65]
GeoPandas	1.0.1	Geographic data processing	[66]
Git	2.42.0	Version control	[67, 68]
GitHub	–	Repository hosting and deployment	[69]
hashlib	–	Cryptographic hashing (PiGuard)	[70]
HTML5	–	Client-side frontend structure	[71, 72]
JavaScript	ES6	Client-side scripting	[73]
JSON	2.0.9	Data interchange format	[74]
Jupyter Notebook	7.2.0	Interactive computing environment	[75]
Matplotlib	3.9.2	Data visualization	[76]
MySQL	8.0	Database management (PiGuard)	[77]
NLTK	3.8.1	Natural language processing	[59]
NumPy	1.26.2	Numerical computing and array processing	[78]
OpenAI API	1.54.1	Generative AI integration	[62]
Pandas	2.1.3	Data manipulation and analysis	[79]
Pillow	11.0.0	Image processing	[80]
Pydantic	2.10.2	Data validation	[81]
Python 3	3.11.6	Core programming language	[60, 82]
Regular Expression (re)	2.2.1	Pattern matching and text manipulation	[82]
Seaborn	0.13.0	Statistical data visualization	[83]
Streamlit	1.31	Web application framework	[61]
Visual Studio Code	1.85	Integrated Development Environment	[84]

project [67, 68]. The research tasks were performed using two complementary approaches:

- **Integrated IDE Workflow:** Using the VS Code Jupyter extension, which allows for advanced debugging, IntelliSense code completion, and version control integration directly within the notebook interface.
- **Local Server Workflow:** Starting a local Jupyter server and accessing the environment via a standard web browser interface, providing a lightweight and robust execution environment independent of the IDE.

This dual-layered setup ensured both the agility of interactive development and the stability required for processing large hospital datasets during the validation phases.

3.3. Evolution of Applied Language Models

Given the rapid advancement of Artificial Intelligence during the research period, the language models applied in our methodology evolved continuously. Initial processing attempts in 2022 relied strictly on traditional NLP techniques (using the NLTK library and regular expressions), which proved insufficient for handling the high variability of unstructured clinical narratives.

Subsequently, a preliminary comparative study presented at the XXXVI. Neumann Colloquium evaluated the standalone capabilities of OpenAI’s GPT-3.5-turbo and Google’s Vertex AI (text-bison-32k) models against strict rule-based extraction [86]. The insights from these early experiments necessitated the development of the hybrid architecture.

To provide a clear chronological context of the technical progression, Table 2 summarizes the different GPT and alternative models utilized across the various phases of the project.

3.4. Automation of MERIS Assessment

The foundation of the system is the computational automation of MERIS, originally developed by Sædder et al. [87]. This scale, ranging from 0 to 36.2 points, evaluates the risk of concomitant medication use based on the patient’s estimated glomerular filtration rate (eGFR) and the specific combination of applied drugs. Therapies under 14 points are considered low-risk, between 14 and 26 points as high-risk, and above 26 points as extreme-risk.

To automate this, we developed a dedicated Python class. The method requires two arguments: the eGFR value (integer or null, assuming healthy kidney function if null) and a

Table 2. Chronological evolution of applied large language models in the research

Model	Timeframe	Research Phase / Application
GPT-3.5-turbo & Vertex AI	2023 (Early phase)	Proof-of-concept testing and comparative analysis of standalone AI capabilities on retrospective hospital data [86].
GPT-4o	2023 – 2024	Initial development of the AUGMERIS web application’s text-processing module. Limited by frequent JSON formatting errors (hallucinations).
GPT-4.1	2024 – Present	Deployed in the global web application and the on-premise PiGuard system. Utilizes <i>Structured Outputs</i> to strictly enforce schema compliance and eliminate parsing errors.
GPT-4o-mini	2024 (Late phase)	Baseline model used for the In Silico validation (synthetic stress test) due to its high processing speed and cost-efficiency.
GPT-5.3-mini	2025	Used to repeat the In Silico validation, demonstrating the generational improvement in accuracy and error-tolerance compared to the legacy mini model.

list of Anatomical Therapeutic Chemical (ATC) codes representing the active medications. The algorithm calculates three subscores:

Kidney Function Subscore: Based strictly on the eGFR value.

Medication Quantity Subscore: Derived from the total count of the simultaneously administered medications (polypharmacy).

Medication Quality Subscore: Calculated by evaluating the harm and interaction potentials of the 53 specific drug classes defined in the original MERIS publication [49].

To facilitate programmatic execution, the original 53 drug classes were mapped meticulously to their corresponding hierarchical ATC codes [88], allowing the algorithm to correctly apply inclusion and exclusion rules during the evaluation process. To demonstrate the translation of complex clinical guidelines into programmatic logic, Figure 3 illustrates the Python implementation for determining the medication quality subscore. This function evaluates both the harm and interaction potentials of the identified drug classes, incorporating specific accumulation limits defined by the MERIS protocol.

```

1 def class_score(self, ATCs):
2     """Calculates the Medication Quality Subscore"""
3     drugs = self.collect_classes(ATCs)
4     score = 0
5     counter = [0, 0, 0, 0, 0]
6     for drug in drugs:
7         # Evaluate harm potential
8         harm = self.drug_classes[drug][1]
9         if harm == 0.25:
10             counter[0] += 1
11             if counter[0] < 4:
12                 score += harm
13         elif harm == 0.5:
14             counter[1] += 1
15             if counter[1] < 9:
16                 score += harm
17         elif harm == 1:
18             counter[2] += 1
19             if counter[2] < 8:
20                 score += harm
21         # Evaluate interaction potential
22         interaction = self.drug_classes[drug][2]
23         if interaction == 0.25:
24             counter[3] += 1
25             if counter[3] < 13:
26                 score += interaction
27         elif interaction == 0.5:
28             counter[4] += 1
29             if counter[4] < 3:
30                 score += interaction
31     return score

```

Figure 3. Python method demonstrating the calculation of the MERIS quality subscore, applying specific harm and interaction accumulation limits.

3.5. Development of Text Processing Modules

3.5.1. The AUGMERIS Hybrid Architecture

To address these limitations, GPT models – such as OpenAI’s GPT-4.1 – introduce the capability to extract subtle relationships and identify implicit patient risk factors within narrative content, drastically reducing human interpretation time [20]. However, because

generative outputs are probabilistic rather than deterministic, they must be combined with domain-specific rules to mitigate misclassification risks.

AUGMERIS overcomes this challenge through a hybrid architecture where explicit clinical rules serve as strict boundaries for generative outputs. This hybridization exploits the strengths of both paradigms: rule-based CDS ensures strict compliance with established clinical guidelines, while AI-driven NLP enables the rapid parsing of diverse, unstructured datasets that would otherwise require substantial manual review [3]. Ultimately, this synergy expands diagnostic coverage across extensive EHR corpora and mitigates operational bottlenecks like time constraints and staff shortages, efficiently flagging patients at high risk for adverse drug interactions or contraindications.

3.5.2. The Four-Step Extraction Pipeline

Translating this conceptual architecture into a functional system requires a robust module capable of handling the highly unstructured nature of free-text therapy descriptions found in EHRs. To achieve this, the engineered text-processing workflow operates in a distinct, sequential pipeline shown in Figure 4.

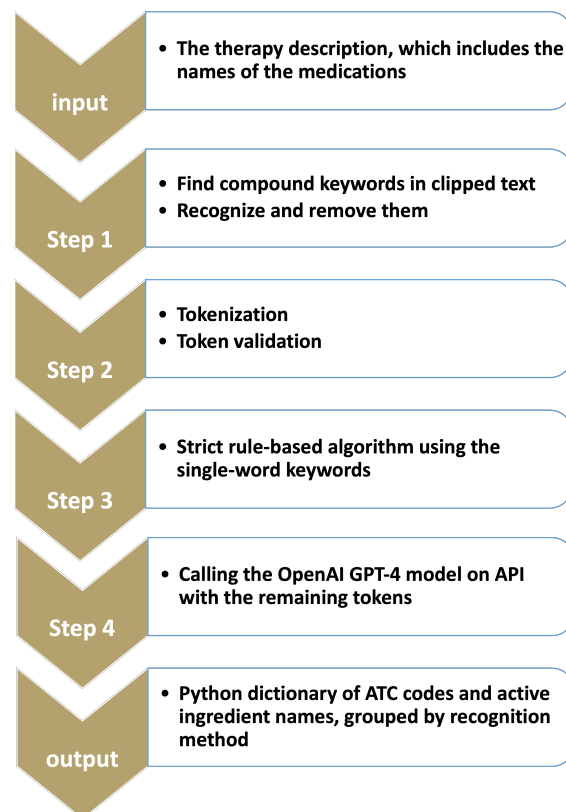


Figure 4. Text processing architecture (TPA) of the AUGMERIS system [54]

Step 1: Finding compound keywords in clipped text This initial step tries to identify multiword drug names and common abbreviations in the raw text, which often contains truncated or clipped entries. The algorithm eliminates the names of compounded drugs because splitting them could lead to incorrect identification.

Step 2: Tokenization and Validation: The raw, free-text therapy description is first tokenized using the NLTK library. These generated tokens are then validated to filter out irrelevant characters while retaining potential drug names.

Step 3: Strict Rule-Based Matching: A deterministic algorithm iterates through the validated tokens, searching for exact and complex compound matches against a custom dictionary. This dictionary is compiled from the Hungarian National Centre for Public Health and Pharmacy (NNGYK) [89] and the National Health Insurance Fund (NEAK) Public Pharmaceutical Database (PUPHA) [90]. Successfully recognized sequences are mapped to their respective ATC codes and removed from the active token pool.

Step 4: AI-Assisted Recognition: The remaining, unrecognized tokens — which frequently contain typographical errors, clinical abbreviations, or foreign drug names — are routed via API to OpenAI’s generative models. The integration of this step evolved significantly during the research. Early comparative testing demonstrated that while standalone models like Google’s Vertex AI and OpenAI’s GPT-3.5 could identify medications, they lacked the strict structural reliability required for automated scoring [86].

During the initial development of the web application, the GPT-4o model was employed. However, relying purely on "system prompt" instructions to format the output as an ATC classifier proved unreliable, frequently resulting in "invalid JSON" errors that disrupted the computational pipeline.

To resolve this, the system was upgraded to the GPT-4.1 model to utilize OpenAI’s *Structured Outputs* feature. This transition allowed us to strictly enforce a predefined JavaScript Object Notation (JSON) schema for the output (using the Pydantic library), practically eliminating JSON parsing errors and conversational hallucinations. The definition of the expected structured output schema is shown in Figure 5. The integration of this finalized approach is demonstrated in Figure 6, showing how the model is forced to return a valid schema containing only the necessary ATC code. Finally, these AI-generated classifications are merged with the deterministically recognized drugs to form the complete input dataset for the MERIS calculator.

```

1 #Model for the expected drug information in the OpenAI response
2 class DrugSchema(BaseModel):
3     atc: str
4     agent_name: str
5
6 #Model for the expected response structure from the OpenAI API
7 class ResponseSchema(BaseModel):
8     drugs: List[DrugSchema]
9     junk: List[str]

```

Figure 5. Pydantic definition of the expected structured output schema for the OpenAI API response, ensuring strict adherence to the required format.

```

1 def call_the_robot(token_list):
2     """Sends the unrecognized tokens to the OpenAI API"""
3     client = OpenAI(api_key=st.secrets["OpenAIkey"])
4     messages=[
5         # The system prompt controls the model's behavior
6         {"role": "system", "content": system_prompt},
7         # The user prompt provides the actual data
8         {"role": "user", "content": str(token_list)},
9     ]
10    response = client.responses.parse(
11        model="gpt-4.1", # Specify the model
12        input=messages, # The input messages for the model
13        text_format=ResponseSchema, # Define response format
14    )
15    try:
16        answer={
17            'drugs' : dict([(d.atc, d.agent_name) for d in response.
18                output_parsed.drugs]),
19            'junk' : response.output_parsed.junk
20        }
21    except:
22        answer = {'drugs': [], 'junk': token_list}
23        print(token_list, response.output_parsed)
24    return answer

```

Figure 6. OpenAI API call configured for strict, deterministic ATC classification.

3.6. In Silico Validation on Synthetic Data

Following the stabilization of the text-processing architecture via OpenAI’s *Structured Outputs*, an In Silico validation experiment was conducted using a controlled, synthetically generated dataset ($N = 100$). The primary objective of this stress-test was to isolate the parser’s extraction performance from the inherent ambiguity and variability of actual EHR narratives, ensuring the pipeline’s robustness before exposing it to live clinical data.

Because live API calls are resource-intensive, this high-volume, repetitive testing was initially performed using the cost-effective and highly optimized GPT-4o-mini model. Later, to quantify the rapid technological advancement in natural language processing, the exact same experiment was repeated using the newer GPT-5.3-mini model.

The dataset was constructed by randomly sampling 5 to 12 medication names per case from an independent drug repository, establishing a definitive "ground truth" for each quality subscore calculation. To simulate the linguistic patterns of routine clinical documentation, the names were truncated to remove packaging information, and various dosage instructions were appended. Furthermore, to test the system’s resilience against human clerical errors, simulated typographical errors (character swaps, duplications, or omissions) were injected into approximately 8% of the drug names. The configuration parameters for the synthetic dataset generation are summarized in Table 3.

Table 3. Parameters for the synthetic dataset generation pipeline.

Parameter	Value / Description
Data Source	Independent drug repository (not used in parsing)
Preprocessing	Keyword-based packaging truncation
Total Sample Size (N)	100 cases
Medications per Case	5 – 12 (randomized)
Dosage Instruction Presence	100%
Noise Injection Rate (Typos)	~8% probability per drug name
Error Types	Swap, Duplicate, Omit

To quantify the performance of the parser and compare the efficiency of different LLM versions, the following statistical metrics were utilized:

Accuracy (%): The percentage of cases where the calculated quality subscore exactly matched the ground truth subscore.

Mean Absolute Error (MAE): Calculated as the average of the absolute differences between the ground truth and the predicted values. It provides a linear measure of the average error magnitude [91].

Root Mean Square Error (RMSE): Used to penalize larger deviations more heavily than smaller ones. A significant gap between RMSE and MAE indicates the presence of larger outlier errors in the extraction process [91].

Spearman’s Rank Correlation Coefficient (ρ): Given that the quality score distribution is non-parametric, this coefficient was employed to assess the strength and direction of the monotonic relationship between the real and estimated risk scores [92, 93].

3.7. Internal Validation

Before public deployment, an internal retrospective validation experiment was conducted to assess the accuracy and clinical reliability of the hybrid text-processing module.

3.7.1. Setting & Data Collection

The data pool consisted of anonymized case descriptions from Karolina Hospital in Hungary (year 2022). Access to this retrospective dataset was authorized by the institutional approval of the Petz Aladár University Teaching Hospital of Győr-Moson-Sopron County (approval number: 9/1841-3/2023). After cleaning data to ensure the presence of both blood count results (for eGFR) and therapy descriptions, a working dataset of 2,999 records was established. From this, a sample of 45 cases was randomly selected, equally distributed across the three MERIS risk categories (low, high, extreme) based on the algorithm’s initial automated assessment.

3.7.2. Evaluation Protocol

Two independent clinical pharmacy experts (Dr. Nóra Gyimesi and Dr. Lóránd Erdélyi) were tasked with evaluating the 45 sample cases. They manually categorized the documented drugs into the MERIS risk classes without using the automated algorithm. Furthermore, an attempt was made to quantify the subjective need for clinical intervention by asking the experts to score the relevance and urgency of required interventions on a four-point scale. A composite subjective score was calculated by multiplying the proposed number of interventions, relevance, and urgency.

3.7.3. Data Analysis

The primary outcome was the correlation between the MERIS scores generated by the AI-powered module and those calculated manually by the experts, analyzed using the Spearman correlation coefficient [92, 93] ($p < 0.05$).

3.8. Prospective Pilot Testing (Web Application)

To test the algorithm’s robustness against highly diverse, global inputs, the text processing and calculator modules were integrated into a web application using the Streamlit framework. A screenshot of the user interface is shown in Figure 7. The application was designed to accept free-text therapy descriptions and eGFR values, returning the calculated MERIS score with a breakdown of the subscores. The application was made publicly accessible to allow healthcare professionals worldwide to test the system with their own data, providing insights into its performance across different documentation styles and drug nomenclatures.

3.8.1. Deployment and Tracking

The application, named AUGMERIS, was made publicly accessible. A secure, private backend database was established to track usage metrics anonymously. The database recorded input values (therapy descriptions and eGFR), the country of access, the return values of the evaluation functions, the recognized drugs marked with their identification methods (local or AI), and the application’s response time.

3.8.2. Data Analysis

Descriptive statistics were used to summarize performance. A two-proportion Z-test was utilized to statistically compare the rate of AI-assisted drug identifications in the global web app dataset against the domestic Hungarian EHR dataset. Additionally, a Spearman rank-order correlation test was conducted to evaluate the consistency of medication usage frequencies between the two datasets.

AUGMERIS calculator [?]

Calculation #389

Renal function (Please enter eGFR value.) [?]

47 ✕ - +

Therapy description (Please enter the names of the active ingredients or the drugs with comma separator. Do not enter anything else.) [?]

Augmentin, Nebivolol, Lantus, Rivotril, Milurit, Laresin, Berodual, Detraleks

OK

Expand the bar to view how the MERIS algorithm operates. ▼



eGFR subscore: 5 -- (depends on kidney function)

Quantity subscore: 5 -- (depends on the number of drugs)

Quality subscore: 6.25 -- (depends on the classification of the drugs)

MERIS score: 16.25 **HIGH RISK** [?]



The response time was 2.99s

Expand the bar to view the medications that the AI has recognized and taken into the calculation. ^

	Name	ATC code	Processing type
1	neбиволol	C07AB12	locally
2	amoxicillin és enzim-inhibitor	J01CR02	locally
3	fenoterol and ipratropium bromide	R03AL01	locally
4	insulin glargine	A10AE04	locally
5	olmesartan medoxomil	C09CA08	locally
6	allopurinol	M04AA01	locally
7	klonazepám	N03AE01	locally
8	Diosmin, combinations	C05CA53	AI

The list of tokens we couldn't use in the calculation:

Figure 7. User interface of the AUGMERIS web application demonstrating the input fields and the generated medication risk score. [54]

3.9. Impact Analysis (Clinical Implementation)

To measure the real-world clinical impact, the system was transferred from a theoretical web tool into an active clinical support system named **PiGuard**, provided by GRNWALtd., an innovative healthcare IT enterprise. A screenshot of the user interface from the demo version of the on-premise PiGuard system is shown in Figure 8. It should be noted that the author is a part-owner of the company, which supplied the software under a free-of-charge loan agreement for the duration of the study. This multicenter, non-interventional clinical study was conducted based on the authorization of the Scientific and Research Ethics Committee of the Medical Research Council (ETT TUKEB) (approval number: BM/8122-1/2025).

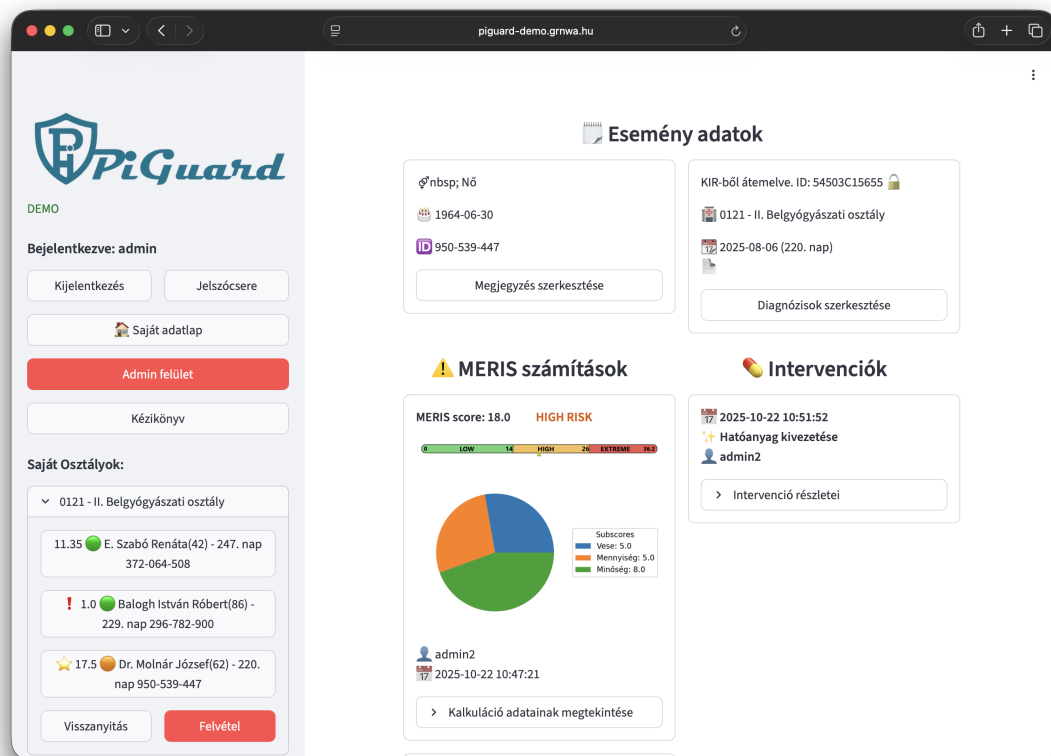


Figure 8. The on-premise PiGuard system interface, illustrating the risk map and the clinical intervention logging features.

3.9.1. Architecture and Deployment

PiGuard was deployed as an "On-Premise" Streamlit application across five different hospital sites. The implementation phase presented substantial technical challenges, as the system had to be independently configured and maintained across six heterogeneous server

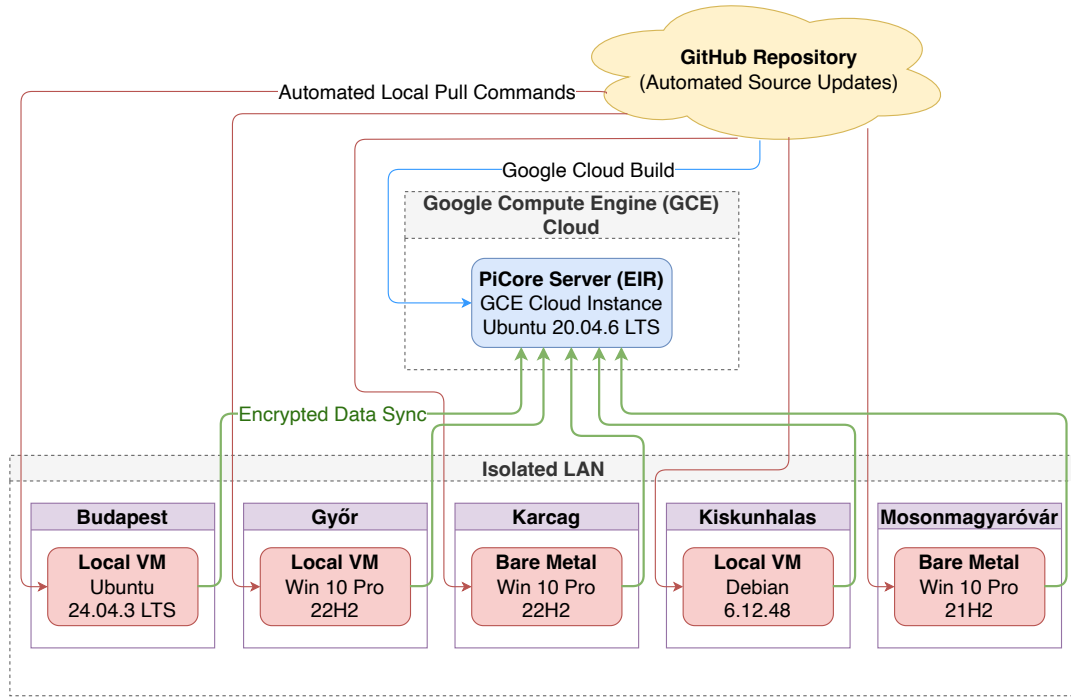


Figure 9. Network topology of the PiGuard system, showcasing the isolated hospital nodes and the central Enterprise Information Repository (EIR) synchronization.

environments (including the central EIR server). The comprehensive network architecture, illustrating the secure connection between the isolated hospital LANs and the central PiCore server, is depicted in Figure 9. As a core security requirement, all hospital-based instances were deployed within isolated internal local area networks (LANs). These nodes are not accessible from the public internet [94, 95]; therefore, I developed and implemented custom automated routines to handle local software updates from the GitHub repository and to manage secure data synchronization with the PiCore server via controlled internal protocols.

The infrastructure reflects a high degree of technical diversity, comprising three Linux-based and three Windows-based environments on physical and virtual machines (VM), as summarized in Table 4. This hybrid architecture required the manual configuration of diverse host types, ranging from dedicated physical hardware to virtualized instances, ensuring platform-independent stability for the clinical decision support modules.

3.9.2. Workflow Integration

The system captures patient therapy and eGFR data, processes the unstructured text utilizing the highly reliable GPT-4.1 model with *Structured Outputs*, computes the MERIS score, and presents it to the clinical pharmacist. The pharmacist then uses the PiGuard

Table 4. Technical specifications and diversity of the deployed server infrastructure

Site / Node Role	Infrastructure	Operating System
Budapest	Local VM Instance	Ubuntu 24.04.3 LTS
Győr	Local VM Instance	Windows 10 Pro 22H2
Karcag	Dedicated Bare Metal	Windows 10 Pro 22H2
Kiskunhalas	Local VM Instance	Debian 6.12.48-1
Mosonmagyaróvár	Dedicated Bare Metal	Windows 10 Pro 21H2
PiCore (EIR Server)	Google Cloud Engine (GCE)	Ubuntu 20.04.6 LTS

interface to document their clinical interventions (e.g., dose modification, drug discontinuation).

3.9.3. Data Aggregation

Anonymized data points — specifically the pre-intervention MERIS score, the standardized intervention type, and the post-intervention MERIS score — are synchronized via a secure, encrypted channel to a central EIR server named PiCore. This allows for performing statistical analysis and calculating the "Delta MERIS" to objectively quantify the risk reduction achieved by pharmacist interventions. The PiCore server also provides a user interface for real-time data visualization, which is shown by the screenshot in Figure 10.

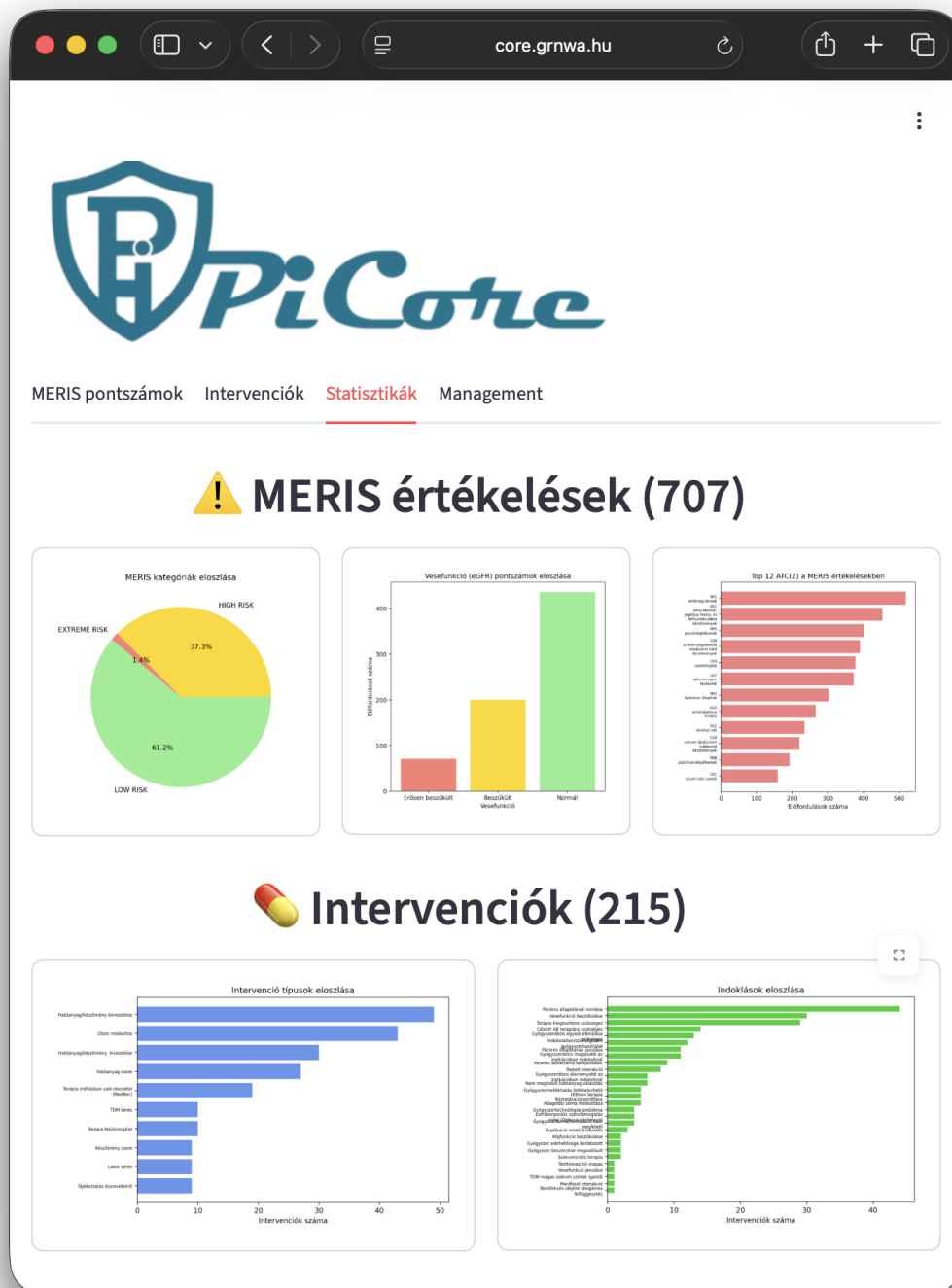


Figure 10. The PiCore Enterprise Information Repository (EIR) server interface, illustrating the centralized data aggregation and analysis features.

4. Results

4.1. In Silico Performance Comparison of LLM Generations

The results of the In Silico validation experiment conducted on the synthetic dataset ($N = 100$) provide quantitative evidence of the performance evolution between successive LLM generations. The primary objective was to assess how the transition from the legacy architecture (GPT-4o-mini) to the updated model (GPT-5.3-mini) affects the accuracy and reliability of the medication risk assessment.

A comprehensive statistical summary of the comparison is presented in Table 5. The data shows a clear trend of improvement across all evaluated metrics.

Table 5. Comparative performance metrics of GPT-4o-mini and GPT-5.3-mini on the synthetic dataset.

Metric	GPT-4o-mini	GPT-5.3-mini	Improvement
N	52	63	+21.1%
Acc (%)	52.0%	63.0%	+11.0%
MAE	0.5725	0.3475	-39.3%
RMSE	0.9457	0.6476	-31.5%
Bias	-0.4975	-0.3025	-39.2%
ρ	0.9424	0.9703	+3.0%
OA (%)	92.98%	95.74%	+3.0%

Abbreviations: N = exact match count; Acc = exact match rate; MAE = mean absolute error; RMSE = root mean square error; Bias = average deviation; ρ = Spearman correlation; OA = overall accuracy.

The transition to the newer model resulted in a substantial reduction in extraction errors. The Mean Absolute Error (MAE) decreased by approximately 39.3%, while the Root Mean Square Error (RMSE) dropped from 0.9457 to 0.6476. The narrower gap between MAE and RMSE in the newer model suggests that GPT-5.3-mini is more effective at suppressing large outlier errors. Furthermore, the exact match rate increased from 52% to 63%, demonstrating significantly higher precision in identifying medication names and dosages.

The visual correlation between the reference ground truth quality subscores and the parser’s calculated values is illustrated in Figure 11.

As shown in the scatter plots, GPT-5.3-mini exhibits a much tighter alignment with the $y = x$ identity line. While a systematic negative bias (underestimation) persists in both models—with 32 cases of underestimation vs. 5 cases of overestimation in the newer model—the magnitude of this bias was reduced by 39.2%. This indicates that while the

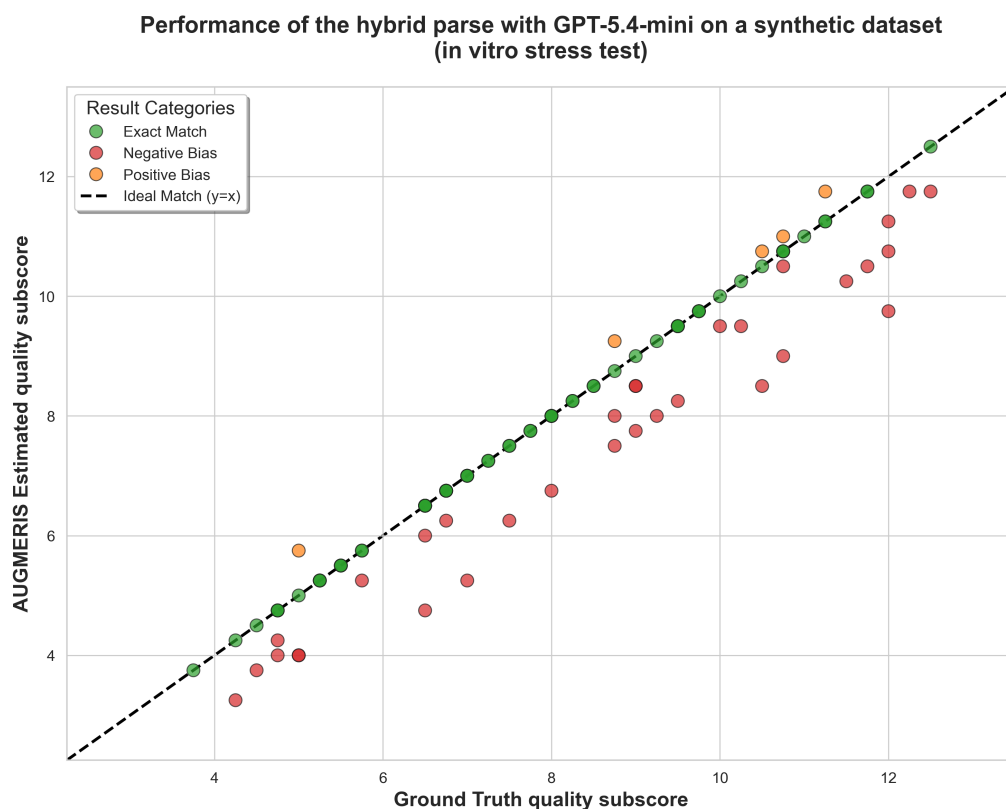
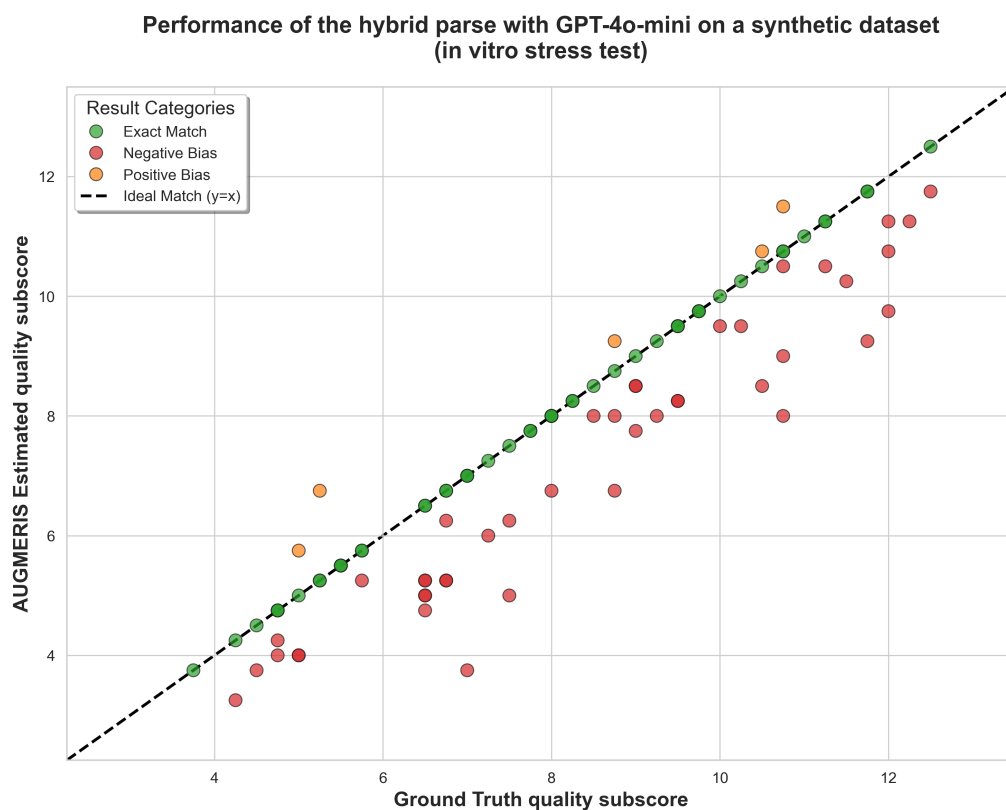


Figure 11. Scatter plots comparing predicted MERIS scores against ground truth values for the legacy (a) and updated (b) models. The tighter clustering around the identity line in (b) reflects the higher Spearman correlation ($\rho = 0.97$).

system remains conservative, the newer generation is far more successful at resolving complex strings and typographical errors.

4.2. Internal Validation of the Text-Processing Module

The validation experiment comparing the AI-driven AUGMERIS calculations to expert manual assessments yielded positive results. Of the 45 randomly selected cases, 11 involved patients who died during hospitalization. The two independent experts handled these terminal cases differently: the first expert processed the last assessable medication regimen before death, while the second expert excluded these cases entirely.

Using Spearman correlation analysis [92, 93], the correlation coefficient between the AI's calculation and the first expert's assessment was 0.463 ($p < 0.05$). For the second expert, the correlation coefficient was 0.496 ($p < 0.05$). While both correlations are moderate, there was no statistically significant difference between the opinions of the two human experts ($p = 0.846$). This indicates that the hybrid algorithm's performance is comparable to the variance seen between human professionals evaluating complex, poor-quality clinical texts.

The secondary hypothesis—that the calculated MERIS value would correlate strongly with the subjective composite score of pharmacist interventions (number $\times$ relevance $\times$ urgency)—could not be confirmed. The subjective nature of the intervention scoring rules proved insufficiently defined to support robust statistical correlation in a retrospective setting.

4.3. Prospective Pilot Testing: Global Usage Data

Following the public release of the AUGMERIS web application, significant international interest was recorded. The tracking database logged trial calculations originating from 15 different countries: Argentina, Brazil, Croatia, Denmark, France, Germany, Hungary, the Netherlands, New Zealand, Norway, Singapore, Slovenia, Spain, Switzerland, and the United Kingdom. Among these, Hungary, New Zealand, and Germany emerged as the most active users, contributing the vast majority of the inputs. Consequently, the detailed performance evaluation was focused on these three primary datasets. The remaining 12 countries contributed supplementary pilot data, ranging from only 1 to 4 requests each.

4.3.1. Risk Distribution and Performance

The average MERIS value across all calculations was 11.87 points. The distribution revealed 228 low-risk, 137 high-risk, and 18 extreme-risk cases. Reduced kidney function was noted in 277 cases, 65 of which were critically low. The average response time of the application was 2.38 seconds. The strict rule-based algorithm handled cases entirely locally in 89 instances (average response time: 0.62s), while the GPT model was invoked in 244 instances (average response time: 3.00s).

4.3.2. AI Recognition Rates

A key finding emerged when comparing the performance of the local Hungarian dictionary against global inputs. In the global web app dataset, the AI was required to identify 25.82% ($N = 2,386$) of the medications due to typos or foreign brand names. The list of the top 15 most frequently identified drug groups (at ATC level 2) is shown in Table 6. In contrast, when processing the domestic Hungarian EHR dataset during internal validation, only 10.34% ($N = 28,269$) of medications required AI assistance. A two-proportion Z-test confirmed this difference is highly statistically significant ($Z = 8.67, p < 0.0001$). Specifically, the AI recognition rates in the most active countries were 20.57% for Hungary, 26.26% for Germany, and 29.64% for New Zealand.

Despite the higher reliance on AI for international data, a Spearman rank-order correlation test comparing the frequency of specific ATC level 2 drug groups between the global and domestic datasets showed a strong, positive, and statistically significant correlation ($\rho = 0.8047, p < 0.0001$), suggesting that global prescribing habits for major disease categories remain relatively consistent.

4.4. Impact Analysis: PiGuard Clinical Implementation

The deployment of the PiGuard system across five clinical sites provided quantifiable evidence of the system's impact on clinical pharmacy workflows. The central EIR server collected data from 703 individual MERIS evaluations and 215 recorded pharmacist interventions.

4.4.1. Risk Distribution and Pharmacist Focus

The automated baseline assessment categorized 61.3% of the evaluated patients into the "low risk" category (mean MERIS score: 12.39, standard deviation: 6.95). However, an analysis of the actual clinical interventions revealed that 54.6% of all pharmacist

Table 6. Top 15 most frequent ATC level 2 groups compared between the domestic (EHR) and global (APP) datasets.

ATC	Group Name	EHR (%)	APP (%)
B01	ANTITHROMBOTIC AGENTS	11.61%	4.60%
A02	DRUGS FOR ACID RELATED DISORDERS	7.98%	4.44%
B05	BLOOD SUBSTITUTES & PERFUSION SOLUTIONS	7.38%	0.24%
J01	ANTIBACTERIALS FOR SYSTEMIC USE	7.09%	4.40%
C03	DIURETICS	6.78%	6.88%
C09	AGENTS ACTING ON RENIN-ANGIOTENSIN SYSTEM	5.64%	7.88%
C07	BETA BLOCKING AGENTS	5.49%	5.64%
A10	DRUGS USED IN DIABETES	4.40%	5.36%
N05	PSYCHOLEPTICS	4.17%	2.84%
A12	MINERAL SUPPLEMENTS	3.59%	1.20%
N02	ANALGESICS	3.30%	9.04%
C01	CARDIAC THERAPY	3.04%	1.48%
C10	LIPID MODIFYING AGENTS	2.87%	6.84%
R03	DRUGS FOR OBSTRUCTIVE AIRWAY DISEASES	2.84%	2.56%
N06	PSYCHOANALEPTICS	2.52%	4.00%
Cumulative representation		78.70%	67.40%

interventions occurred in patient cohorts categorized as "high" or "extreme" risk, as illustrated in Figure 12. This demonstrates that the algorithm successfully highlights the "Hot-Spots"—patients who objectively require the most urgent pharmaceutical care.

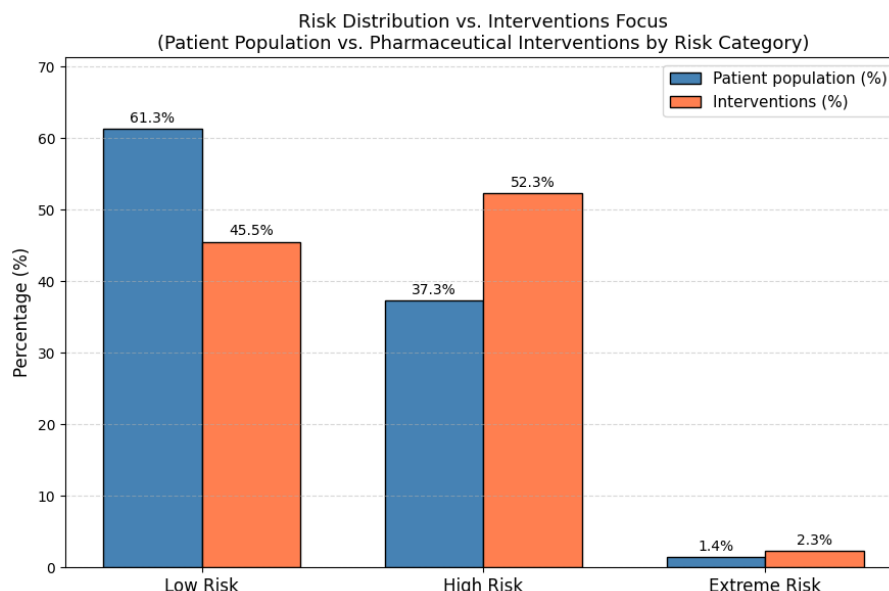


Figure 12. Comparison of patient population risk distribution and clinical pharmacist intervention focus.

4.4.2. Intervention Characteristics

The system successfully linked 82 high-risk events directly to documented pharmacist interventions. For patients with a high *quantity subscore* (polypharmacy) and a high *eGFR subscore* (impaired renal function), the most frequent intervention was "Dose modification." Conversely, for patients with a high *quality subscore* (severe interactions or high-harm drugs), the most frequent interventions were "Discontinuation of active ingredient/product" ($N = 9$) and "Introduction of active ingredient/product" ($N = 11$).

4.4.3. Delta MERIS (Risk Reduction)

The chronological tracking of events (Pre-Intervention MERIS → Pharmacist Intervention → Post-Intervention MERIS) allowed for the calculation of the "Delta MERIS". Initial sequence extractions successfully identified complete chains where pharmacist interventions led directly to measurable risk reduction. For example, in cases where the intervention was "Discontinuation of active ingredient/product", the MERIS scores dropped significantly (e.g., from 23.60 to 17.75, yielding a Delta MERIS of -5.85; and from 23.85 to 18.75, yielding a Delta of -5.10). Overall, in the complete sequences identified, 50.0% of

interventions resulted in an immediate, quantifiable reduction in the calculated medication risk score.

5. Discussion

The primary purpose of this research was to develop, validate, and operationalize the AUGMERIS system as a hybrid medication risk assessment tool. The results across all phases of the study demonstrate that combining deterministic rule-based algorithms with generative AI effectively bridges the gap between unstructured clinical documentation and standardized risk scoring, offering a tangible solution to a persistent healthcare IT problem.

5.1. Performance of the Hybrid Text-Processing Module

Before evaluating the system on real-world clinical data, the In Silico validation using a synthetic dataset ($N = 100$) demonstrated the fundamental robustness of the text-processing architecture under controlled noise conditions. By injecting simulated typographical errors and complex dosage strings, we successfully isolated the parser’s extraction performance from the inherent ambiguity of real EHRs. The comparison between LLM generations revealed a significant evolutionary leap in extraction reliability. The transition from the legacy GPT-4o-mini to the GPT-5.3-mini model yielded a 39.3% reduction in Mean Absolute Error (MAE) and increased the exact match rate from 52% to 63%. Furthermore, the narrowing gap between MAE and RMSE in the newer model indicates enhanced resilience against large outlier errors, proving the system’s capacity to handle human clerical errors gracefully. While a slight systematic conservative bias (underestimation) remained, its magnitude was substantially reduced, resulting in a highly robust Spearman correlation ($\rho = 0.97$). Ultimately, this phase confirmed the module’s safety profile and its capability to resolve complex strings before being exposed to live patient data.

Building on this baseline technical verification, the internal validation experiment confirmed that the hybrid text-processing module performs effectively in practice. The moderate, yet statistically significant correlation between the AI-generated MERIS scores and those calculated manually by clinical pharmacists validates our first hypothesis: AI-assisted text processing can extract actionable data even from the poorest quality EHR sources. The variance observed between the AI and human experts, particularly in complex cases involving prolonged hospitalization and patient death, highlights a fundamental difference in processing logic. While the automated algorithm conservatively evaluated the worst-case parameters recorded during the stay, human experts applied clinical judgment to select the most temporally relevant data. This underscores the premise that AUGMERIS is designed as a decision-support assistant, not an autonomous decision-maker; human oversight remains essential to interpret the clinical timeline.

The failure to establish a correlation between the MERIS value and the retrospective, subjective scoring of intervention necessity (relevance $\times$ urgency) exposed a methodological limitation in the internal validation phase. Retrospectively assigning subjective weights to clinical actions without a standardized framework [96–99] proved unreliable, emphasizing the need for the structured, prospective intervention tracking that was later implemented in the PiGuard phase.

5.2. Insights from Global Pilot Testing

The public deployment of the AUGMERIS web application revealed an unanticipated level of international interest, particularly from Germany and New Zealand. This global engagement confirms that the challenge of identifying high-risk medication profiles in unstructured data is a universal pain point in clinical pharmacy.

The data analytics from the web application provided crucial insights into the behavior of the hybrid model across different linguistic and geographic contexts. The highly significant difference in AI reliance — requiring GPT intervention for 25.82% of global drugs versus only 10.34% domestically — demonstrates that the rule-based engine is heavily constrained by its foundational Hungarian dictionaries (NNGYK and PUPHA). However, it also proves the immense value of the GPT fallback layer. Instead of the system failing when encountering foreign brand names or local abbreviations, the generative AI successfully bridged the gap, allowing the application to function globally. The strong correlation in ATC group frequencies between the datasets further implies that users were testing the application with genuine, guideline-driven clinical scenarios.

5.3. Clinical Impact and the PiGuard System

5.3.1. User Trust, Transparency, and Clinical Impact

Ultimately, the excitement around AI-driven pharmacy tools often overshadows practical concerns regarding user trust and operational adoption rates. Healthcare providers frequently oppose opaque processes that limit oversight [55]. To address this, the AUGMERIS framework encourages clinical engagement through embedded explainable AI, making every identified risk traceable back to specific evidential pathway, whether derived from structured lab results or processed narrative descriptions. Rather than relying purely on a probabilistic confidence score, pharmacists can query both the source text and the applied rules behind each flagged case. This informed oversight strengthens professional accountability while maintaining the workflow speed advantages inherent to automation.

Furthermore, by aligning its architecture with validated MERIS criteria instead of ad hoc AI decisions, AUGMERIS improves regulatory compliance, a crucial factor in moving innovative tools past pilot studies into broader adoption [5]. The balance achieved here between computational scalability and clinician interpretability offers an instructive case study for broader CDSS research seeking similar integration across medical specialties.

5.3.2. Interdisciplinary Integration and Technical Robustness

The transition from a theoretical framework to a functional, multi-site on-premise deployment represents a significant interdisciplinary achievement within this research. While clinical pharmacy and software engineering are traditionally distinct fields, the independent development and management of the PiGuard infrastructure—encompassing heterogeneous server environments and complex network topologies—demonstrates a novel synthesis of these domains.

The necessity of configuring and maintaining six distinct server nodes across varied operating systems (as detailed in Table 4) serves as evidence of the system’s technical robustness and platform-independent design. Crucially, the deployment of hospital-based instances within isolated LANs directly addresses a major barrier in the deployment of healthcare AI: the conflict between cloud-based computational needs and strict patient data privacy requirements [94, 95].

By ensuring that these nodes remain inaccessible from the public internet, the architecture effectively minimizes the risk of data leaks while maintaining high-level clinical decision support functionality. This approach aligns with current research advocating for transparent, source-controlled, and data-secure medical AI implementations [33]. This self-managed technical lifecycle—from initial Python-based algorithmic design to the physical orchestration of on-premise servers—underscores the scalability of the AUGMERIS framework and its readiness for sustainable, real-world practice.

5.3.3. PiGuard Deployment and Efficacy

The transition from a theoretical web calculator to the active, on-premise PiGuard system represents the most significant practical achievement of this research. By deploying the application within the secure LAN environments of five hospitals, the critical hurdle of healthcare data privacy was successfully overcome.

The impact analysis results clearly demonstrate the system’s clinical utility. While 62% of the total hospital population was flagged as low risk by the MERIS algorithm, 54% of actual pharmacist interventions were concentrated in the high and extreme risk categories. This finding validates our core hypothesis: integrating a risk map into routine workflows

successfully directs pharmacists' limited time toward the patients who need it most, thereby increasing overall intervention efficiency. Furthermore, the PiGuard system's ability to track chronological sequences and calculate the "Delta MERIS" provides, for the first time, an objective, numerical metric to prove the efficacy of clinical pharmacist interventions. A reduction in a patient's MERIS score directly correlates to a decrease in polypharmacy, the avoidance of severe drug interactions, or the adjustment of dosages to match renal function, all of which fundamentally improve patient safety.

5.4. Strengths and Limitations

5.4.1. Architectural Advantages and Structural Strengths

The primary strength of the AUGMERIS architecture lies in its flexibility and adaptability to legacy EHR systems. By accepting free-text inputs and correcting typos and varied nomenclatures on the fly, it dramatically lowers the barrier to user adoption. Furthermore, the localized deployment model of PiGuard ensures strict General Data Protection Regulation (GDPR) compliance while still enabling centralized, anonymized research. Crucially, AUGMERIS mitigates the inherent risks of generative models by filtering their outputs through the strict MERIS criteria coded into its rule engine. This approach curtails risky extrapolations and implicitly reinforces the system's role as a clinical assistant rather than an autonomous decision-maker, ensuring that a professional's competence is always required to act upon AI-suggested risks.

5.4.2. Methodological and Operational Limitations

Despite these structural advantages, deploying this architecture into live healthcare environments poses distinct challenges. A primary limitation is the heavy reliance on input quality [3, 5, 20]; poor EHR documentation inevitably undermines model accuracy regardless of algorithmic sophistication. For instance, refining NLP of medical narratives containing localized idioms or obscure abbreviations remains a persistent hurdle [3].

Furthermore, generative AI models like GPT-4.1 are inherently probabilistic. While adept at inferring missing details, this "fill-in" behavior carries a residual risk of "hallucination" generating speculative elements that do not reflect reality [100]. Although the implementation of Structured Outputs successfully eliminated formatting inconsistencies and forced the model to reliably return valid ATC code structures, the semantic risk of misclassification remains particularly acute when processing rare or highly specialized medication names that are not well-represented in the training data. Consequently, AI-generated results must always be treated as supportive rather than definitive. Finally, the current web application

data lacks clinical ground truth; it reflects usage interaction patterns rather than verified clinical outcomes.

5.4.3. Navigating Rapid Technological Evolution in Medical AI

A significant methodological takeaway from this research is the profound challenge of developing clinical software in an era of unprecedented, rapid technological evolution. Over the three-year course of this project, the underlying Generative AI landscape transformed fundamentally, dictating continuous architectural pivots.

In the early phases, developing the text-processing pipeline required extensive, heuristic "Prompt Engineering" techniques. We had to construct complex instructions to coerce models like GPT-3.5 and the initial GPT-4o into acting as strict, deterministic ATC classifiers. Despite these efforts, the models remained inherently probabilistic, leading to frequent JSON formatting errors and hallucinations that disrupted the automated workflow.

However, within a matter of months, the introduction of API-level constraints—specifically OpenAI's *Structured Outputs* feature in the GPT-4.1 and the subsequent *mini* models—obviated the need for these fragile workarounds. By guaranteeing schema compliance at the API level, a problem that previously required complex software engineering was solved natively by the service provider.

This rapid paradigm shift highlights a critical lesson for future medical AI research: clinical software architectures must be designed with extreme modularity. Developers must anticipate the continuous, almost monthly obsolescence and replacement of foundational AI models. What requires extensive algorithmic compensation today may become a native, guaranteed API feature tomorrow. Therefore, hybrid systems like AUGMERIS, which decouple the strict clinical logic from the interchangeable language model layer, offer the most sustainable path forward in this volatile technological environment.

5.5. International Overview and Parallel Experiences

To fully evaluate the significance and real-world applicability of the AUGMERIS system, it must be situated within the broader context of international research on Artificial Intelligence in clinical pharmacy. As highlighted by comprehensive reviews in the international literature, integrating AI into CDSS is a global priority, yet it is fraught with persistent challenges [3, 5].

The international consensus confirms that the heterogeneity and low quality of unstructured EHRs constitute the primary bottleneck for automated risk scoring globally [5]. Purely deterministic systems, while compliant and reproducible, fail to capture critical

clinical nuances buried in free-text notes. Conversely, unconstrained probabilistic models (such as standalone Large Language Models) are highly susceptible to hallucinations and lack the necessary guardrails for high-stakes medical decision-making [6]. The hybrid architecture developed in the AUGMERIS project, which embeds generative NLP capabilities within the strict deterministic bounds of the MERIS criteria, directly addresses this universal challenge. This alignment is strongly supported by the global pilot testing data presented in this dissertation. The fact that the GPT-based fallback layer was required for 25.82% of international inputs (compared to 10.34% domestically) mathematically proves that the necessity for adaptive, hybrid text-processing is not a local anomaly, but a fundamental requirement for the global scalability of clinical AI tools.

Furthermore, international research consistently emphasizes that clinical acceptance of AI tools is heavily dependent on explainability, workflow integration, and the mitigation of automation bias [5]. Clinicians are reluctant to adopt "black-box" solutions where the rationale for a risk alert cannot be traced. AUGMERIS operationalizes these international mandates through the PiGuard interface, which provides dual provenance for every alert. By displaying both the triggered deterministic rule and the parsed narrative excerpt, the system fosters trust and ensures that the AI acts strictly as a clinical assistant rather than an autonomous arbiter. The quantifiable reduction in risk scores (Delta MERIS) achieved during the PiGuard deployment further corroborates international findings that integrating risk stratification directly into pharmacist workflows significantly optimizes intervention efficiency and resource allocation.

5.5.1. Comparing AUGMERIS with the ScreenGPT Project

Beyond clinical decision support for healthcare professionals, the implementation of LLMs in healthcare can be conceptually divided into two distinct operational poles: the normalization of complex inputs (Business-to-Business (B2B)/clinical focus) and the translation of medical knowledge into accessible outputs (Business-to-Consumer (B2C)/patient focus). To further contextualize the application of LLMs in healthcare, it is instructive to compare the clinician-facing AUGMERIS system with patient-facing preventive tools. The key differences and shared characteristics are summarized in the visual matrix presented in Table 7.

During my doctoral research, I participated as the lead programmer in the development of the ScreenGPT project, which provides an excellent parallel for this comparison [42].

ScreenGPT is an AI-driven tool designed to support the primary and secondary prevention of cervical cancer through empathetic, patient-facing communication. A pilot study involving 112 participants demonstrated that while general awareness of LLMs is high, their application for health screening requires careful validation of the users' health literacy [42].

Table 7. Visual matrix: comparison of AUGMERIS and ScreenGPT projects

Feature	AUGMERIS	ScreenGPT
Target Audience	Professionals (B2B)	Patients/Laypersons (B2C)
Direction	Input Normalization	Output Translation
Objective	Risk Stratification	Prevention & Education
Validation	Objective (Delta MERIS)	Subjective (User Experience)
Core Tech	Hybrid (Rules + GPT)	Generative Conversational
Safety	Deterministic Guardrails	Professional Validation

Both AUGMERIS and ScreenGPT share a fundamental technological foundation: they leverage the advanced language comprehension capabilities of GPT to navigate complex medical domains. Moreover, both projects recognize the inherent risks of generative model, specifically hallucinations and automation bias, and emphasize that AI must serve as a supportive tool rather than a replacement for professional healthcare.

However, the two projects address diametrically opposed communication bottlenecks in healthcare, revealing significant methodological and operational differences:

Target Audience and Objective: AUGMERIS is a Business-to-Business (B2B) clinical tool designed for highly trained healthcare professionals (clinical pharmacists and physicians). Its objective is to prioritize high-risk patients and prevent acute adverse drug events in a hospital setting. In contrast, ScreenGPT is a Business-to-Consumer (B2C) tool aimed at laypersons, focusing on primary and secondary prevention (e.g., lifestyle choices and screening awareness) prior to the onset of acute illness.

Data Processing Direction: AUGMERIS operates on the *input side* of healthcare data. It ingests messy, unstructured, and low-quality EHR notes and normalizes them into structured, quantifiable risk variables (MERIS scores) for professional decision-making. ScreenGPT operates on the *output side*. It takes structured clinical guidelines and complex medical protocols and translates them into accessible, empathetic, and personalized educational narratives for patients.

Validation and Metrics: The validation of AUGMERIS relies on objective clinical parameters. The correlation with expert pharmacist assessments, adherence to deterministic pharmacological rules, and the calculation of measurable risk reduction (Delta MERIS). The validation of ScreenGPT relied primarily on user-reported experience. Furthermore, two independent medical practitioners assessed the system’s responses to the 30 most frequently asked questions for clinical accuracy [42].

In conclusion, the parallel development and understanding of both input-processing clinical tools (AUGMERIS) and output-generating educational tools (ScreenGPT) demonstrate a comprehensive approach to modernizing patient safety and preventive care. While they serve different endpoints of the healthcare spectrum, both projects validate the international consensus: the safe integration of Artificial Intelligence in medicine requires strict methodological guardrails, domain-specific tuning, and an unwavering focus on enhancing human decision-making.

5.6. Professional Recognition and Industry Feedback

The practical innovation and clinical relevance of the AUGMERIS framework were further validated beyond the academic and clinical spheres through significant industry recognition. In March 2026, the AUGMERIS calculator was awarded the professional Grand Prize—the Strix Award—at Hungary’s first comprehensive corporate artificial intelligence competition during the AI Summit [101, 102].

The professional jury identified the system as one of the year’s most defining AI-based developments, specifically highlighting its contribution to patient safety through the automated analysis of medication risks. This recognition is particularly noteworthy as the entry documentation was based on the original AUGMERIS web application, emphasizing that the underlying hybrid text-processing model provides a robust solution for extracting precise information from unstructured and often incomplete clinical narratives. The award also reflects the success of the multi-institutional collaboration between Semmelweis University, Hospital Ltd., Széchenyi István University, and Aarhus University, confirming that the system bridges the gap between complex data science and everyday clinical practice.

5.7. Future Directions

Building on the clinical validation and the professional recognition described above, future development will follow two main strategic paths: technical refinement and systemic integration.

5.7.1. Algorithmic and Data Expansion

Future work will focus on expanding the core dictionaries of the strict rule-based algorithm. By integrating broader international databases, such as the WHO International Nonproprietary Name (INN) lists and specialized repositories like Drugs.com, the system’s reliance on generative AI API calls can be further reduced. This will not only increase

processing speed and reduce operational costs but also further mitigate the inherent risks of large language model hallucinations.

5.7.2. System Integration and Scalability

The successful deployment of the PiGuard infrastructure has paved the way for deeper integration into comprehensive Hospital Information Systems (HIS). The objective is to transition from a standalone tool to a standard, Software as a Service (SaaS)-based module within routine clinical pharmacy workflows. By utilizing the "Delta MERIS" metric as a key performance indicator, the system can provide continuous, objective proof of risk reduction, facilitating its adoption in diverse healthcare environments globally.

5.7.3. Commercialization and Technology Transfer

Beyond technical and systemic advancements, strategic steps are being taken toward the commercial utilization of the AUGMERIS system. Initial discussions have been conducted with the Centre of Science & Innovations Vice-Rector and Business Development at Semmelweis University to outline the framework for technology transfer. To facilitate market entry and broader clinical adoption, the establishment of a dedicated project company (spin-off) is currently in the planning phase. This new enterprise will be founded through a joint ownership structure comprising Semmelweis University, Hospitali Kft., and the principal investigator. The financial and operational framework, including profit sharing, will be managed on an equity basis, ensuring a balanced and mutually beneficial partnership to drive the global implementation of the system.

6. Conclusions

The research presented in this dissertation successfully navigated the complex intersection of data science and clinical pharmacy practice to address a critical gap in medication safety. The primary achievement is the development and validation of the AUGMERIS system, which successfully automates the application of the Danish MERIS algorithm.

Furthermore, a rigorous In Silico stress-test using a synthetic dataset demonstrated the parser's exceptional structural reliability and error-tolerance. Facing simulated typographical errors and unstructured dosage instructions, the upgraded LLM architecture achieved a 95.7% accuracy and a 0.97 correlation with ground-truth values, proving that the underlying data-extraction logic is highly robust before being exposed to real-world clinical variables.

By engineering a novel, hybrid text-processing architecture that combines the strict reliability of national drug databases with the adaptive power of GPT, we overcame the persistent challenge of parsing unstructured, low-quality clinical documentation. The internal validation demonstrated that this AI-assisted data extraction performs with an accuracy comparable to that of human clinical professionals.

The public deployment of the AUGMERIS web application confirmed a significant global demand for objective, automated risk stratification tools. It also highlighted the necessity of flexible NLP models, as the generative AI seamlessly compensated for international variations in drug nomenclature that traditional, static databases could not handle.

Most importantly, the deployment of the secure, On-Premise PiGuard system across multiple hospital sites provided concrete evidence of the tool's clinical impact. The system effectively functions as a "Hot-Spot" detector, successfully directing clinical pharmacists toward high-risk patients. The tracking of "Delta MERIS" scores provides objective proof that pharmacist interventions, guided by this decision support system, result in immediate and quantifiable reductions in medication-related risks.

Ultimately, AUGMERIS represents a scalable, highly adaptable solution that enhances the efficiency of clinical pharmacists and significantly improves patient safety. Through continuous algorithmic refinement and broader integration into hospital information systems, it possesses the potential to become a fundamental component of modern, data-driven pharmacotherapy management.

7. Summary

Medication-related harm is a major healthcare challenge, often exacerbated by the fact that vital patient data is buried in unstructured, free-text EHRs. This dissertation introduces AUGMERIS, an AI-supported CDSS designed to automate the calculation of MERIS criteria. The primary objective was to overcome the limitations of traditional deterministic systems by utilizing a hybrid architecture that combines strict pharmacological rules with the adaptive NLP capabilities of GPT.

The research progressed through multiple phases. Initially, a three-step text-processing pipeline was engineered, which processes clinical narratives. This module was retrospectively validated against expert clinical pharmacist assessments. Subsequently, the system was deployed globally as a Streamlit-based web application to test its robustness across diverse linguistic inputs. Finally, an on-premise version of the system was implemented across five hospital sites to evaluate its real-world clinical impact.

Internal validation demonstrated that the hybrid AI algorithm performs with an accuracy comparable to human experts on poor-quality EHR notes. The global pilot test revealed a universal need for adaptive NLP, as the generative AI fallback was required for far more international inputs. Most significantly, the clinical deployment of PiGuard proved the system's operational value. Although only 38% of the evaluated patients were categorized as high or extreme risk, this cohort accounted for 54% of all pharmacist interventions, confirming the tool's efficacy as a "Hot-Spot" detector.

To contextualize the system's clinical focus on input normalization, the dissertation also contrasts AUGMERIS with ScreenGPT, a patient-facing preventive AI tool. While ScreenGPT relies on empathetic output generation and subjective health literacy metrics, AUGMERIS is anchored in deterministic clinical guardrails to strictly mitigate LLM hallucinations. Ultimately, this research validates that integrating hybrid, explainable AI into routine clinical pharmacy workflows significantly optimizes resource allocation, enhances professional accountability, and provides a scalable, data-driven solution for improving global medication safety.

8. References

- [1] P. C. N. E. Association. *PCNE Classification For Drug-Related Problems*. Report. Pharmaceutical Care Network Europe, 2020. URL: https://www.pcne.org/upload/files/417_PCNE_classification_V9-1_final.pdf (visited on 04/16/2026).
- [2] T. I. P. F. (FIP). *Medication Review And Medicines Use Review: A Toolkit For Pharmacists*. Report. The International Pharmaceutical Federation, 2022. URL: <https://www.fip.org/file/5100> (visited on 04/16/2026).
- [3] M. Elhaddad and S. Hamam. “AI-Driven Clinical Decision Support Systems: An Ongoing Pursuit Of Potential”. In: *Cureus* 16.4 (2024), e57728. DOI: 10.7759/cureus.57728.
- [4] G. Golden, C. Popescu, S. Israel, K. Perlman, C. Armstrong, R. Fratila, M. Tanguay-Sela, and D. Benrimoh. “Applying Artificial Intelligence To Clinical Decision Support In Mental Health: What Have We Learned?” In: *Health Policy And Technology* 13.2 (2024), p. 100844. DOI: 10.1016/j.hlpt.2024.100844.
- [5] F. Magrabi, E. Ammenwerth, J. Brender McNair, N. F. De Keizer, H. Hyppönen, P. Nykänen, M. Rigby, P. J. Scott, T. Vehko, Z. S.-Y. Wong, and A. Georgiou. “Artificial Intelligence In Clinical Decision Support: Challenges For Evaluating AI And Practical Implications”. In: *Yearb Med Inform* (2019), pp. 128–134. DOI: 10.1055/s-0039-1677903.
- [6] A. Haim, M. M. Katson, M. Cohen-Shelly, S. M. Peretz, D. P. Aran, and S. M. Shelly. “Evaluating GPT -4 As A Clinical Decision Support Tool In Ischemic Stroke Management”. In: (2024). DOI: 10.1101/2024.01.18.24301409.
- [7] H.-Q. See, J.-N. Chan, S.-J. Ling, S.-C. Gan, C.-O. Leong, and C.-W. Mai. “Advancing Pharmacy Service Using Big Data – Are We Fully Utilising The Big Data’s Potential Yet?” In: *Journal Of Pharmacy & Pharmaceutical Sciences* 21.1 (2018), pp. 217–221. DOI: 10.18433/jpps29869.
- [8] S.-F. Chen, S. Loguercio, K.-Y. Chen, S. E. Lee, J.-B. Park, S. Liu, H. J. Sadaei, and A. Torkamani. “Artificial Intelligence For Risk Assessment On Primary Prevention Of Coronary Artery Disease”. In: *Current Cardiovascular Risk Reports* 17 (2023), pp. 215–231. DOI: 10.1007/s12170-023-00731-4.
- [9] A. Choudhury. “Factors In Influencing Clinicians ’ Willingness To Use An AI-based Clinical Decision Support System”. In: *Frontiers In Digital Health* 4 (2022), p. 920662. DOI: 10.3389/fdgth.2022.920662.

- [10] B. Bajgain, D. Lorenzetti, J. Lee, and K. Sauro. “Determinants Of Implementing Artificial Intelligence-based Clinical Decision Support Tools In Healthcare: A Scoping Review Protocol”. In: *British Medical Journal Open* 13 (2023), e068373. DOI: 10.1136/bmjopen-2022-068373.
- [11] A. G. S. 2. B. C. U. E. Panel, D. M. Fick, T. P. Semla, J. Beizer, N. Brandt, R. Dombrowski, C. E. DuBeau, W. Eisenberg, J. J. Epplin, N. Flanagan, et al. “American Geriatrics Society 2015 Updated Beers Criteria For Potentially Inappropriate Medication Use In Older Adults”. In: *Journal Of The American Geriatrics Society* 63.11 (2015), pp. 2227–2246. DOI: 10.1111/jgs.13702.
- [12] D. O’Mahony, A. Cherubini, A. R. Guiteras, M. Denking, J.-B. Beuscart, G. Onder, A. Gudmundsson, A. J. Cruz-Jentoft, W. Knol, G. Bahat, et al. “STOPP/START Criteria For Potentially Inappropriate Prescribing In Older People: Version 3”. In: *European Geriatric Medicine* 14.4 (2023), pp. 625–632. DOI: 10.1007/s41999-023-00777-y.
- [13] N. Kumar, C. B. Knowler, D. Strumpman, and B. Bajorek. “Facilitating Medication Misadventure Risk Assessment In The Emergency Medical Unit”. In: *Journal Of Pharmacy Practice And Research* 41 (2 2011), pp. 108–112. DOI: 10.1002/j.2055-2335.2011.tb00675.x.
- [14] A. E. M. J. H. Linkens, V. Milosevic, P. H. M. van der Kuy, V. H. Damen-Hendriks, C. Mestres Gonzalvo, and K. P. G. M. Hurkens. “Medication-related Hospital Admissions And Readmissions In Older Patients: An Overview Of Literature”. In: *International Journal Of Clinical Pharmacy* 42.5 (2020), pp. 1243–1251. DOI: 10.1007/s11096-020-01040-1.
- [15] S. Walker, N. McAuslane, L. Liberti, J. Leong, and S. Salek. “A Universal Framework For The Benefit-Risk Assessment Of Medicines: Is This The Way Forward?” In: *Therapeutic Innovation & Regulatory Science* 49.1 (2015), pp. 17–25. DOI: 10.1177/2168479014547421.
- [16] M. Christensen and A. Lundh. “Medication Review In Hospitalised Patients To Reduce Morbidity And Mortality”. In: *Cochrane Database Of Systematic Reviews* 2 (2016). DOI: 10.1002/14651858.CD008986.pub3.
- [17] T. Graabæk and L. J. Kjeldsen. “Medication Reviews By Clinical Pharmacists At Hospitals Lead To Improved Patient Outcomes: A Systematic Review”. In: *Basic & Clinical Pharmacology & Toxicology* 112.6 (2013), pp. 359–373. DOI: 10.1111/bcpt.12062.

- [18] J. Gallagher, S. McCarthy, and S. Byrne. “Economic Evaluations Of Clinical Pharmacist Interventions On Hospital Inpatients: A Systematic Review Of Recent Literature”. In: *International Journal Of Clinical Pharmacy* 36 (2014), pp. 1101–1114. DOI: 10.1007/s11096-014-0008-9.
- [19] S. Simoens. “Factors Affecting The Cost Effectiveness Of Antibiotics”. In: *Chemotherapy Research And Practice* 2011.1 (2011), p. 249867. DOI: 10.1155/2011/249867.
- [20] H. C. Çubukçu, D. İ. Topcu, and S. Yenice. “Machine Learning-based Clinical Decision Support Using Laboratory Data”. In: *Clinical Chemistry And Laboratory Medicine* 62.5 (2024), pp. 793–823. DOI: 10.1515/cc1m-2023-1037.
- [21] S. V. S. Rana. “Metamorphosis Of Human Health Risk Assessment With Artificial Intelligence”. In: *Journal Of Environmental Biology* 45.1 (2024), pp. vii–viii. DOI: 10.22438/jeb/45/1/Ed-3.
- [22] N. Gray, H. Page, I. Buchan, and D. W. Joyce. “Risk And Uncertainty Communication In Deployed AI-based Clinical Decision Support Systems: A Scoping Review”. In: (2024). DOI: 10.1101/2024.12.06.24318489.
- [23] Z. Huang, H. Y.-F. Lim, J. T. Ow, S. H.-L. Sun, and A. Chow. “Doctors’ Perception On The Ethical Use Of AI-enabled Clinical Decision Support Systems For Antibiotic Prescribing Recommendations In Singapore”. In: *Frontiers In Public Health* 12 (2024), p. 1420032. DOI: 10.3389/fpubh.2024.1420032.
- [24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. “Attention Is All You Need”. In: *Advances In Neural Information Processing Systems* 30 (2017). DOI: 10.48550/arXiv.1706.03762.
- [25] C. Manning and H. Schütze. *Foundations Of Statistical Natural Language Processing*. MIT press, 1999.
- [26] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer. “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, pp. 252–259.
- [27] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang. “Squad: 100,000+ Questions For Machine Comprehension Of Text”. In: *Proceedings Of The 2016 Conference On Empirical Methods In Natural Language Processing*. 2016, pp. 2383–2392.
- [28] D. Jurafsky and J. H. Martin. *Speech And Language Processing: An Introduction To Natural Language Processing, Computational Linguistics, And Speech Recognition With Language Models*. 3rd. 2026.

- [29] B. Meskó. “The Impact Of Multimodal Large Language Models On Health Care’s Future”. In: *J Med Internet Res* 25 (2023), e52865. ISSN: 1438-8871. DOI: 10.2196/52865.
- [30] H. Xiao, F. Zhou, X. Liu, T. Liu, Z. Li, X. Liu, and X. Huang. “A Comprehensive Survey Of Large Language Models And Multimodal Large Language Models In Medicine”. In: *Information Fusion* 117 (2025), p. 102888. DOI: 10.1016/j.inffus.2024.102888.
- [31] J. Li, A. Dada, B. Puladi, J. Kleesiek, and J. Egger. “ChatGPT In Healthcare: A Taxonomy And Systematic Review”. In: *Computer Methods And Programs In Biomedicine* 245 (2024), p. 108013. DOI: 10.1016/j.cmpb.2024.108013.
- [32] H. Kim, H. Hwang, J. Lee, S. Park, D. Kim, T. Lee, C. Yoon, J. Sohn, J. Park, O. Reykhart, et al. “Small Language Models Learn Enhanced Reasoning Skills From Medical Textbooks”. In: *NPJ Digital Medicine* 8.1 (2025), p. 240. DOI: 10.1038/s41746-025-01653-8.
- [33] S. Griewing, F. Lechner, N. Gremke, S. Lukac, W. Janni, M. Wallwiener, U. Wagner, M. Hirsch, and S. Kuhn. “Proof-of-concept Study Of A Small Language Model Chatbot For Breast Cancer Decision Support—a Transparent, Source-controlled, Explainable And Data-secure Approach”. In: *Journal Of Cancer Research And Clinical Oncology* 150.10 (2024), p. 451.
- [34] M. Magnini, G. Aguzzi, and S. Montagna. “Open-source Small Language Models For Personal Medical Assistant Chatbots”. In: *Intelligence-based Medicine* 11 (2025), p. 100197. DOI: 10.1016/j.ibmed.2024.100197.
- [35] F. Wang, Z. Zhang, X. Zhang, Z. Wu, T. Mo, Q. Lu, W. Wang, R. Li, J. Xu, X. Tang, et al. “A Comprehensive Survey Of Small Language Models In The Era Of Large Language Models: Techniques, Enhancements, Applications, Collaboration With LLMs, And Trustworthiness”. In: *ACM Transactions On Intelligent Systems And Technology* 16 (6 2025). DOI: 10.1145/3768165.
- [36] A. E. Weidmann and E. W. Watson. “Novel Opportunities For Clinical Pharmacy Research: Development Of A Machine Learning Model To Identify Medication Related Causes Of Delirium In Different Patient Groups”. In: *International Journal Of Clinical Pharmacy* 46.4 (2024), pp. 992–995. DOI: 10.1007/s11096-024-01707-z.
- [37] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Nee-lakantan, P. Shyam, G. Sastry, A. Askell, et al. “Language Models Are Few-shot Learners”. In: *Advances In Neural Information Processing Systems* 33 (2020), pp. 1877–1901. DOI: 10.48550/arXiv.2005.14165.

- [38] T. D. Maarseveen, T. Meinderink, M. J. Reinders, J. Knitza, T. W. Huizinga, A. Kleyer, D. Simon, E. B. van den Akker, and R. Knevel. “Machine Learning Electronic Health Record Identification Of Patients With Rheumatoid Arthritis: Algorithm Pipeline Development And Validation Study”. In: *JMIR Medical Informatics* 8.11 (2020), e23930. DOI: 10.2196/23930.
- [39] S. I. Pitts, N. Barasch, A. T. Maslen, B. A. Thomas, L. P. Dorissaint, K. G. Decker, S. Kazi, Y. Yang, and A. R. Chen. “Understanding CancelRx: Results Of End-to-End Functional Testing, Proactive Risk Assessment, And Pilot Implementation”. In: *Applied Clinical Informatics* 10 (2019), pp. 336–347. DOI: 10.1055/s-0039-1688698.
- [40] S. Arain, R. Thalapparambath, and F. H. Al Ghamdi. “COVID-19 Pandemic: Response Plan By The Johns Hopkins Aramco Healthcare Inpatient Pharmacy Department”. In: *Research In Social And Administrative Pharmacy* 17 (2021), pp. 2009–2011. DOI: 10.1016/j.sapharm.2020.05.016.
- [41] J. Ayers, A. Poliak, M. Dredze, E. Leas, Z. Zhu, J. Kelley, D. Faix, A. Goodman, C. Longhurst, M. Hogarth, and D. Smith. “Comparing Physician And Artificial Intelligence Chatbot Responses To Patient Questions Posted To A Public Social Media Forum”. In: *Jama Internal Medicine* 183 (6 2023), p. 589. DOI: 10.1001/jamainternmed.2023.1838.
- [42] V. Angyal, **Á. Bertalan**, P. Domján, and E. Dinya. “Exploring The Possibilities And Limitations Of Customized Large Language Model To Support And Improve Cervical Cancer Screening”. In: *BMC Medical Informatics And Decision Making* 25.1 (2025), p. 242. DOI: 10.1186/s12911-025-03088-3.
- [43] S. Epelbaum, Y. M. Saade, C. F. Roze, E. Roze, S. Ferrieux, C. Arbizu, M. Nogues, C. Azuar, B. Dubois, S. T. d. Montcel, and M. Teichmann. “A Reliable And Rapid Language Tool For The Diagnosis, Classification, And Follow-up Of Primary Progressive Aphasia Variants”. In: *Frontiers In Neurology* 11 (2021). DOI: 10.3389/fneur.2020.571657.
- [44] N. Yuan, A. Oesterle, P. Botting, S. Chugh, C. Albert, J. Ebinger, and D. Ouyang. “High-Throughput Assessment Of Real-World Medication Effects On QT Interval Prolongation: Observational Study”. In: *JMIR Cardio* 7 (2023), p. 1. DOI: 10.2196/41055.
- [45] W. H. Organization. *Medication Without Harm: Policy Brief*. World Health Organization, 2024.

- [46] M. Pirmohamed, S. James, S. Meakin, C. Green, A. K. Scott, T. J. Walley, K. Farrar, B. K. Park, and A. M. Breckenridge. “Adverse Drug Reactions As Cause Of Admission To Hospital: Prospective Analysis Of 18 820 Patients”. In: *British Medical Journal (BMJ)* 329.7456 (2004), pp. 15–19. DOI: 10.1136/bmj.329.7456.15.
- [47] D. Formica, J. Sultana, P. Cutroneo, S. Lucchesi, R. Angelica, S. Crisafulli, Y. Ingrassiotta, F. Salvo, E. Spina, and G. Trifirò. “The Economic Burden Of Preventable Adverse Drug Reactions: A Systematic Review Of Observational Studies”. In: *Expert Opinion On Drug Safety* 17.7 (2018), pp. 681–695. DOI: 10.1080/14740338.2018.1491547.
- [48] E. A. Saedder, B. Brock, L. P. Nielsen, D. K. Bonnerup, and M. Lisby. “Identifying High-risk Medication: A Systematic Literature Review”. In: *European Journal Of Clinical Pharmacology* 70.6 (2014), pp. 637–645. DOI: 10.1007/s00228-014-1668-z.
- [49] E. A. Saedder, B. Brock, L. P. Nielsen, D. K. Bonnerup, and M. Lisby. “Classification Of Drugs With Different Risk Profiles”. In: *Dan Med J* 62.8 (2015), A5118. URL: <https://pubmed.ncbi.nlm.nih.gov/26239589/>.
- [50] T. W. Thøgersen, E. A. Saedder, and M. Lisby. “Is A High Medication Risk Score Associated With Increased Risk Of 30-Day Readmission? A Population-Based Cohort Study From CROSS-TRACKS”. In: *Journal Of Patient Safety* 18.4 (2022), e714–e721. DOI: 10.1097/PTS.0000000000000939.
- [51] O. Uemura, H. Yokoyama, K. Ishikura, Y. Gotoh, H. Satō, H. Sugiyama, M. Honda, and S. Matsuo. “Performance In Adolescents Of The Two Japanese Serum Creatinine Based Estimated Glomerular Filtration Rate Equations, For Adults And Paediatric Patients: A Study Of The Japan Renal Biopsy Registry And Japan Kidney Disease Registry From 2007 To 2013”. In: *Nephrology* 22 (6 2017), pp. 494–497. DOI: 10.1111/nep.12982.
- [52] A. S. Levey, L. A. Stevens, C. H. Schmid, Y. Zhang, A. F. Castro III, H. I. Feldman, J. W. Kusek, P. Eggers, F. Van Lente, T. Greene, et al. “A New Equation To Estimate Glomerular Filtration Rate”. In: *Annals Of Internal Medicine* 150.9 (2009), pp. 604–612. DOI: 10.7326/0003-4819-150-9-200905050-00006.
- [53] K. K. Viktil and H. S. Blix. “The Impact Of Clinical Pharmacists On Drug-related Problems And Clinical Outcomes”. In: *Basic & Clinical Pharmacology & Toxicology* 102.3 (2008), pp. 275–280. DOI: 10.1111/j.1742-7843.2007.00206.x.

- [54] Á. Bertalan, V. Angyal, P. Domján, E. A. Sædder, G. Király, L. Erdélyi, N. Gyimesi, and D. Elek. “Development Of An Artificial Intelligence Powered Medication Risk Score Calculator Application”. In: *Basic & Clinical Pharmacology & Toxicology* (2025), e70109. DOI: 10.1111/bcpt.70109.
- [55] N. A. Aziz, A. Manzoor, M. D. M. Qureshi, M. A. Qureshi, and W. Rashwan. “Explainable AI In Healthcare: Systematic Review Of Clinical Decision Support Systems”. In: (2024). DOI: 10.1101/2024.08.10.24311735.
- [56] T. Richards and A. Treuille. *Streamlit For Data Science: Create Interactive Data Apps In Python*. Expert insight. Packt Publishing, Limited, 2023. ISBN: 9781803248226.
- [57] P. Singh and H. Singh. *Generative AI With LangChain: Build Smart AI Apps Using LangChain And Python Tools (English Edition)*. BPB Publications, 2025. ISBN: 9789365896497.
- [58] K. Høj, H. S. Pedersen, A. S. B. Lundberg, F. Bro, L. P. Nielsen, and E. A. Sædder. “External Validation Of The Medication Risk Score In Polypharmacy Patients In General Practice: A Tool For Prioritizing Patients At Greatest Risk Of Potential Drug-related Problems”. In: *Basic & Clinical Pharmacology & Toxicology* 129.4 (2021), pp. 319–331. DOI: 10.1111/bcpt.13636.
- [59] S. Bird, E. Klein, and E. Loper. *Natural Language Processing With Python: Analyzing Text With The Natural Language Toolkit*. Sebastopol, CA: O’Reilly Media, 2025.
- [60] G. Van Rossum and F. L. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2025.
- [61] Streamlit. *Streamlit Documentation*. Report. Streamlit Inc., 2025. URL: <https://docs.streamlit.io/en/stable/> (visited on 04/16/2026).
- [62] OpenAI. *OpenAI API Reference*. 2025. URL: <https://platform.openai.com/docs/api-reference/introduction> (visited on 04/16/2026).
- [63] Apache Software Foundation. *Apache HTTP Server Version 2.4 Documentation*. 2024. URL: <https://httpd.apache.org/docs/2.4/> (visited on 03/10/2026).
- [64] World Wide Web Consortium. *Cascading Style Sheets (CSS) Snapshot 2023*. W3C Working Group Note. W3C, 2023. URL: <https://www.w3.org/TR/css-2023/> (visited on 03/10/2026).
- [65] python-visualization. *Folium Documentation*. Tech. rep. Folium, 2025. URL: <https://python-visualization.github.io/folium/latest/> (visited on 02/23/2026).

- [66] G. D. Team. *GeoPandas Documentation*. Tech. rep. GeoPandas, 2025. URL: <https://geopandas.org/en/stable/docs.html> (visited on 02/23/2026).
- [67] S. Chacon and B. Straub. *Pro Git*. 2nd. Apress, 2014.
- [68] Git Development Community. *Git: Distributed Version Control System*. 2024. URL: <https://git-scm.com/> (visited on 04/18/2026).
- [69] GitHub, Inc. *GitHub Docs - Official Product Documentation*. 2024. URL: <https://docs.github.com/> (visited on 03/10/2026).
- [70] Python Software Foundation. *Hashlib — Secure Hashes And Message Digests*. 2024. URL: <https://docs.python.org/3/library/hashlib.html> (visited on 03/10/2026).
- [71] WHATWG (Web Hypertext Application Technology Working Group). *HTML Living Standard*. 2024. URL: <https://html.spec.whatwg.org/multipage/> (visited on 03/10/2026).
- [72] S. Faulkner, A. Eicholz, T. Leithead, A. Danilo, and S. Moon. *HTML 5.1 2Nd Edition*. W3C Recommendation. W3C, 2017. URL: <https://www.w3.org> (visited on 03/10/2026).
- [73] MDN Web Docs. *JavaScript - Dynamic Client-side Scripting*. 2024. URL: <https://developer.mozilla.org> (visited on 03/10/2026).
- [74] Python Software Foundation. *Json — JSON Encoder And Decoder*. 2024. URL: <https://docs.python.org/3/library/json.html> (visited on 03/15/2024).
- [75] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grollenberg, S. Corlay, et al. “Jupyter Notebooks-a Publishing Format For Reproducible Computational Workflows.” In: *ELPUB*. 2016, pp. 87–90. DOI: 10.3233/978-1-61499-649-1-87.
- [76] J. Hunter, D. Dale, E. Firing, M. Droettboom, and Matplotlib development team. *Matplotlib 3.9.2 Documentation*. 2024. URL: <https://matplotlib.org> (visited on 03/10/2026).
- [77] Oracle Corporation. *MySQL 8.0 Reference Manual*. 2024. URL: <https://dev.mysql.com> (visited on 03/10/2026).
- [78] C. R. Harris, K. J. Millman, S. J. van der Walt, et al. “Array Programming With NumPy”. In: *Nature* 585.7825 (2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

- [79] The pandas development team. *Pandas-dev/Pandas: Pandas 2.1.3*. 2023. DOI: 10.5281/zenodo.10101831. URL: <https://pandas.pydata.org> (visited on 03/10/2026).
- [80] A. Clark and Pillow Contributors. *Pillow (PIL Fork) 11.0.0 Documentation*. 2024. URL: <https://pillow.readthedocs.io> (visited on 03/10/2026).
- [81] S. Colvin, E. Jolibois, H. Ramezani, A. Garcia Badaracco, T. Dorsey, D. Montague, S. Matveenko, M. Trylesinski, S. Runkle, D. Hewitt, A. Hall, and V. Plot. *Pydantic Validation*. 2024. URL: <https://docs.pydantic.dev/latest/> (visited on 03/15/2026).
- [82] P. S. Foundation. *Python Documentation*. 2025. URL: <https://www.python.org/doc/> (visited on 04/16/2026).
- [83] M. L. Waskom. “Seaborn: Statistical Data Visualization”. In: *Journal Of Open Source Software* 6.60 (2021), p. 3021. DOI: 10.21105/joss.03021.
- [84] Microsoft. *Visual Studio Code Documentation - Jupyter Notebooks In VS Code*. 2024. URL: <https://code.visualstudio.com/docs/datascience/jupyter-notebooks> (visited on 04/18/2026).
- [85] F. Pérez and B. E. Granger. “IPython: A System For Interactive Scientific Computing”. In: *Computing In Science & Engineering* 9.3 (2007), pp. 21–29. DOI: 10.1109/MCSE.2007.53.
- [86] **Á. Bertalan**, L. Erdélyi, and E. Dinya. “AI Nyelvi Modellek Alkalmazása Szöveges Kórlapok Feldolgozására A Gyógyszerelés Kockázati Értékelésének Vizsgálatához”. In: *Orvosi Informatika. A XXXVI. Neumann Kollokvium Konferencia-kiadványa* (2023), pp. 1–6. DOI: 10.5281/zenodo.10429438.
- [87] E. A. Saedder, M. Lisby, L. P. Nielsen, J. Rungby, L. V. Andersen, D. K. Bonnerup, and B. Brock. “Detection Of Patients At High Risk Of Medication Errors: Development And Validation Of An Algorithm”. In: *Basic & Clinical Pharmacology & Toxicology* 118.2 (2016), pp. 143–149. DOI: 10.1111/bcpt.12473.
- [88] M. A. D. Gunadi. “Random Forest-based Compound Atc Classification Using Structural And Physiochemical Information”. In: *American Journal Of Biomedical Science And Research* 17 (3 2022), pp. 16–22. DOI: 10.34297/ajbsr.2022.17.002344.
- [89] National Institute of Pharmacy and Nutrition. *Gyógyszeradatbázis*. 2025. URL: <https://ogyei.gov.hu/gyogyszeradatbazis/> (visited on 02/23/2025).

- [90] National Health Insurance Fund of Hungary. *Publikus Gyógyszertörzs*. 2025. URL: https://www.neak.gov.hu/felso_menu/szakmai_oldalak/gyogyszer_segedeszkoz_gyogyfurdo_tamogatas/egeszsegugyi_vallalkozasoknak/pupha/Publikus_Gyogyszertorzs (visited on 02/23/2025).
- [91] T. Chai and R. R. Draxler. “Root Mean Square Error (RMSE) Or Mean Absolute Error (MAE)? – Arguments Against Avoiding RMSE In The Literature”. In: *Geoscientific Model Development* 7.3 (2014), pp. 1247–1250. DOI: 10.5194/gmd-7-1247-2014.
- [92] C. Spearman. “The Proof And Measurement Of Association Between Two Things”. In: *International Journal Of Epidemiology* 39.5 (2010), pp. 1137–1150. ISSN: 0300-5771. DOI: 10.1093/ije/dyq191.
- [93] E. Dinya. *Biometria Az Orvosi Gyakorlatban (Biometrics In Medical Practice)*. Budapest: Medicina Könyvkiadó Zrt., 2001, p. 484.
- [94] A. Grünig, J. Kriebel, J. Varghese, T. Herrmann, S. Sandmann, and C. Bruns. “Implementation And User Evaluation Of An On-Premise Large Language Model In A German University Hospital Setting: Cross-Sectional Survey”. In: *JMIR AI* 5 (2026), e84362. ISSN: 2817-1705. DOI: 10.2196/84362.
- [95] P. K. Barman, R. Rathour, P. Mishra, and R. Patel. “A Privacy-Preserving AI Symptom Checker Using A Self-Hosted LLM For Basic Healthcare Guidance”. In: *TechRxiv* 2025.1022 (2025). DOI: 10.36227/techrxiv.176114174.43989310/v1.
- [96] M. Bouzeid, J. Clarenne, C. Mongaret, H. Pluchart, S. Chanoine, C. Vermorel, P. Bedouch, M. Bourdelin, B. Charpiat, O. Conort, J. Gravoulet, A. Janoly-Dumenil, M. Juste, J.-L. Bosson, and S. VIP–Act-IP©group. “Using National Data To Describe Characteristics And Determine Acceptance Factors Of Pharmacists’ interventions: A Six-year Longitudinal Study”. In: *International Journal Of Clinical Pharmacy* 45.2 (2023), pp. 430–441. DOI: 10.1007/s11096-022-01526-0.
- [97] A. Ihbe-Heffinger, C. Langebrake, C. Hohmann, K. Leichenberg, H. Hilgarth, M. Kunkel, M. Lueb, and T. Schuster. “Prospective Survey-based Study On The Categorization Quality Of Hospital Pharmacists’ interventions Using DokuPIK”. In: *International Journal Of Clinical Pharmacy* 41.2 (2019), pp. 414–423. DOI: 10.1007/s11096-019-00785-8.
- [98] K. A. Maes, R. M. Tremp, G. W. group on clinical pharmacy <http://www.gsasa.ch/seiten/aktivitaeten/klinische-aktivitaeten/>, K. E. Hersberger, and M. L. Lampert. “Demonstrating The Clinical Pharmacist’s Activity: Validation Of An Intervention

- Oriented Classification System”. In: *International Journal Of Clinical Pharmacy* 37 (2015), pp. 1162–1171. DOI: 10.1007/s11096-015-0179-z.
- [99] T. H. Vo, T. L. Hoang, E. M. Faller, and D. T. Nguyen. “Development And Validation Of The Vi-Med® Tool For Medication Review”. In: *Journal Of Applied Pharmaceutical Science* 10.2 (2020), pp. 086–096. DOI: 10.7324/JAPS.2020.102014.
- [100] J. Waldo and S. Boussard. “GPTs And Hallucination: Why Do Large Language Models Hallucinate?” In: *Queue* 22.4 (2024), pp. 19–33. ISSN: 1542-7730. DOI: 10.1145/3688007.
- [101] V. Zákányi. *Átadták A STRIX AI Award Díjait: Ezek Lettek 2025 Legjobb Vállalati AI-projektjei*. 2026. URL: <https://gyartastrend.hu/cikk/atadtak-a-strix-ai-award-dijait-ezek-lettek-2025-legjobb-vallalati-ai-projektjei> (visited on 05/04/2026).
- [102] PHARMINDEX Online. *Szakmai Nagydíjat Nyert A Gyógyszerbiztonságot Fokozó Úttörő Magyar MI-alkalmazás*. 2026. URL: <https://www.pharmindex-online.hu/> (visited on 04/17/2026).

9. Bibliography of the Candidate's Publications

9.1. Publications Related to the Dissertation

Á. Bertalan, V. Z. Angyal, P. Domján, E. A. Sædder, G. Király, L. Erdélyi, N. Gyimesi, and E. Dinya. “Development of an Artificial Intelligence Powered Medication Risk Score Calculator Application”. In: *BASIC & CLINICAL PHARMACOLOGY & TOXICOLOGY* 137 (2025). ISSN: 1742-7835. DOI: 10.1111/bcpt.70109. URL: <https://m2.mtmt.hu/api/publication/36346440>.

(IF: 3.6)

Á. Bertalan, L. Erdélyi, and E. Dinya. “AI nyelvi modellek alkalmazása szöveges kórlapok feldolgozására a gyógyszerelés kockázati értékelésének vizsgálatához (Application of AI Language Models for Processing Textual Medical Records to Investigate Medication Risk Assessment)”. In: *Orvosi informatika. A XXXVI. Neumann Kollokvium konferencia-kiadványa* (2023), pp. 1–6. DOI: 10.5281/zenodo.10429438. URL: <https://m2.mtmt.hu/api/publication/34519507>.

Á. Bertalan, V. Z. Angyal, G. Király, and E. Dinya. “AUGMERIS a kibővített gyógyszerelés-kockázati értékelő applikáció (AUGMERIS the Augmented Medication Risk Score Calculator Application)”. In: (2024), pp. 221–224. URL: <https://m2.mtmt.hu/api/publication/37411047>.

Á. Bertalan and E. Dinya. “A klinikai gyógyszerész szerepe az elektronikus gyógyszeradminisztrációban (The Role of the Clinical Pharmacist in Electronic Medication Administration)”. In: *IME* 23 (2024), pp. 5–9. ISSN: 1588-6387. DOI: 10.53020/IME-2024-101. URL: <https://m2.mtmt.hu/api/publication/34854816>.

V. Z. Angyal, Á. Bertalan, P. Domján, and E. Dinya. “ScreenGPT – A mesterséges intelligencia alkalmazásának lehetőségei és korlátai a primer, szekunder és terciér prevencióban (ScreenGPT – Opportunities and Limitations of Artificial Intelligence Applications in Primary, Secondary, and Tertiary Prevention)”. In: *ORVOSI HETILAP* 165 (2024), pp. 629–635. ISSN: 0030-6002. DOI: 10.1556/650.2024.33029. URL: <https://m2.mtmt.hu/api/publication/34817394>.

(IF: 1.0)

V. Z. Angyal, Á. Bertalan, P. Domján, and E. Dinya. “CN10 Evaluating prompt engineering techniques for artificial intelligence-generated cervical cancer screening information: A comparative content analysis”. In: *ANNALS OF ONCOLOGY* 36 (2025), S1518–S1519. ISSN: 0923-7534. DOI: 10.1016/j.annonc.2025.08.1086. URL: <https://m2.mtmt.hu/api/publication/36704634>.

V. Z. Angyal, Á. Bertalan, and E. Dinya. “A ScreenGPT rendszer megbízhatóságának vizsgálata méhnyakrák prevencióval kapcsolatos információátadás támogatásában (Investigating the Reliability of the ScreenGPT System in Supporting Information Delivery Related to Cervical Cancer Prevention)”. In: *V. Életmódorvosi és II. Mozgás Receptre Konferencia*. 2025, pp. 6–7. URL: <https://m2.mtmt.hu/api/publication/36381127>.

V. Z. Angyal, Á. Bertalan, P. Domján, H. J. Feith, and E. Dinya. “A generatív mesterséges intelligencia használatának vizsgálata a primer és szekunder egészségügyi prevencióban Magyarországon (Investigating the Use of Generative Artificial Intelligence in Primary and Secondary Health Prevention in Hungary)”. In: *ORVOSI HETILAP* 167 (2026), pp. 628–636. ISSN: 0030-6002. DOI: 10.1556/650.2026.33534. URL: <https://m2.mtmt.hu/api/publication/37025022>.

(IF: 1.0)

V. Z. Angyal, Á. Bertalan, P. Domján, and E. Dinya. “Exploring the possibilities and limitations of customized large language model to support and improve cervical cancer screening”. In: *BMC MEDICAL INFORMATICS AND DECISION MAKING* 25 (2025). ISSN: 1472-6947. DOI: 10.1186/s12911-025-03088-3. URL: <https://m2.mtmt.hu/api/publication/36226867>.

(IF: 5.5)

Related IF 11,1

9.2. Other Publications

Á. Bertalan, L. Erdélyi, and E. Dinya. “Az elektronikus betegdokumentáció klinikai gyógyszerészeti vonatkozásai (Clinical Pharmacy Aspects of Electronic Health Records)”. In: (2022), pp. 20–24. URL: <https://m2.mtmt.hu/api/publication/37411073>.

V. Z. Angyal, Á. Bertalan, P. Domján, H. J. Feith, and E. Dinya. “Development of a Questionnaire for Assessing the Use of ChatGPT in Primary and Secondary Disease Prevention”. In: *FRONTIERS IN PUBLIC HEALTH* 13 (2026). DOI: 10.3389/fpubh.2025.1709611. URL: <https://m2.mtmt.hu/api/publication/36831463>.

(IF: 4.1)

Á. Bertalan. “Gyógyszerészek lehetőségei egy kisebb városi kórházban (Opportunities for Pharmacists in a Smaller City Hospital)”. In: *FOCUS MEDICINAE* 21 (2019), pp. 33–34. ISSN: 1419-0478. URL: <https://m2.mtmt.hu/api/publication/37413687>.

Á. Bertalan, L. Erdélyi, G. Király, and E. Dinya. “Elektronikus lázlap és betegszintű gyógyszerelés a Karolina Kórházban (Electronic Health Record and Patient-Level Medication Management at the Karolina Hospital)”. In: *GYÓGYSZERÉSZET* 68 (2024), pp. 86–91. ISSN: 0017-6036. URL: <https://m2.mtmt.hu/api/publication/37411082>.

Not Related IF 4,1

$\sum$ IF: 15,2

9.3. Book

E. Dinya, Z. Tóth, Z. Sándor, and Á. Bertalan. *BIOMETRIA A KLINIKUMBAN: Python programozási nyelv és alkalmazása (Biometry in the Clinic: Python Programming Language and its Application) /under release/*. in print. Medicina Kiadó, 2026.

9.4. Conferences

Á. Bertalan, B. Hankó, and S. Sebbők. “Helyes gyógyszeralkalmazás gyógyszerészi minimálintervenciós modellje (Correct Drug Use with a Pharmacist-Led Minimal Intervention Model)”. In: *LIV. Rozsnyay Mátyás Emlékverseny*. 2019.

Á. Bertalan, S. Gábor, and M. Knausz. “Antibiotikum Stewardship tevékenység megvalósítása egy városi kórházban (Implementation of Antibiotic Stewardship Activities in a City Hospital)”. In: *Magyar Infektológiai és Klinikai Mikrobiológiai Társaság 47. Kongresszusa*. 2019.

Á. Bertalan and I. Bertalan. “Hogyan tervezzünk intelligens betegbiztonsági rendszert? (How to Design an Intelligent Patient Safety System?)” In: *Congressus Pharmaceuticus Hungaricus*. 2020.

Á. Bertalan, L. Erdélyi, and E. Dinya. “Az elektronikus betegdokumentáció klinikai gyógyszerészeti vonatkozásai (Clinical Pharmacy Aspects of Electronic Health Records)”. In: *XXXV. Neumann Kollokvium*. 2022.

Á. Bertalan and E. Dinya. “The Retrospective Evaluation of the Pharmaceutical Therapies at Karolina Hospital via Text Mining Technology”. In: *PhD Scientific Days*. 2023.

Á. Bertalan. “AI egészségügyi szempontból (AI from a Healthcare Perspective)”. In: *Kórházi Gyógyszerészek XXIV. Kongresszusa*. 2023.

Á. Bertalan, L. Erdélyi, and E. Dinya. “AI nyelvi modellek alkalmazása szöveges kórlapok feldolgozásában gyógyszerelés kockázati értékelés céljából (Application of AI Language Models in Processing Text-Based Medical Records for Medication Risk Assessment)”. In: *XXXVI. Neumann Kollokvium*. 2023.

Á. Bertalan. “LEHETŐSÉGEK ÉS KOCKÁZATOK A MESTERSÉGES INTELLIGENCIA EGÉSZSÉGÜGYI ALKALMAZÁSÁBAN (OPPORTUNITIES AND RISKS IN THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN HEALTHCARE)”. In: *Magyar Egészségügyi Menedzser Klub évváró konferenciája*. 2023.

Á. Bertalan and E. Dinya. “Elektronikus lázlap és betegszintű gyógyszerelés a Karolina Kórházban (Electronic Health Record and Patient-Level Medication Management at the Karolina Hospital)”. In: *LVIII. Rozsnyay Mátyás Emlékverseny*. 2023.

Á. Bertalan, L. Erdélyi, and E. Dinya. “A dán gyógyszerelés kockázati index számítása mesterséges intelligencia alkalmazásával az egészségügyi dokumentumok feldolgozásában (Calculating the Danish Medication Risk Index with Artificial Intelligence in Processing Healthcare Documents)”. In: *Semmelweis Symposium*. 2023.

Á. Bertalan, G. Király, V. Angyal, and E. Dinya. “Development of a Web Application for MERIS evaluation from real-world data”. In: *PhD Scientific Days*. 2024.

Á. Bertalan, V. Angyal, G. Király, and E. Dinya. “AUGMERIS a kibővített gyógyszerelés-kockázati értékelő applikáció (AUGMERIS the Augmented Medication Risk Score Calculator Application)”. In: *XXXVII. Neumann Kollokvium*. 2024.

Á. Bertalan, L. Erdélyi, V. Angyal, P. Domján, and E. Dinya. “A klinikai gyógyszerészeti intervenciók és a Medication Risk Score (MERIS) értékek korrelációja (Correlation of Clinical Pharmacy Interventions and Medication Risk Score (MERIS) Values)”. In: *Kórházi Gyógyszerészek XXV. Kongresszusa*. 2025.

L. Erdélyi, Á. Bertalan, V. Angyal, P. Domján, and E. Dinya. “Gyakorlatorientált klinikai gyógyszerészeti intervenció rögzítő rendszer a PAEOK-ban (Practice-Oriented Clinical Pharmacy Intervention Recording System in the PAEOK)”. In: *Kórházi Gyógyszerészek XXV. Kongresszusa*. 2025.

10. Acknowledgements

I am profoundly grateful to my supervisor, Prof. Dr. Elek Dinya, for his firm guidance and for consistently keeping my attention anchored on the most critical focal steps of this research.

I would like to express my gratitude to Gyula Király, Managing Director of Hospitaly Ltd., who helped me navigate the previously uncharted territory of the IT industry by serving as both my employer and a professional and ethical mentor.

Special thanks go to my fellow researcher, Viola Angyal, who guided me through the bureaucratic hurdles of the doctoral education journey with great dedication. I also thank her for her valuable insights regarding the AUGMERIS project and for providing me with the opportunity to gain invaluable experience by involving me in the ScreeGPT project.

I am thankful to my colleague, Dr. Lóránd Erdélyi, PhD, for his words of encouragement, our inspiring collaborative thinking, and our joint efforts as co-founders of GRNWA Ltd.

I would like to acknowledge Péter Domján for his supportive words, for setting a positive example with his unbreakable perseverance, and for his contributions to our numerous joint publications.

I am deeply indebted to Director General Dr. Júlia Kinga Jenei, whose understanding and support were essential in harmonizing my responsibilities as a chief pharmacist with my research duties.

I also wish to thank the clinical pharmacist colleagues across the five research sites for their enthusiastic participation in the data collection process.

I would like to express my appreciation to all the staff members of Semmelweis University as well. Their helpful and cooperative attitude facilitated the smooth resolution of numerous administrative and practical challenges, ranging from securing necessary signatures to executing research-related procurements.

I left it for last, but I would like to express my deepest gratitude to my family, especially to my wife, for her endless patience, unwavering support, and remarkable willingness to compromise throughout this arduous journey.

Funding

The research was supported by the Ministry of Culture and Innovation from the National Research, Development and Innovation Fund, within the framework of the Cooperative Doctoral Programme (KDP-2023) of the National Research, Development and Innovation Office.