UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

HOCHSCHULE HEILBRONN

(172393) Research Project

Realistische Daten für ein lebendiges Lehre-KIS

Robin Hefner[1]

Florian Hauptmann[2]

14. Juni 2026

Eingereicht bei Prof. Dr. Petra Knaup

[1]206488, rohefner@stud.hs-heilbronn.de
[2]225690, fhauptmann@stud.hs-heilbronn.de

Inhaltsverzeichnis

Abkürzungsverzeichnis

AI: Artificial Intelligence

CGM: Continuous Glucose Monitors

CIEL: Columbia International eHealth Laboratory

CSV: Comma-separated values

EHR: Electronic Health Record

EMR: Electronic Medical Record

FHIR: Fast Healthcare Interoperability Resources

GAN: Generative Adversarial Network

GDPR: General Data Protection Regulation

HIPAA: Health Insurance Portability and Accountability Act

HIS: Hospital Information System

HL7: Health Level Seven International

IT: Information Technology

JSON: JavaScript Object Notation

LLM: Large Language Model

PDF: Portable Document Format

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyzes

REST: Representational State Transfer

SQL: Structured Query Language

UI: User Interface

XML: Extensible Markup Language

Abbildungsverzeichnis

Tabellenverzeichnis

Realistic data for a dynamic teaching hospital information system

Robin Hefner[1], Florian Hauptmann[1]

[1]Institute of Medical Informatics, Heidelberg University, Heidelberg, Germany;

Abstract – Background: Operational Hospital Information Systems (HIS) are essential for modern medical and medical informatics training. However, traditional educational HIS environments face legal and structural limitations, often relying on unpopulated databases or static datasets that fail to reflect dynamic, longitudinal clinical workflows.
Objective: Within the Heidelberg "studiKIS" project, this research project aims to develop and systematically evaluate a methodology for providing realistic, privacy-compliant, dynamic data, including synthetic patient records and continuous wearable sensor streams, for a dedicated educational HIS.
Methods: A literature review on PubMed and IEEE Xplore (March 2026) identified state-of-the-art frameworks and integration standards. Core quality criteria were defined through six qualitative interviews with medical and IT experts. Prototypically, rule-based (Synthea) and Large Language Model-based (Gemini 3.1 Pro) pipelines were engineered alongside a temporal ingestion script to simulate continuous hospital workflows in the open-source HIS Bahmni. Continuous glucose monitoring sensor streams were integrated via HL7 FHIR. Finally, 120 blinded patient records across three modalities (Real clinical data from MIMIC-III, Synthea, and LLM) were evaluated by clinical experts using four criteria on a 5-point Likert scale and thematic qualitative analysis.
Results: Synthea-generated records achieved significantly higher ratings in longitudinal consistency (c2) compared to LLM profiles ($p<0.001$). No statistically significant differences were found for clinical plausibility (c1), chronological sequence (c3), and overall impression (c4). Qualitative analysis revealed distinct source-specific deficits: real data suffered from poor readability and chronological compression, Synthea datasets exhibited repetitive clinical patterns, and LLM data suffered from factual hallucinations and guideline violations. Empirical sensor streams were visualized as trend graphs, though dynamic zooming and event annotation are natively lacking.
Conclusion: Providing dynamic, "living" clinical data via HL7 FHIR pipelines offers meaningful benefits over static teaching methods. Rule-based frameworks excel in life-span consistency but introduce rigid stereotypes, whereas LLMs generate flexible but often hallucinated records. Future work should focus on scalable translation scripts for rule-based systems and knowledge-grounded AI architectures (such as RAG and multi-agent critic frameworks) to eliminate clinical inaccuracies.

Keywords - Hospital Information System; HIS; Synthetic Clinical Data; Electronic Health Record; HL7 FHIR; Synthea; Large Language Models; Medical Sensors; Medical Education

I INTRODUCTION

Hospital Information Systems (HIS) are the core infrastructure of the modern public health system. Accordingly, the development of a suitable educational HIS for academic training in human medicine and medical informatics is essential. To create a fully functional and realistic educational HIS, the project investigates the generation of structured, longitudinal synthetic clinical data alongside the integration of continuous sensor data streams from medical wearables.

The ability to proficiently navigate and use operational HIS is a fundamental competency in modern medical education (Bichel-Findlay et al., 2023). As healthcare becomes increasingly digitized, students must not only understand traditional clinical workflows but also learn to manage heterogeneous data formats, use various technical interfaces and evaluate the growing role of patient-generated health data in the treatment context (Ferreira et al., 2025). Practical, hands-on experience with realistic, evolving patient histories and continuous physiological monitoring is therefore important to effectively prepare future physicians for the complexities of modern, data-driven patient care (Mohan et al., 2016).

Despite this educational necessity, traditional training environments face significant structural and legal limita-

tions. The widespread use of authentic electronic health records (EHRs) is strictly restricted by data privacy regulations such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) (R. J. Chen et al., 2021). Consequently, academic HIS often rely on unpopulated databases or static synthetic datasets (Choi et al., 2021). These static approaches represent merely a cross-section at a specific point in time, failing to adequately reflect the dynamic, longitudinal nature of clinical workflows or the continuous evolution of clinical parameters required to analyze disease progressions (Majumder et al., 2017).

To bridge this pedagogical gap, this project aims to develop a methodology for providing dynamic data for a dedicated educational HIS within the Heidelberg "studiKIS" project. The primary objective is to synthetically generate structured clinical data that continuously evolves over time, enabling students to observe and analyze realistic, longitudinal treatment trajectories. Furthermore, the project integrates real-world sensor data to teach students how to process diverse data formats and technical interoperability standards, such as HL7 FHIR. By engineering this dynamic data pipeline and systematically evaluating its clinical plausibility, the project yields a privacy-compliant, highly realistic simulation environment that bridges the gap between theoretical knowledge and practical clinical application.

II BACKGROUND

Virtual Hospital Information System

In academic medical and medical informatics education, students require hands-on experience with operational system environments (Bichel-Findlay et al., 2023). Traditional educational HIS often rely on empty databases or static datasets, which do not reflect realistic clinical workflows (Choi et al., 2021). A dynamic virtual HIS addresses this gap by continuously simulating routine operations (Medlock et al., 2023). The progressive injection of patient transitions, diagnostic orders, and documentation updates provides a realistic environment for training in real-time data analysis, system troubleshooting, and workflow integration (Mohan et al., 2016).

Bahmni

Bahmni is an open-source Hospital Information System and Electronic Medical Record (EMR). Originally designed for low-resource environments (Syzdykova et al., 2017), it is increasingly used in educational and research settings (Medlock et al., 2023). It integrates open-source components such as OpenMRS for clinical data management, Odoo for inventory and billing, and dcm4chee as a Picture Archiving and Communication System (Syzdykova et al., 2017). In the Heidelberg "studiKIS" project, Bahmni serves as the central infrastructure to teach clinical workflows and data architectures.

Synthetic Clinical Data

Synthetic clinical data provides an alternative to real patient records, mitigating privacy constraints imposed by regulations such as the GDPR and the HIPAA (R. J. Chen et al., 2021). By replicating the structural and statistical properties of EHRs, synthetic datasets facilitate software testing, algorithm benchmarking, and education without re-identification risks (Walonoski et al., 2018). However, conventional synthetic datasets are often static and lack the temporal dynamics of real clinical pathways (Dahmen & Cook, 2019).

Synthea

Synthea is an open-source, rule-based simulation framework for generating synthetic, longitudinal medical data. It models patient lifecycles using discrete-state-machine modules based on public health statistics, clinical guidelines, and demographic data. Synthea outputs standard-compliant records, primarily in the HL7 FHIR format. Its default configuration models the North American healthcare system. (Walonoski et al., 2018)

Generative AI and Large Language Models in EHR Synthesis

Rule-based systems provide deterministic control over clinical workflows but require extensive manual modeling and lack the variance of real-world clinical narratives. Large Language Models (LLMs) offer a probabilistic approach to synthesize unstructured and semi-structured clinical data (Thirunavukarasu et al., 2023). LLMs can generate contextual case vignettes and longitudinal records. Key challenges include mitigating algorithmic hallucinations and ensuring schema compliance (Agrawal et al., 2022; Thirunavukarasu et al., 2023).

Interoperability Standards and Semantic Terminologies

Healthcare interoperability requires standards for data exchange between heterogeneous systems. The Health Level Seven (HL7) Fast Healthcare Interoperability Resources (FHIR) standard provides a RESTful framework using structured JSON or XML (Mandel et al., 2016). Semantic interoperability requires mapping these structures to standardized clinical ontologies (Lehne et al., 2019).

Controlled vocabularies, such as the Columbia International eHealth Laboratory (CIEL) concept dictionary, map local HIS terminologies to global standards, ensuring the universal interpretability of synthesized concepts (Syzdykova et al., 2017).

Medical Sensors and Wearable Data Streams

Continuous monitoring via medical sensors and consumer wearables is an established method in patient care, especially for chronic disease management (Majumder et al., 2017). Devices like Continuous Glucose Monitors (CGMs) produce high-frequency, time-series data that provides insights into a patient's real-time health status (Nguyen & White, 2022). Integrating these continuous data streams into a HIS presents structural and scalability challenges (Majumder et al., 2017). Unlike episodic documentation, sensor data necessitates low-latency ingestion pipelines and standardized data models, such as HL7 FHIR Observation resources, to enable data persistence and visualization without compromising system performance (Lehne et al., 2019).

III METHODS

This section describes the methodology for evaluating synthetic clinical data generation and sensor data integration into the Heidelberg "studiKIS", following a four-phase approach.

Phase 1: Literature Review

The first phase consisted of a literature review that focused on identifying papers in two primary domains:

1. **Synthetic Clinical Data Generation:** Rule-based architectures and LLM-based agents for creating longitudinal patient records.

2. **Sensor Data Integration:** Standards and protocols for integrating real-time sensor data from medical devices into Hospital Information Systems.

Eligibility criteria

Papers were included if they met the criteria for their specific domain:

1. **Synthetic Data:** Inclusion required the use of LLMs, Generative AI, or state-machine models for generating synthetic clinical health data.

2. **Sensor Data:** Inclusion required the collection and integration of sensor data into a HIS or EHR. Studies solely utilizing sensor data to train machine learning algorithms without HIS integration were excluded.

Search Strategy

Literature was retrieved from PubMed and the IEEE Xplore Digital Library in March 2026. The retrieval used two search strings for the two primary domains:

1. **Synthetic Data:** For IEEE Xplore and PubMed: (("synthetic data generation" OR "synthetic clinical data" OR "synthetic data" OR "synthetic patient" OR "synthetic patient data" OR "synthetic EHR" OR "Synthea") **AND** ("large language models" OR "LLM" OR "generative AI" OR "autonomous agents" OR "agent-based" OR "machine learning") **AND** ("clinical workflows" OR "patient history" OR "medical records" OR "case vignettes" OR "longitudinal" OR "simulated patients")). No restrictions on publication year or language were applied.

2. **Sensor Data:** For PubMed: (("Hospital Information Systems" [MeSH] OR "Electronic Health Records" [MeSH]) AND ("sensor*" OR "sensor data" OR "wearable*") AND ("storage" OR "archiving" OR "persistence" OR "NoSQL" OR "SQL" OR "Time-series database" OR "openEHR" OR "FHIR" OR "repository")) NOT ("Blockchain" OR "Security" OR "Cryptography" OR "Privacy Preservation" OR "Cybersecurity").

   For IEEE Xplore, the identical string without *[MeSH]* filters was used. Only English and German articles were considered, with no publication year restrictions.

Synthesis methods

Depending on the underlying domain, two different methodological approaches were planned to synthesize the findings.

1. **Synthetic Data:** Included papers were categorized by their primary synthesis engine or framework. Additionally, the medical quality requirements reported by the authors were synthesized.

2. **Sensor Data:** Included papers were categorized by the used transmission protocols and interfaces.

Phase 2: Definition of Quality Criteria

The second phase focused on defining quality criteria for the synthetic patient data and integrating the sensor

data into the HIS. To establish these criteria, qualitative interviews were conducted with medical experts and IT professionals who had practical experience in academic teaching. These interviews followed a structured guideline targeting medical, information technology, and didactic metrics. Questions were tailored to each participant's specific domain, for instance, medical experts were not questioned on IT requirements. After the interviews, the answers were analyzed, and any requirement for synthetic patient data and integration of sensor data into the HIS cited by two or more experts was formally established as a quality criterion.

Phase 3: Prototypical Implementation

The third phase consisted of two tracks: generating synthetic patient histories (via rule-based and LLM-based approaches) and preparing real sensor data. For both data types, a prototypical ingestion pipeline into the HIS was developed.

Generation of Synthetic Patient Data

Rule-Based Approach: The open-source framework Synthea (Walonoski et al., 2018) was used for rule-based data generation. The repository was forked to generate localized German datasets, including regional names, addresses, and healthcare entities.

To limit the number of virtual clinicians, a mapping strategy was designed to substitute Synthea-generated providers with a fixed pool of clinicians existing in the HIS across five specialized departments (Cardiology, Neurology, Chest Pain Unit, Stroke Unit, and Neurological Rehabilitation).

LLM-Based Approach: An LLM-based pipeline utilizing Gemini 3.1 Pro (High) was developed to generate structured clinical records directly as FHIR bundles. The prompts were refined using an iterative approach to enforce OpenMRS-compatibility and strict formatting rules.

Real-World Sensor Data

Empirical continuous glucose monitoring data was collected from four participants over two weeks using FreeStyle Libre 3 (Nguyen & White, 2022) sensors. To enable direct import into the HIS and preserve historical timestamps, the data was mapped to a low-level HL7 FHIR Observation schema supported by Bahmni. This process used a corresponding continuous glucose concept identified within the OpenMRS concept dictionary. Finally, the data was visualized on the patients dashboard.

Ingestion Pipeline for Patient Records

Both the synthetic records and the real-world sensor data were ingested into the HIS via Bahmni's native FHIR API. The correct JSON schema for a successful import was reverse-engineered by exporting and analyzing the payload of a manually created reference patient.

Simulating a Continuous Clinical Workflow

To simulate continuous clinical workflows, a temporal ingestion pipeline was developed. FHIR-formatted patient data typically comprises multiple objects, ranging from demographic records to clinical encounters and observations. While a static approach imports these objects simultaneously, this pipeline partitions and ingests them at scheduled intervals to mirror the chronological intervals of real-world hospital visits.

Phase 4: Evaluation Framework

The fourth phase focused on establishing an evaluation framework to assess the efficacy and structural quality of the synthesized datasets against real-world clinical baselines. The medical experts who participated in the initial quality criteria interviews served as evaluators for this process.

In contrast to the synthetic data evaluation, the real-world sensor streams were evaluated qualitatively, focusing on the frontend visualization performance within the HIS environment.

Data Cohort Selection and Homogenization

Patient data from three distinct sources were used in the evaluation: rule-based generated (Synthea), generated through agentic generative models (LLMs), and real clinical records derived from anonymized patient records extracted from the public MIMIC-III dataset (Johnson et al., 2016). A uniform 40/40/40 split was created to ensure statistical balance.

MIMIC-III records were pseudo-anonymized with German identities matching the recorded biological sex. To account for variations in information depth, the records were homogenized to feature five clinical dimensions: allergies, diagnoses, medications, procedures, and laboratory/vital signs.

All records were presented uniformly, effectively blinding the evaluators to the underlying data source.

Data Collection Instruments

To evaluate the synthetic patient data, a structured questionnaire based on the quality criteria established in Phase 2 was developed. Each criterion was quantitatively assessed using a five-point Likert scale (ranging from 1

to 5). Additionally, a text field was provided for each criterion, enabling evaluators to add qualitative comments regarding the evaluation for the respective quality criteria.

Data Analysis

Quantitative Analysis Descriptive statistics (mean, standard deviation, and median) were calculated for each quality criterion across the three data sources (Real-world, Synthea, and LLM). To evaluate global differences between the data sources across the criteria, non-parametric Kruskal-Wallis tests were performed.

For criteria displaying statistically significant global differences, post-hoc pairwise comparisons were conducted using Mann-Whitney U tests. To control for the family-wise error rate, a conservative Bonferroni correction was applied.

Qualitative Analysis To evaluate the free-text comments provided by the evaluators, a thematic analysis was conducted. Qualitative feedback was systematically categorized to identify source-specific characteristics, errors, and patterns across the three data streams.

IV RESULTS

Literature Review and Synthesis Findings

The literature review identified relevant articles across both primary domains:

- **Synthetic Data Generation:** The study selection process is systematically documented in the PRISMA flow chart (Page et al., 2021) presented in Figure A.1. The initial database search retrieved 125 articles. Title and abstract screening excluded 72 papers. Full-text screening of the remaining 53 articles excluded 26, leaving 27 for the final analysis. These papers address Large Language Models and Transformers (n=9), Generative Adversarial Networks for tabular and image data (n=4), agent-based simulation via Synthea (n=2), and probabilistic or interactive frameworks (n=2).

  Based on the literature, core requirements for generated patient data are categorized into seven domains: *statistical fidelity* (e.g., high-dimensional disease codes reaching correlation probabilities of $R^2 > 0.9$) (X. Chen et al., 2025; Theodorou et al., 2023), *structural integrity* (hierarchical consistency and modeling informative missingness) (Josling et al., 2026), *temporal dynamics* (realistic inter-visit gaps) (Josling et al., 2026; Theodorou et al., 2023), *clinical utility* (validated via the "Train-Synthetic-Test-Real" framework) (Theodorou et al., 2023; Jung et al., 2025), *data privacy* (differential privacy to mitigate re-identification risks) (Tsao et al., 2023; Hahn et al., 2022) (X. Chen et al., 2025), *governance/ethics* (adherence to FAIR and CARE principles) (Tsao et al., 2023), and *active bias mitigation* (Theodorou et al., 2023).

- **Sensor Data Integration:** The search string yielded 269 articles (A.2). Title and abstract screening excluded 251 papers. Full-text screening of the remaining 18 articles excluded 4, leaving 14 for the final analysis. These papers utilize the HL7 FHIR standard (n=11) and classical REST interfaces for data transmission (n=3). Continuous data streaming into the HIS was not described in any study. Integration relied exclusively on batch-processing via server pull mechanisms.

Defined Quality Criteria from Expert Interviews

Six qualitative interviews were conducted with three medical experts and three IT professionals. The qualitative analysis of the interviews identified four core dimensions of synthetic data quality for academic teaching:

1. **Clinical Plausibility (c1):** The medically logical interplay of symptoms, diagnoses, and treatments, ranging from correct standard therapies to plausible medical errors.

2. **Longitudinal Consistency (c2):** Coherent progression of chronic conditions and follow-up examinations over time without illogical jumps.

3. **Chronological-Logical Sequence (c3):** Realistic time intervals between clinical events (e.g., from initial diagnosis through intervention to follow-up care).

4. **Overall Impression (c4):** The global synthesis of the patient record, which dictates student acceptance and didactic value.

Beyond clinical dimensions, the analysis identified key technical and educational requirements. Technically, systems must use standards like HL7 FHIR for interoperability and scale datasets to match learning objectives (from basic navigation to advanced analytics). Educationally, case complexity must adapt to student skill levels, offering simple, isolated cases for beginners and complex, longitudinal records for advanced learners. Crucially, all generated data must remain strictly relevant to the scenario.

Regarding empirical physiological sensor streams, the experts noted that evaluating raw data primarily assesses device precision rather than pedagogical value, as the data originates from certified hardware. Consequently, the focus shifted to the visual presentation and functional integration of sensor data within the studiKIS patient dashboard. The interviews defined four frontend requirements:

- **Graphical Trend Visualization:** High-frequency sensor streams require rendering as continuous curves or trend graphs rather than static tables to facilitate clinical pattern recognition.
- **Adaptive Contextual Granularity:** The interface must provide interactive zoom and scaling features to adjust the temporal resolution dynamically.
- **Anomaly Highlighting:** Physiological measurements outside established norms require visual flagging to indicate critical status changes.
- **Temporal Event Annotation:** The dashboard should support chronological marking of patient activities (e.g., dietary intake, medication) to correlate lifestyle events with physiological responses.

Prototypical Implementation

The implementation phase established technical pipelines for rule-based simulation, generative AI workflows, and empirical sensor data integration.

Generation of Synthetic Patient Data

Rule-Based Approach: By adjusting the configuration properties of the Synthea framework, the established workflow generated localized German patient datasets in FHIR format. However, technical evaluation of the pipeline components revealed functional limitations:

- The automated provider-pooling strategy failed to function as intended. Instead, the concept was tested by manually replacing the providers in a simplified Synthea-generated patient.
- The structural complexity of Synthea's full FHIR bundles led to repeated rejections by the HIS. Technical analysis identified semantic gaps and documentation constraints within the Bahmni FHIR module, revealing the operational scalability limits of the rule-based ingestion pipeline. Consequently, importing complete, Synthea-generated patient profiles as monolithic bundles proved unfeasible.

LLM-Based Approach: The LLM-based pipeline was implemented, resulting in an optimized, final prompt that reliably generates patient records in FHIR format. By structuring the generation process around a predefined set of medications and medical conditions, the prompt ensured full OpenMRS compatibility. Furthermore, this constraint-based approach mitigated LLM hallucinations, yielding consistent synthetic data.

Real-World Sensor Data

Continuous glucose sensor streams from four participants were parsed and integrated into the HIS via the FHIR interface. Within the OpenMRS concept dictionary, "*Fasting blood glucose measurement (mg/dL)*" was selected as the most appropriate representation for the sensor stream data. Modeling this high-frequency time-series data as low-level FHIR Observation resources bypassed standard clinical ordering workflows, enabling the direct ingestion of historical records. This data was subsequently visualized dynamically within the native frontend dashboard (Figure 2).

Ingestion Pipeline for Patient Records

Using the correct JSON structure for the FHIR interface enabled the import of both synthetic patient data and real-world sensor data into the HIS. However, an initial limitation arose regarding several Synthea-generated concepts that were not natively supported by Bahmni. This discrepancy was resolved by importing the CIEL dictionary, which contained the majority of the required concepts.

Simulating a Continuous Clinical Workflow

The temporal ingestion pipeline used a decoupled, database-driven architecture. A partitioning script parsed multi-object FHIR bundles into individual clinical transactions, assigned each a scheduled time, and stored them in a local SQLite database. An execution script, running as a cron job on the HIS server, then polled the database to execute these transactions exactly at their scheduled times.

Evaluation Framework

To circumvent limitations when importing complete Synthea-generated profiles into the HIS, the records were converted into an alternative format for review. Custom Python scripts rendered the complex FHIR bundles into uniform PDFs. Restricting the data to the five specified clinical dimensions, this pipeline generated an evaluation dataset of 120 unique, visually indistinguishable patient records across three modalities (Synthea, LLM, and real-world). To streamline the assessment and facilitate data

analysis, the review was conducted via a secure web application featuring a dual-pane interface: the randomized patient PDF on the left, and a five-point Likert scale with qualitative feedback fields for each quality criterion on the right. The application also prevented duplicate submissions and cached session progress to ensure data integrity.

Evaluation Results

Synthetic Data: Quantitative Analysis

A Kruskal-Wallis test ($\alpha = 0.05$) evaluated differences across the three data sources (Real-world, Synthea, LLM) across four criteria (Table 1). No statistically significant differences were found for $c1$, $c3$, and $c4$. However, $c2$ showed a significant global difference ($p = 0.002$), aligning with the visual distribution in Figure 1, where Synthea achieved the highest proportion of positive ratings and LLMs shifted toward negative scores. Mann-Whitney U tests for $c2$ with $\alpha = 0.05$ revealed that Synthea scored significantly higher than LLMs ($p < 0.001$) and Real data ($p = 0.027$). After applying a conservative Bonferroni correction ($\alpha = 0.0167$), the difference between Synthea and LLMs remained highly significant, while the difference between Synthea and Real data did not. No significant difference was found between Real data and LLMs ($p = 0.634$). Regarding data completeness, all 120 datasets were fully evaluated by two reviewers, with a subset of 30 datasets evaluated by a third reviewer.

Synthetic Data: Qualitative Analysis

Thematic analysis of the evaluators' free-text comments revealed distinct, source-specific characteristics:

- **Real-World Data (Real):** Had poor readability, local abbreviations, and unspecific codes (like "NOS"). It also showed unrealistic timing, where completely different clinical events happened on the exact same day.
- **Synthetic Data (Synthea):** Was well-structured but showed repetitive patterns (e.g., too many dental cases) and logical errors, such as contraindicated prescriptions during pregnancy.
- **LLM-Generated Data (LLM):** Had the lowest quality. It showed broken timelines, made-up laboratory units, and unrealistic treatment steps (e.g., delayed diabetes medication or wrong stroke protocols).

Sensor Data: Dashboard Visualization

The integration of empirical sensor data within the Bahmni-based patient dashboard yielded the following results for the four defined quality criteria:

For **Graphical Trend Visualization**, the platform renders high-frequency time-series data as continuous physiological curves within the native frontend. For **Anomaly Highlighting**, the system flags and color-codes values deviating from normal clinical ranges, provided these thresholds are predefined in the laboratory concept configuration.

Adaptive Contextual Granularity (dynamic scaling or zooming) is not supported natively within the UI. Adjusting the temporal scope requires modifying the system's backend configuration to change the number of displayed patient visits, followed by a page reload. Data granularity is fixed prior to ingestion. The infrastructure lacks mechanisms for post-ingestion aggregation or dynamic downsampling.

Regarding **Temporal Event Annotation**, the dashboard lacks a native mechanism to mark timestamps or annotate external clinical events. Consequently, correlating physiological fluctuations with patient activities directly within the interface is not supported.

V DISCUSSION

Interpretation of Findings

Interpretation of Quality Criteria

Comparing the literature and interview findings reveals shared core requirements for synthetic data evaluation but divergent priorities regarding its application. Both sources align on the need for temporal and clinical coherence, as the literature focuses on temporal dynamics and clinical utility (Josling et al., 2026; Theodorou et al., 2023; Jung et al., 2025) that match the experts' demands for chronological sequencing, longitudinal consistency, and clinical plausibility. Furthermore, the experts' requirement for interoperability standards like HL7 FHIR mirrors the literature's emphasis on governance, ethics, and FAIR principles (Tsao et al., 2023).

However, their primary evaluation metrics differ. While the literature prioritizes quantitative metrics such as statistical fidelity (X. Chen et al., 2025; Theodorou et al., 2023), structural integrity (Josling et al., 2026), and mathematical privacy guarantees (Tsao et al., 2023; Hahn et al., 2022; X. Chen et al., 2025), the interview results focus on qualitative and pedagogical utility for academic teaching. For educational purposes, experts prioritize narrative coherence, realistic clinical elements like medical errors, and adaptable case complexity. Consequently, while statistical parameters ensure technical validity and

Table 1.: Descriptive statistics and Kruskal-Wallis test results for group differences.

Criterion	Data Source	M	SD	Mdn	p-value
c1	Real data	3.02	1.24	3.0	0.070
	Synthea	3.15	0.93	3.0	
	LLM	2.80	0.95	3.0	
c2	Real data	3.07	1.42	3.0	**0.002****
	Synthea	3.49	1.00	3.0	
	LLM	2.93	1.13	3.0	
c3	Real data	3.16	1.45	3.0	0.112
	Synthea	3.57	1.00	4.0	
	LLM	3.26	1.15	3.0	
c4	Real data	2.97	1.26	3.0	0.165
	Synthea	3.11	0.95	3.0	
	LLM	2.82	0.90	3.0	

Note: M = Mean, SD = Standard Deviation, Mdn = Median.
** $p < 0.01$. Significant differences in c2 (Kruskal-Wallis test).

privacy, educational utility depends strictly on narrative realism and didactic flexibility.

Interpretation of Quantitative and Qualitative Group Differences in Patient Data

Statistically significant differences were limited to criterion $c2$ (Longitudinal Consistency), where Synthea-generated patient data outperformed LLM-generated records and demonstrated a positive trend against real-world data. These results may be attributed to Synthea's underlying generative framework, which is designed to simulate a patient's entire lifespan, beginning at birth, tracking all clinical encounters, and concluding at death (Walonoski et al., 2018). Consequently, longitudinal consistency is inherent to these records. In contrast, LLM-based patient profiles contained fewer encounters, observations, and diagnoses. Furthermore, while real-world data captured higher encounter volumes, it spanned a shorter historical duration. These factors likely explain Synthea's significant improvement over LLM profiles and its positive, although non-significant, trend against real-world benchmarks.

However, this comprehensive data generation model also introduced a major point of criticism in the qualitative findings. Evaluators noted that Synthea-generated records exhibited repetitive patterns, such as an over-representation of dental cases. This artifact reflects the lifespan-simulation methodology, where frequent, routine dental checkups naturally outnumber visits to hospitals.

For criteria $c1$, $c3$, and $c4$, the generated data was statistically comparable to each other and to authentic records. However, a wide score distribution across all modalities indicated high variance and low inter-rater reliability. Qualitative observations explain this variance: real-world records suffered from poor readability and unspecific coding, likely stemming from localized hospital abbreviations. Synthea records exhibited repetitive patterns and occasional lapses in clinical plausibility, and LLM data contained factual errors and hallucinations. Evaluators most likely prioritized these distinct deficits differently based on their clinical focus, penalizing structural disorganization, logical inconsistencies, or hallucinations with varying severity, resulting in low inter-rater reliability.

Interpretation of Sensor Data Visualization Quality

The Bahmni patient dashboard natively renders assigned laboratory values as continuous graphs. By mapping the sensor data to an appropriate concept and importing it via the FHIR interface, the streams were visualized, satisfying core expert requirements for real-time data representation. Furthermore, Bahmni automatically highlights out-of-range values, provided the range is defined in the concept. However, this flagging mechanism appears to be persisted at the database level during data ingestion rather than evaluated dynamically at runtime during visualization. Consequently, values imported via the FHIR

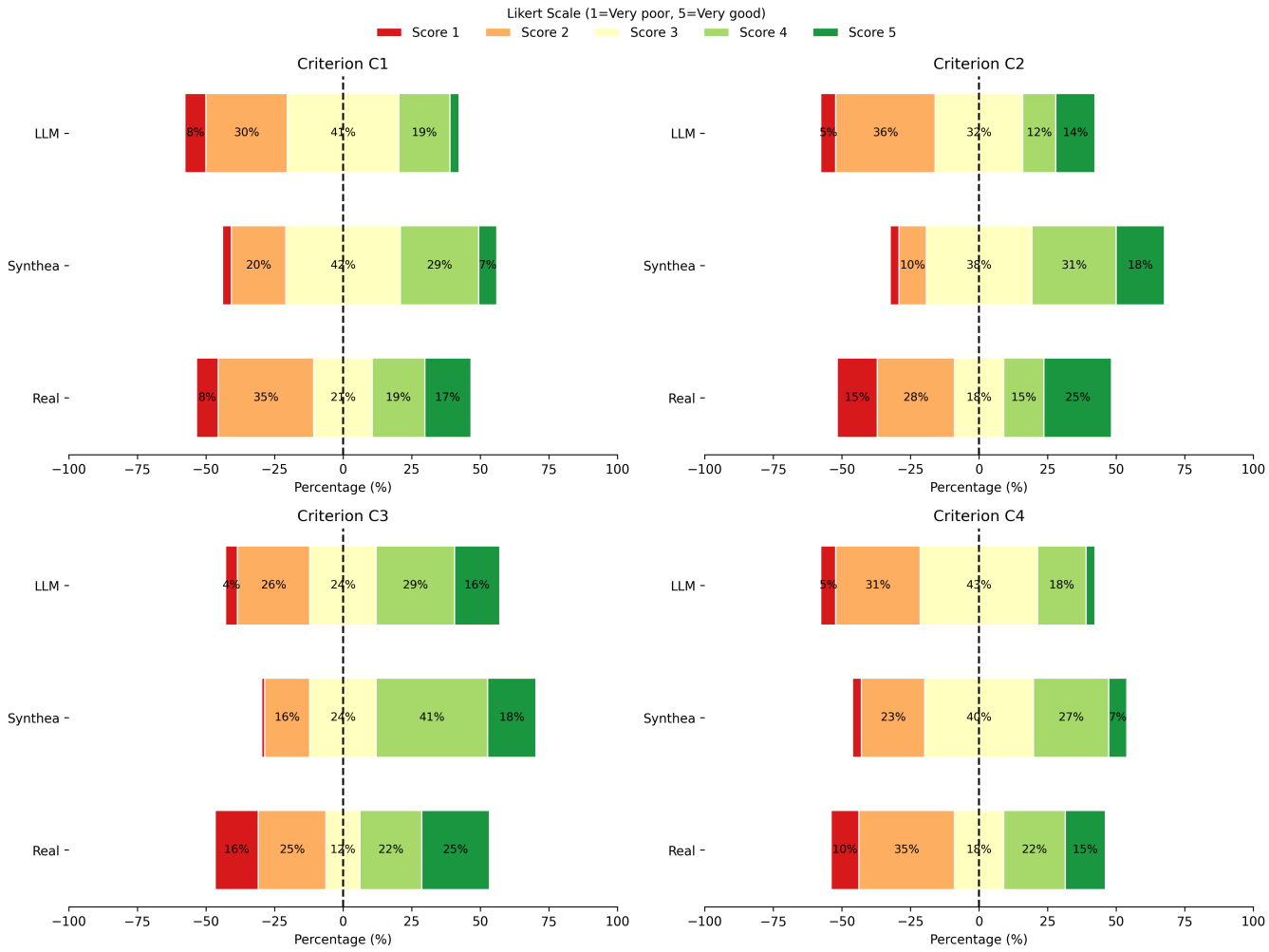


Figure 1.: Percentage distribution of ratings (1 = dark red, 5 = dark green) centered around the neutral rating of 3 (yellow). Positive ratings shift right; negative ratings shift left.

interface lacked out-of-range indicators despite breaching the defined ranges, whereas manually entered values were flagged appropriately. Incorporating range parameters within the FHIR payload might circumvent this barrier, but due to time constraints, this hypothesis could not be verified.

Consequently, the platform inherently satisfies the *Graphical Trend Visualization* criterion and could potentially satisfy *Anomaly Highlighting*, provided programmatic triggering can be achieved via the ingestion pipeline. However, the remaining two criteria, *Adaptive Contextual Granularity* and *Temporal Event Annotation*, are not supported out of the box. Prospective enhancements are detailed in Section "Advanced UI Component Engineering for Sensor Stream Dashboarding".

Limitations

Literature Review

To ensure comprehensive coverage of both the clinical and technical dimensions of the research, the literature review focused on two primary databases: PubMed and IEEE Xplore. While a structured methodology was applied, the search did not strictly adhere to systematic review standards or full PRISMA guidelines. Because the primary objective was the technical integration of data into a virtual HIS, the literature search served only to establish a functional baseline. Consequently, expanding the database selection and executing a full systematic review fell outside the project's scope.

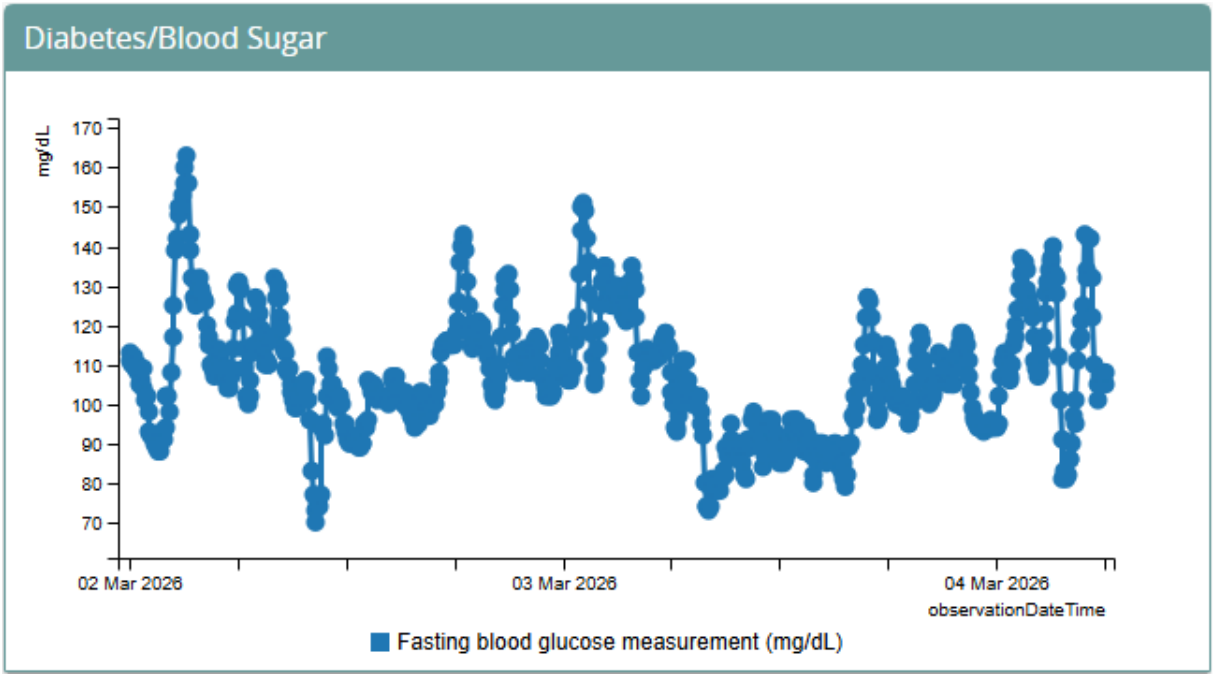


Figure 2.: The visualization of the continuous sensor data in the Bahmni patient dashboard.

Limitations regarding Data Completeness

Due to time constraints, 90 patient records were not evaluated by the third reviewer, introducing a risk of verification bias. However, given the fixed project timeline and the availability of two complete evaluation rounds for the entire sample, the analysis proceeded.

Technical Limitations

Because Synthea-generated patient records were too structurally complex for direct HIS import, an initial troubleshooting strategy was designed to analyze error logs and programmatically strip non-compliant fields. However, this approach was quickly deemed too time-intensive and raised scalability concerns, as manually auditing individual profiles cannot guarantee compatibility across subsequent patient records. Nonetheless, given the significant performance advantage observed for Synthea data regarding criterion $c2$ over LLM-generated records, developing a translation script to map Synthea's output to a Bahmni-compliant FHIR schema may present a key area for future work. Such an integration pipeline requires careful calibration. A naive truncation technique that blindly removes non-compliant fields risks degrading the clinical realism that drove Synthea's performance metrics. Hence, it may be necessary to develop alternative import strategies rather than simple data truncation to preserve dataset integrity.

Tied to the complexity of the Synthea files was the current inability to create a script, that replaced all newly generated Synthea providers with a predefined set of existing HIS clinicians. However, manual validation confirmed that provider replacement was successful within simplified Synthea profiles, and the Bahmni user interface correctly displayed the assigned personnel alongside their corresponding diagnoses. Consequently, while an automated provider-mapping script is theoretically possible, its utility is dependent on the script adapting the Synthea-generated FHIR file to a FHIR structure accepted by Bahmni.

In contrast, generating compliant FHIR structures for LLM-generated patient profiles is significantly more straightforward, as the output schema can be strictly enforced through targeted prompting. A limitation remains the reproducibility of LLM outputs across different model versions. Nevertheless, the detailed prompt architecture is expected to minimize variance when deploying alternative models, preserving the framework's generalizability. Furthermore, hallucinations persisted in the LLM-generated patient data. Proposed solutions to this problem are outlined in the section "Knowledge-Grounded Architectures for Generative AI".

Future Work

Based on the technical challenges, qualitative findings, and architectural trade-offs, these four areas for future research have been defined:

Automated Ingestion and Interoperability Pipelines for Rule-Based Simulation

Future work should develop a scalable translation script to map native Synthea FHIR bundles to Bahmni-accepted profiles. This pipeline must address two architectural gaps:

- **Dynamic Concept Harvesting:** Active database synchronization using the CIEL concept dictionary to programmatically inject missing clinical concepts into the HIS backend prior to ingestion.
- **Automated Provider Pooling:** A script to map dynamically generated virtual clinician identities to a fixed pool of pre-existing HIS users across specialized departments.

Future studies must evaluate the trade-off between structural simplification and medical realism to ensure that system compatibility does not compromise clinical complexity.

Knowledge-Grounded Architectures for Generative AI

To mitigate clinical hallucinations and guideline violations in LLM-generated records, future architectures should transition to knowledge-grounded generation:

- **Retrieval-Augmented Generation (RAG):** Coupling the generative engine with a vector database of localized clinical guidelines to ground probabilistic patient transitions in authentic pathways.
- **Multi-Agent Critic Frameworks:** A dual-agent system where a generative model synthesizes the history, while a specialized "Clinical Critic" agent verifies medical facts, units, and chronologies prior to database export.

Advanced UI Component Engineering for Sensor Stream Dashboarding

To address limitations in interactive granularity and contextual layout within the sensor dashboard, the following frontend extensions are proposed:

- **Interactive Frontend Scaling:** Zoom and pan components to adjust temporal resolutions directly within the browser, avoiding database downsampling or manual reloads.
- **Temporal Event Annotation:** A mechanism to anchor discrete lifestyle metrics (e.g., dietary intake) onto tracking graphs, allowing students to correlate behaviors with physiological responses.

Implementing these features in Bahmni might require core codebase alterations, posing maintainability risks during framework updates. This could lead to either repetitive manual code synchronization upon every release or upstream adoption by the core Bahmni development team, so that the changes persist in upcoming versions. Due to the lack of long-term sustainability for local customizations, evaluating alternative virtual HIS platforms that natively support these requirements may be necessary.

Improving the Continuous Clinical Workflow

A two-step approach, which partitions patient records into individual transactions and executes imports via a cron job at scheduled times, simulates a real-world clinical workflow. However, determining appropriate temporal intervals between distinct encounters and diagnoses remains an open question, as does identifying the optimal baseline for the initiation of a patient record.

Because Synthea-generated profiles track an individual's history from birth, subsequent clinical events are distributed across decades. Simulating continuous ingestion from birth is practically unfeasible, as real-time execution would mirror an actual lifespan. A more viable alternative could involve a hybrid ingestion strategy: statically importing a historical baseline of encounters and diagnoses, followed by dynamic, continuous workflows. Nevertheless, the temporal fidelity of synthetic encounters requires further evaluation. While criterion $c3$ (Chronological-Logical Sequence) evaluates this progression in this article, neither Synthea nor LLM-generated datasets currently excel in this domain. Consequently, further research is required to optimize chronological logic before deploying these synthetic profiles in medical education.

Alternatively, to maximize control over educational datasets, course instructors could predefine patient records aligned directly with specific lesson plans. Instructors would configure the *Scheduled-Time* attribute to dictate exactly when a virtual patient is imported into the HIS. This workflow could be driven by a dedicated web application integrated with the ingestion database, allowing the cron job to extract these targeted encounters. Developing and evaluating such a pedagogical management platform could be the focus of a subsequent study.

Methodological Scaling and Multi-Reviewer Validation

Finally, to address the risk of verification bias stemming from the incomplete third evaluation round of 90 datasets, future evaluation cycles must enforce full multi-reviewer coverage across the entire case matrix. Additionally, research should systematically evaluate inter-rater disagreement to understand how subjective expert priorities

influence the validation of synthetic clinical data.

VI CONCLUSION

This research project demonstrates the feasibility of generating synthetic patient data through rule-based approaches using Synthea and generative AI approaches using LLMs. Synthea-generated records showed significant improvements over LLM-generated records concerning longitudinal consistency and a positive trend over real-world patient data. By sequentially importing these records into a HIS using HL7 FHIR standards, the system simulates the continuous nature of real-world clinical documentation. Additionally, continuous data streams from wearables were imported into the HIS and displayed using the patient's native dashboard. These dynamic structures allow students to analyze realistic disease progression and patient-generated data without privacy risks.

Despite these advantages, the current implementation has several limitations. Rule-based profiles from Synthea produced repetitive patterns, and their FHIR bundles were rejected by the HIS due to structural complexity. Conversely, Large Language Models generated simpler, compatible FHIR data but suffered from broken timelines and clinical hallucinations. Furthermore, the sensor integration lacked interactive frontend features like temporal scaling, and the evaluation was limited by low inter-rater reliability.

Future research should focus on technical and methodological refinements. Key priorities include developing scalable translation scripts to map complex rule-based outputs to HIS-compliant schemas, alongside implementing knowledge-grounded AI architectures, such as Retrieval-Augmented Generation and multi-agent frameworks, to actively enforce guideline compliance and eliminate hallucinations. Additionally, dedicated frontend engineering is required to enable dynamic downsampling and event marking within the sensor dashboard. Addressing these technical barriers and expanding future evaluations to full multi-reviewer validation will help establish reliable, dynamic simulation environments for medical education.

BIBLIOGRAPHY

Agrawal, M., Hegselmann, S., Lang, H., Kim, Y. & Sontag, D. (2022). *Large language models are few-shot clinical information extractors.* doi: https://doi.org/10.48550/arXiv.2205.12689

Bichel-Findlay, J., Koch, S., Mantas, J., Abdul, S. S., Al-Shorbaji, N., Ammenwerth, E., ... Wright, G. (2023). Recommendations of the international medical informatics association (imia) on education in biomedical and health informatics: Second revision. *International Journal of Medical Informatics*, *170*, 104908. doi: https://doi.org/10.1016/j.ijmedinf.2022.104908

Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F. K. & Mahmood, F. (2021, Juni). Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*, *5* (6), 493–497. doi: https://doi.org/10.1038/s41551-021-00751-8

Chen, X., Wu, Z., Shi, X., Cho, H. & Mukherjee, B. (2025, Juli). Generating synthetic electronic health record data: a methodological scoping review with benchmarking on phenotype data and open-source software. *Journal of the American Medical Informatics Association: JAMIA*, *32* (7), 1227–1240. doi: 10.1093/jamia/ocaf082

Choi, J., Bove, L., Tarte, V. & Choi, W. (2021, 01). Impact of simulated electronic health records on informatics competency of students in informatics course. *Healthcare Informatics Research*, *27*, 67-72. doi: 10.4258/hir.2021.27.1.67

Dahmen, J. & Cook, D. (2019, März). SynSys: A synthetic data generation system for healthcare applications. *Sensors (Basel)*, *19* (5), E1181. doi: https://doi.org/10.3390/s19051181

Ferreira, J. C., Elvas, L. B., Correia, R. & Mascarenhas, M. (2025). Empowering health professionals with digital skills to improve patient care and daily workflows. *Healthcare*, *13* (3). Zugriff auf `https://www.mdpi.com/2227-9032/13/3/329` doi: 10.3390/healthcare13030329

Hahn, W., Schütte, K., Schultz, K., Wolkenhauer, O., Sedlmayr, M., Schuler, U., ... Wolfien, M. (2022, August). Contribution of Synthetic Data Generation towards an Improved Patient Stratification in Palliative Care. *Journal of Personalized Medicine*, *12* (8), 1278. doi: 10.3390/jpm12081278

Johnson, A., Pollard, T. & Mark, R. (2016, September). MIMIC-III Clinical Database. *PhysioNet*. Zugriff auf `https://doi.org/10.13026/C2XW26` (Version 1.4) doi: 10.13026/C2XW26

Josling, G., Diouf, I. & Khanna, S. (2026, Januar). A novel pipeline for realistic synthetic longitudinal EHR data generation. *Research Square*, rs.3.rs–8497559. doi: 10.21203/rs.3.rs-8497559/v1

Jung, H., Kim, Y., Seo, J., Choi, H., Kim, M., Han, J., ... Kim, Y.-H. (2025, Dezember). Clinical Assessment of Fine-Tuned Open-Source LLMs in Cardiology: From Progress Notes to Discharge Summary. *Journal of Healthcare Informatics Research*, *9* (4), 686–702. doi: 10.1007/s41666-025-00203-x

Lehne, M., Sass, J., Essenwanger, A., Schepers, J. & Thun, S. (2019, August). Why digital medicine depends on interoperability. *NPJ Digit. Med.*, *2* (1), 79. doi: https://doi.org/10.1038/s41746-019-0158-1

Majumder, S., Mondal, T. & Deen, M. J. (2017, Januar). Wearable sensors for remote health monitoring. *Sensors (Basel)*, *17* (1), E130. doi: https://doi.org/10.3390/s17010130

Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S. & Ramoni, R. B. (2016). SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Inform Assoc*, *23* (5), 899–908. doi: https://doi.org/10.1093/jamia/ocv189

Medlock, S., Ploegmakers, K. J., Cornet, R. & Pang, K. W. (2023). Use of an open-source electronic health record to establish a "virtual hospital": A tale of two curricula. *International Journal of Medical Informatics*, *169*, 104907. doi: https://doi.org/10.1016/j.ijmedinf.2022.104907

Mohan, V., Woodcock, D., Mcgrath, K., Scholl, G., Pranaat, R., Doberne, J. W., ... Ash, J. S. (2016). Using simulations to improve electronic health record use. In *Clinician training and patient safety: Recommendations from a consensus conference. AMIA ... annual symposium proceedings. AMIA symposium* (S. 904–913).

Nguyen, A. & White, J. R. (2022, 10). Freestyle libre 3. *Clinical Diabetes*, *41* (1), 127-128. Zugriff auf `https://doi.org/10.2337/cd22-0102` doi: 10.2337/cd22-0102

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... McKenzie, J. E. (2021). Prisma 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*. Zugriff auf `https://www.bmj.com/content/372/bmj.n160` doi: 10.1136/bmj.n160

Syzdykova, A., Malta, A., Zolfo, M., Diro, E. & Oliveira, J. L. (2017, November). Open-source electronic health record systems for low-resource settings: Systematic review. *JMIR Med. Inform.*, *5* (4), e44.

Theodorou, B., Xiao, C. & Sun, J. (2023, März). Synthesize Extremely High-dimensional Longitudinal Electronic Health Records via Hierarchical Autoregressive Language Model. *Research Square*, rs.3.rs–2644725. doi: 10.21203/rs.3.rs-2644725/v1

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F. & Ting, D. S. W. (2023, August). Large language models in medicine. *Nature Medicine*, *29*

(8), 1930–1940. doi: https://doi.org/10.1038/s41591-023-02448-8

Tsao, S.-F., Sharma, K., Noor, H., Forster, A. & Chen, H. (2023, Mai). Health Synthetic Data to Enable Health Learning System and Innovation: A Scoping Review. *Studies in Health Technology and Informatics*, *302*, 53–57. doi: 10.3233/SHTI230063

Walonoski, J., Kramer, M., Nichols, J., Quina, A., Moesel, C., Hall, D., ... McLachlan, S. (2018, 03). Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *Journal of the American Medical Informatics Association*, *25* (3), 230-238. Zugriff auf `https://doi.org/10.1093/jamia/ocx079` doi: 10.1093/jamia/ocx079

A. Appendix

A.1. Gegenstand und Motivation

- **Gegenstand:** Erzeugung realistischer klinischer Daten für das Heidelberger Lehre-KIS durch die Kombination von generativen Modellen, regelbasierten Systemen und Sensordatenintegration.
- **Bedeutung:** Steigerung der Lehrqualität durch authentische Fallbeispiele, die angehende Fachkräfte auf die Komplexität und Dynamik rechnerbasierter Anwendungssysteme vorbereiten.
- **Problematik:** Bestehende synthetische Datensätze sind oft statisch und bilden die zeitliche Variabilität klinischer Parameter unzureichend ab. Zudem erschweren proprietäre Schnittstellen die Einbindung realer Medizinprodukte in Lehrumgebungen.
- **Motivation:** Das Ziel ist die Schaffung eines „lebendigen“ KIS, das durch die Fusion von algorithmisch erzeugten Krankheitsverläufen und echten physiologischen Datenströmen (Wearables/Sensoren) eine hohe Plausibilität der bereitgestellten Daten erreicht. Durch diesen technologieoffenen Ansatz können sowohl moderne Sprachmodelle als auch bewährte Simulationsmethoden genutzt werden, um die Lücke zwischen statischer Theorie und dynamischer klinischer Praxis zu schließen.

A.2. Zielsetzung

A.2.1. 1. Literaturrecherche:

Identifikation von State-of-the-Art Ansätzen für synthetische Daten und Sensor-Schnittstellen.

1.1 Welche Technologien werden verwendet, um strukturierte klinische Daten synthetisch zu erzeugen?

1.2 Welche Technologien werden verwendet, um Sensor-Daten in das KIS einzubinden?

1.3 Welche Anforderungen werden in der Literatur an synthetische strukturierte klinische Daten gestellt?

A.2.2. 2. Definition Qualitätskriterien:

Festlegung didaktischer und technischer Anforderungen für das Heidelberger Lehre-KIS.

2.1 Befragung von Expert*innen

2.2 Analyse der Ergebnisse: Welche Anforderungen haben Spezialisten an die Daten?

A.2.3. 3. Prototypische Implementierung:

Erzeugung von Datensätzen und Anbindung von Wearables/Hardware sowie Einbindung in das KIS.

3.1 Wie könnten strukturierte klinische Daten synthetisch durch nicht-Sprachmodell-basierte Agenten erzeugt werden?

3.2 Wie könnten strukturierte klinische Daten synthetisch durch Sprachmodelle erzeugt werden?

3.3 Wie können Datensätze (sowohl sensorische als auch synthetische) im KIS eingebunden werden?

3.4 Wie können die Daten dynamisch („lebendig“) in ein KIS eingebunden werden?

A.2.4. 4. Bewertung:

Evaluierung der Ansätze hinsichtlich Eignung, Komplexität und Ressourcenverbrauch.

4.1 Aus welchen Quellen sollen die Patientendaten für die Evaluierung stammen?

4.2 Wie können die verschiedenen Patientendaten einheitlich dargestellt werden?

4.3 Wie kann die Evaluierung effizient durchgeführt werden?

4.4 Gibt es statistische Unterschiede zwischen den Patientengruppen?

4.5 Welche qualitativen Auffälligkeiten zeigen sich zwischen den Patientengruppen?

A.3. Methodik und Ergebnisse

A.3.1. Ziel 1: Literaturrecherche

Frage 1.1: Technologien Synthetische Daten

Methodik

Für die initiale Literaturrecherche wurde der folgende Suchstring für die Datenbanken IEEEXplore und PubMed verwendet:

((„synthetic data generation“ OR „synthetic clinical data“ OR „synthetic data“ OR „synthetic patient“ OR „synthetic patient data“ OR „synthetic EHR“ OR „Synthea“) **AND** („large language models“ OR „LLM“ OR „generative AI“ OR „autonomous agents“ OR „agent-based“ OR „machine learning“) **AND** („clinical workflows“ OR „patient history“ OR „medical records“ OR „case vignettes“ OR „longitudinal“ OR „simulated patients“))

Für die Auswahl der Publikationen wurden die folgenden Ein- und Ausschlusskriterien definiert:

- **Einschlusskriterien:**
  - Generierung von synthetischen Patientendaten durch Sprachmodelle oder nicht-Sprachmodell-basierte Agenten
- **Ausschlusskriterien:**
  - Generierung von Daten,die nicht im Kontext der Medizin stehen
  - Die Methode zur Generierung der Daten wird nicht genannt.

Mit den resultierenden Publikationen wurde ein Titel-, Abstract- und Volltext-Screening durchgeführt.

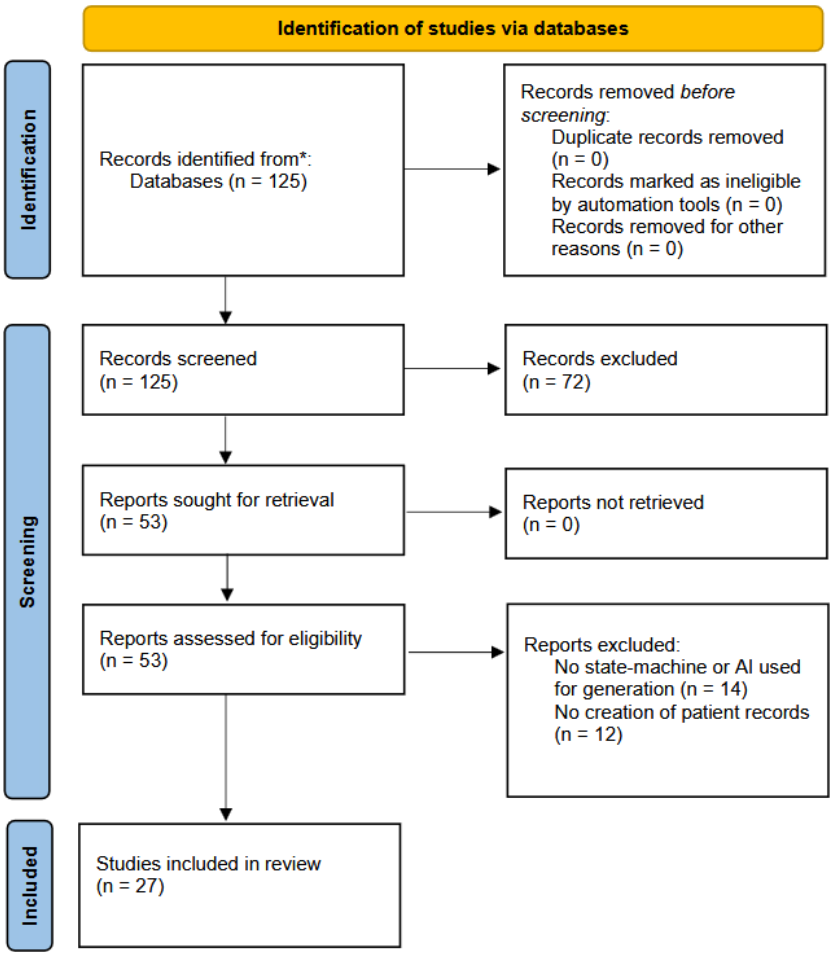


Abbildung A.1.: PRISMA Flussdiagramm (Page et al., 2021) des Auswahlprozesses der Artikel für die Literaturrecherche zur Generierung von synthetischen Patientendaten.

Ergebnisse

Die Analyse der identifizierten Quellen ($n = 27$) erfolgte in zwei Schritten: Zunächst eine Klassifizierung der Artikeltypen und anschließend eine technische Differenzierung der darin beschriebenen Methoden.

Verteilung der Artikeltypen

Die Verteilung der Artikeltypen gliedert sich in zwei Kernbereiche: Reviews und methodische Benchmarks (n=10) dienten primär der Identifikation des aktuellen Stands der Technik sowie dem Vergleich bestehender Frameworks.

Den komplementären Teil bildet die Primärliteratur zu spezifischen Methoden (n=17), welche konkrete Implementierungen oder neu entwickelte Algorithmen beschreibt.

Eingesetzte Technologien

Zur Abbildung der technologischen Breite wurden die Ansätze wie folgt kategorisiert:

- **LLMs & Transformer** ($n = 9$): Fokus auf die Generierung synthetischer Patientenakten durch Sprachmodelle.
- **Generative Adversarial Networks (GANs)** ($n = 4$): Einsatz primär für tabellarische Daten und Bilddaten.
- **Agenten-basierte Simulation (Synthea)** ($n = 2$): Regelbasierte Erzeugung lebenslanger Patientenhistorien.
- **Diffusion & Interaktive Frameworks** ($n = 2$): Neuere Ansätze zur probabilistischen Datenerzeugung.

Beantwortung

Die technologische Landschaft wird primär durch den Einsatz von Large Language Models (LLMs) und Generative Adversarial Networks (GANs) geprägt, ergänzt durch regelbasierte Systeme wie das Synthea-Framework.

Frage 1.2: Technologien Einbindung Sensor Daten

Methodik

Die methodische Grundlage dieser Arbeit bildet eine systematische Literaturrecherche in den Datenbanken PubMed und IEEE Xplore (Stand: März 2026).

Für die Recherche in PubMed wurde folgender Suchstring verwendet:

((„Hospital Information Systems“[MeSH] OR „Electronic Health Records“[MeSH]) AND („sensor∗“ OR „sensor data“ OR „wearable∗“) AND („storage“ OR „archiving“ OR „persistence“ OR „NoSQL“ OR „SQL“ OR „Time-series database“ OR „openEHR“ OR „FHIR“ OR „repository“)) NOT („Blockchain“ OR „Security“ OR „Cryptography“ OR „Privacy Preservation“ OR „Cybersecurity“))

Für die Recherche in IEEE Xplore wurde folgender Search String angewendet:

((„Hospital Information System∗“ OR „Electronic Record∗“) AND („sensor∗“ OR „sensor data“ OR „wearable∗“) AND („storage“ OR „archiving“ OR „persistence“ OR „NoSQL“ OR „SQL“ OR „Time-series database“ OR „openEHR“ OR „FHIR“ OR „repository“)) NOT („Blockchain“ OR „Security“ OR „Cryptography“ OR „Privacy Preservation“ OR „Cybersecurity“))

Für die Auswahl der Primärliteratur wurden die folgenden Ein- und Ausschlusskriterien definiert:

- **Einschlusskriterien:**
  - Nutzung von Sensordaten und deren explizite Integration in ein KIS oder eine elektronische Patientenakte (Electronic Health Record, EHR).
- **Ausschlusskriterien:**
  - Publikationen, die sich ausschließlich mit Datensicherheit, Datenschutz oder Kryptographie befassen.
  - Arbeiten, die lediglich die Erhebung von Sensordaten beschreiben, ohne eine Integration in ein KIS/EHR vorzusehen.

– Beiträge mit einem reinen Fokus auf Machine-Learning-Algorithmen (ohne systemarchitektonischen Bezug).

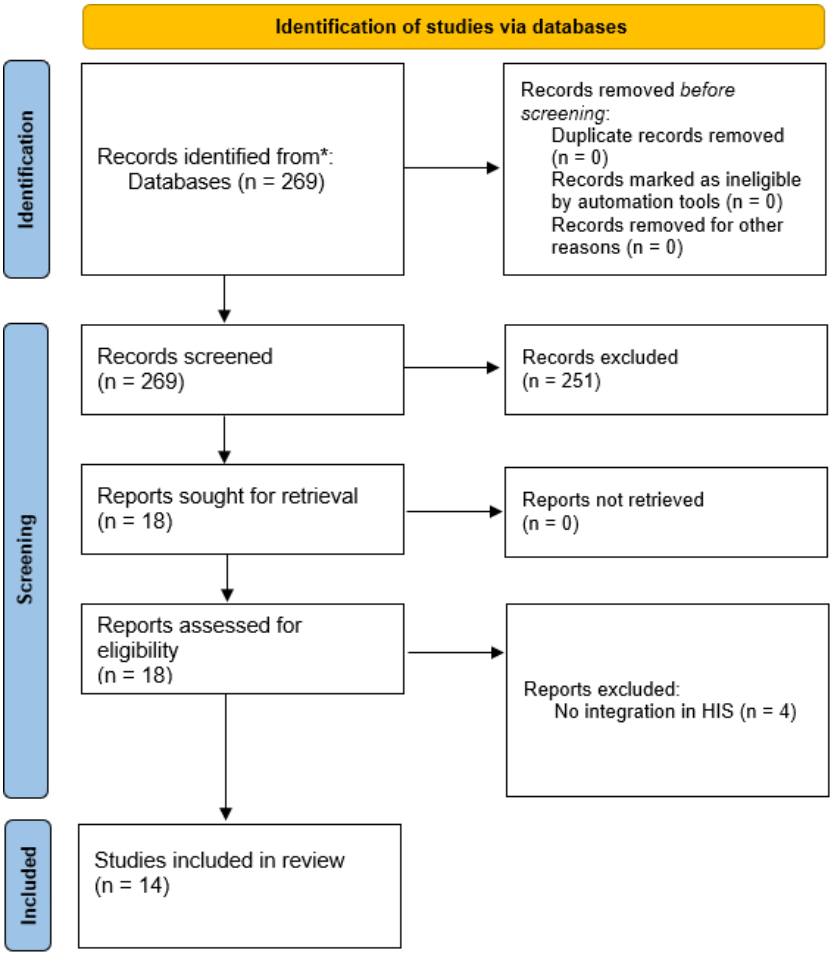


Abbildung A.2.: PRISMA Flussdiagramm (Page et al., 2021) des Auswahlprozesses der Artikel für die Literaturrecherche zur Integration der Sensordaten.

Die Literatursuche und das anschließende Screening führten zu folgenden Ergebnissen:

In PubMed wurden initial 69 Ergebnisse identifiziert. Nach einem Titel- und Abstract-Screening wurden 10 Publikationen als relevant eingestuft. Das anschließende Volltext-Screening führte zum Ausschluss von zwei weiteren Arbeiten, sodass insgesamt 8 Publikationen aus PubMed in die Analyse

einbezogen wurden.

In IEEE Xplore wurden initial 200 Ergebnisse gefunden. Nach dem Titel- und Abstract-Screening verblieben 8 relevante Treffer. Im Zuge des Volltext-Screenings wurden ebenfalls zwei Arbeiten ausgeschlossen, woraus sich 6 finale Publikationen aus IEEE Xplore für die Synthese ergaben.

Die insgesamt $n = 14$ identifizierten Quellen wurden quantitativ analysiert und manuell hinsichtlich der verwendeten Schnittstellenstandards kategorisiert. Der kombinierte Ablauf der Literaturrecherche in PubMed und IEEE Xplore ist im Flussdiagramm in Abbildung A.2 dargestellt.

Ergebnisse

Die Analyse zeigt eine deutliche Dominanz des HL7 FHIR Standards, der in $n = 11$ Artikeln Anwendung findet. Lediglich $n = 3$ Quellen setzen auf klassische REST-Schnittstellen zur Datenübertragung.

Beantwortung

Der Dateneinzug erfolgt primär über den HL7 FHIR Standard oder einfache REST-Schnittstellen. Keine der untersuchten Arbeiten beschreibt ein kontinuierliches Daten-Streaming in das KIS, stattdessen erfolgt die Integration ausschließlich im Batch-Processing-Verfahren durch gezieltes Pulling von den Servern.

Frage 1.3: Anforderungen an Synthetische Daten

Methodik

Die methodische Grundlage bildet eine systematische Literaturrecherche in den Datenbanken PubMed und IEEE Xplore (Stand: März 2026). Aus den identifizierten Quellen wurden die von den Autoren explizit formulierten

Anforderungen manuell extrahiert, um eine ungefilterte Vergleichsbasis für die weiteren Analysen zu schaffen.

Ergebnisse

Aus den Papern ergaben sich die folgenden Anforderungen:

- **Statistische Fidelität:** Synthetische Daten müssen die statistischen Eigenschaften des Originaldatensatzes präzise replizieren, insbesondere die Randverteilungen der Variablen und komplexe Korrelationsstrukturen. Hochdimensionale Krankheitskodes sollten dabei eine Korrelation der Kodewahrscheinlichkeiten von $R2 > 0,9$ erreichen. (X. Chen et al., 2025) (Theodorou et al., 2023)

- **Strukturelle Integrität:** Es wird eine hierarchische Konsistenz über multiple Tabellen hinweg gefordert (z. B. zwischen Patientenstammdaten und Laborwerten). Zudem müssen klinisch bedeutsame Fehlwerte (Informative Missingness) explizit modelliert werden.(Josling et al., 2026)

- **Temporale Dynamik:** Bei longitudinalen Daten müssen sowohl die sequenziellen Abhängigkeiten klinischer Ereignisse als auch die spezifischen Zeitintervalle (Inter-visit Gaps) realistisch abgebildet werden. (Josling et al., 2026) (Theodorou et al., 2023)

- **Klinische Nutzbarkeit:** Die Daten müssen im "Train-Synthetic-Test-Real"(TSTR) Framework validiert werden, um sicherzustellen, dass auf synthetischen Daten trainierte Modelle auf realen Daten performant sind. In spezialisierten Domänen wie der Kardiologie wird zudem eine klinische Kohärenz und medizinische Relevanz gefordert, die durch Fachexpert*innen validiert wurde. (Theodorou et al., 2023) (Jung et al., 2025)

- **Datenschutz und Sicherheit:** Eine zentrale Anforderung ist die Ver-

meidung von Overfitting, um Re-Identifizierungsrisiken zu minimieren. Hierfür wird oft die Integration von Differential Privacy (DP) und die Messung der ϵ-Identifizierbarkeit gefordert. (Tsao et al., 2023) (Hahn et al., 2022) (X. Chen et al., 2025)

- **Governance und Ethik:** Die Generierung sollte den FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable) entsprechen. Bei der Verwendung von Daten marginalisierter Gruppen müssen zudem die CARE-Prinzipien (Collective benefit, Authority to control, Responsibility, Ethics) beachtet werden. (Tsao et al., 2023)

- **Bias-Vermeidung:** Die Generierungsprozesse dürfen bestehende algorithmische Biases aus den Originaldaten nicht verstärken oder perpetuieren. (Theodorou et al., 2023)

Beantwortung

Zusammenfassend lassen sich die in der Literatur formulierten Anforderungen in sieben Kernbereiche unterteilen: statistische Fidelität, strukturelle Integrität, temporale Dynamik, klinische Nutzbarkeit, Datenschutz, Governance/Ethik sowie die aktive Bias-Vermeidung.

A.3.2. Ziel 2: Definition der Qualitätskriterien

Frage 2.1: Befragung von Expert*innen

Methodik

Zur Erhebung der Expertenanforderungen wurde im ersten Schritt ein strukturierter Gesprächsleitfaden konzipiert. Dieser diente als methodische Grundlage für die anschließenden Experteninterviews und fokussierte sich gezielt auf die medizinische, informationstechnische sowie didaktische Qualität der syn-

thetisch generierten Patientendaten und der in das Lehre-KIS importierten Sensordaten.

Link zum Leitfaden: Leitfaden Qualitätskriterien

Im Anschluss an die Ausarbeitung des Leitfadens erfolgte die Durchführung der eigentlichen Experteninterviews. Hierbei wurde ein zielgruppenadaptierter Ansatz gewählt, um die Befragung optimal an die jeweilige Fachexpertise der Interviewpartner anzupassen. Dies bedeutet konkret, dass nicht jedem/jeder Expert*in der vollständige Fragenkatalog vorgelegt wurde: Personen mit primär medizinischem Hintergrund wurden nicht zu informationstechnischen Details befragt, ebenso wie technische Expert*innen von den spezifisch medizinischen Qualitätskriterien ausgenommen waren.

Ergebnisse

Durch den gewählten Ansatz konnten insgesamt sechs qualitative Interviews erfolgreich realisiert werden. Die Kohorte der Befragten setzte sich aus drei Personen mit medizinischem Hintergrund sowie drei Personen mit informationstechnischer Spezialisierung zusammen. Ein essenzielles Auswahlkriterium war ihre ausgewiesene praktische Erfahrung im Bereich der akademischen Lehre. Die Befragung lieferte umfassende, protokollierte Datenbasen für die anschließende Auswertung.

Beantwortung

Die Befragung der Expert*innen wurde erfolgreich über sechs qualitative, zielgruppenadaptierte Interviews auf Basis eines dreigeteilten Leitfadens (Medizin, IT, Didaktik) durchgeführt. Dadurch konnte ein breites, interdisziplinäres Spektrum an praktischer Lehrerfahrung für die nachfolgende Kriterienauswertung erfasst werden.

Frage 2.2: Analyse der Ergebnisse: Welche Anforderungen haben Spezialisten an die Daten?

Methodik

Die methodische Auswertung der erhobenen Daten erfolgte durch eine qualitative Analyse der zuvor angefertigten Interviewprotokolle. Ziel dieser Analyse war es, die Aussagen der Expert*innen systematisch zu strukturieren, Gemeinsamkeiten zu identifizieren und daraus die übergeordneten Kernanforderungen an die Datenqualität abzuleiten.

Ergebnisse

Die qualitative Analyse der Interviewprotokolle führte zur Identifikation von vier primären Hauptkriterien für die Datenqualität der synthetischen Daten:

1. **Klinische Plausibilität:** Beschreibt das medizinisch logische Zusammenspiel von Symptomen, Diagnosen und Behandlungen. Das Spektrum reicht dabei von fehlerfreien Standardtherapien bis zu plausiblen Fehlbehandlungen.
2. **Longitudinale Konsistenz:** Bezieht sich auf den langfristigen, korrekten Krankheitsverlauf über größere Zeiträume hinweg. Chronische Erkrankungen und Folgeuntersuchungen müssen lückenlos aufeinander aufbauen und dürfen keine unlogischen Sprünge aufweisen.
3. **Zeitlich logische Abfolge:** Umfasst die korrekte Reihenfolge sowie die Realitätsnähe der zeitlichen Abstände einzelner klinischer Ereignisse (z. B. die Zeitspanne von der Erstdiagnose über die Intervention bis zur Nachsorge).
4. **Gesamteindruck:** Aggregiert die vorgenannten Aspekte zu einem globalen Qualitätseindruck der Patientenakte, welcher maßgeblich die Akzeptanz und den didaktischen Nutzwert für die Studierenden bestimmt.

Neben den klinischen Dimensionen identifizierte die Analyse auch wesentliche technische und didaktische Anforderungen. In technischer Hinsicht müssen Standards wie HL7 FHIR für die Interoperabilität genutzt werden und Datensätze so skalieren, dass sie den Lernzielen entsprechen. Didaktisch gesehen muss sich die Fallkomplexität an das Kompetenzniveau der Studierenden anpassen: Sie sollte einfache, isolierte Fälle für Anfänger und komplexe, longitudinale Patientenakten für Fortgeschrittene bieten.

In Bezug auf die Sensordaten wurde festgestellt, dass eine Auswertung der Qualität der Sensordaten nicht sinnvoll wäre, da diese von den medizinischen Geräten stammen und man dadurch nur die Qualität des Sensors bewerten würde. Stattdessen hat sich der Fokus der Qualitätsbewertung auf die visuelle Darstellung und die funktionale Integration von Sensordaten in das studiKIS-Patienten-Dashboard verschoben. In den Interviews wurden vier Frontend-Anforderungen definiert:

- **Grafische Trendvisualisierung:** Hochfrequente Sensorströme müssen als kontinuierliche Kurven oder Trendgraphen statt als statische Tabellen dargestellt werden, um das Erkennen klinischer Muster zu erleichtern.
- **Adaptive Granularität:** Die Benutzeroberfläche muss interaktive Zoom- und Skalierungsfunktionen bieten, um die zeitliche Auflösung dynamisch anzupassen.
- **Hervorhebung von Anomalien:** Physiologische Messwerte außerhalb etablierter Normen sollen visuell markiert werden, um auf kritische Statusänderungen hinzuweisen.
- **Zeitliche Ereignisannotation:** Das Dashboard sollte die chronologische Markierung von Patientenaktivitäten (z. B. Nahrungsaufnahme, Medikation) unterstützen, um Ereignisse, die die Messwerte beeinflussen

können, darstellen zu können.

Beantwortung

Die Spezialisten stellen insgesamt vier zentrale Anforderungen an die synthetischen Patientendaten: klinische Plausibilität, longitudinale Konsistenz, eine zeitlich logische Abfolge sowie einen stimmigen Gesamteindruck. Diese extrahierten Kernkriterien definieren die qualitativen Mindestanforderungen, die für eine erfolgreiche und akzeptierte Nutzung in der akademischen Lehre erfüllt sein müssen. Aus technischer Sicht sollen Standards wie HL7 FHIR zur Integration benutzt werden und didaktisch gesehen soll das Niveau der Komplexität der Daten an das Niveau der Studierenden angepasst werden. Bezüglich der Integration der Sensordaten wurden diese vier Kriterien definiert: Grafische Trendvisualisierung, Adaptive Granularität, Hervorhebung von Anomalien und Zeitliche Ereignisannotation.

A.3.3. Ziel 3: Prototypische Implementierung

Frage 3.1: Generierung mittels Synthea

Methodik

Die Patientengenerierung erfolgte auf Basis des öffentlichen Synthea-Repositorys unter Einbeziehung der offiziellen Wiki-Dokumentation zur Modellierung klinischer Abläufe.

Hürden

Im Implementierungsprozess wurden drei zentrale Herausforderungen adressiert:

- **Herkunft der Patienten:** Synthea generiert standardmäßig Patienten, die in den USA verortet sind. Das bedeutet, dass sie im amerikanischen

Gesundheitssystem behandelt werden und amerikanische Namen und Adressen besitzen. Die Lösung war es, das Repository zu forken (Ein Fork ist eine Kopie eines Repositories, welches denselben Code und Sichtbarkeit besitzt wie das ursprüngliche Repository) und die EU/DE Properties, die von Synthea zur Verfügung gestellt werden, zu nutzen und eigene Änderungen einzuführen, um Patienten, die in der EU / Deutschland verortet werden, zu erhalten. Durch die zur Verfügung gestellten Properties von Synthea wurden Krankenhäuser sowie Städte und Krankenkassen aus Deutschland importiert. Die Änderungen, die nachträglich noch erstellt wurden, waren die Einführung deutscher Vor und Nachnamen, sowie deutscher Straßennamen, welche in Synthea auch synthetisch generiert werden.

- **Vorgenerierung einer kompletten Patientenhistorie:** Synthea erstellt für einen Patienten immer eine komplette lückenlose Patientenhistorie. Das bedeutet, dass von Geburt bis ggf. Tod die Geschichte eines Patienten erstellt wird. Für ein möglichst realistisches KIS ist dies jedoch wenig repräsentativ. Als Lösung haben wir hierfür die FHIR Bundle, die von Synthea erstellt werden, in einzelne Transaktionen aufgeteilt. Hierdurch kann die Patientenhistorie aufgeteilt werden und nur bis zu einem bestimmten Punkt ins KIS importiert werden. Wie dies gestaltet werden soll, also bis zu welchem Punkt die Daten importiert werden und wo der Cut-Off gesetzt wird, steht noch offen.

- **Ärzte in Synthea und der Umgang im KIS:** Synthea erstellt für einen Patienten immer eine Liste an Ärzten, die diesen über den Verlauf behandelt haben. Dies würde für das „studiKIS“ bedeuten, dass sich die Menge der Ärzte exponentiell erhöhen würde, da jeder Patient eine

Vielzahl an Ärzten hat, die ihn im Verlauf behandeln. Ein Lösungsansatz wird in den Ergebnissen zu Frage 3.c dargestellt.

Ergebnisse

Es wurde ein funktionsfähiger Workflow etabliert, der die bedarfsgerechte Erzeugung beliebig großer Patientenkollektive im standardisierten FHIR-Format ermöglicht. (Link: Synthea Repository zur Erstellung deutscher Populationen)

Beantwortung

Durch die gezielte Anpassung des Synthea-Frameworks und die Implementierung regionaler Spezifika konnte eine Pipeline geschaffen werden, die lokalisierte und prozessual steuerbare synthetische Patientendaten für die Systemintegration bereitstellt.

Frage 3.2: Generierung mittels LLM

Methodik: Entwicklung eines Prompts, der es ermöglicht, strukturierte klinische Daten als FHIR Bundle zu erzeugen.

Hürden:

- **Halluzination von Krankheiten:** Eine Hürde war es, dass bei der Generierung der Daten vom LLM Verläufe halluziniert wurden, die nicht konsistent waren und wenig Plausibel. Hierfür wurde als Lösungsansatz eine Liste an Diagnosen und Medikamenten dem LLM zur Verfügung gestellt. Hierdurch wurden plausiblere und konsistentere Daten erzeugt.
- **Inkonsistenz im Format:** Bei der Erzeugung von FHIR Bundlen im JSON Format traten Inkonsistenzen in der Formatierung auf. Durch die Erweiterung des Prompts mit Punkt 6. OpenMRS-Kompatibilität und 7. Formatvorgabe konnte dies überwunden werden.

Ergebnisse: Mithilfe des erstellten Prompts können durch die Nutzung des Modells Gemini 3.1 Pro (High)

Link zum Prompt: Prompt FHIR Datengenerierung

Beantwortung: Durch einen Prompt der dem Modell klare Richtlinien gibt können Patientendaten im FHIR Format erstellt werden.

Frage 3.3: Import der Daten ins KIS

Methodik

Der Import der generierten Patientendaten in das bestehende Lehre-KIS erfolgte über die integrierte FHIR-Schnittstelle.

Hürden

Im Rahmen des Importprozesses traten mehrere technische Herausforderungen auf:

- **Fehlende Dokumentation:** Eine wesentliche Erschwerung stellte die unzureichende Dokumentation des FHIR-Moduls von Bahmni dar. Die korrekten Formate und Endpunkte für den Import spezifischer klinischer Informationen der generierten Patientendaten (wie Diagnosen und Medikation) mussten empirisch ermittelt werden.
- **Fehlende Concept Mappings:** Die von Synthea generierten Patientendaten enthielten Diagnosecodes, die standardmäßig nicht als Konzepte in Bahmni hinterlegt waren. Dieses Problem konnte durch die Integration des CIEL-Concept-Dictionaries gelöst werden.

- **Neu generierte Ärzte:** Ein weiteres signifikantes Hindernis beim Importprozess stellte die Handhabung des medizinischen Personals dar. Synthea generiert standardmäßig nicht nur die eigentlichen klinischen Patientendaten, sondern verknüpft diese systematisch mit eigens dafür neu erstellten, virtuellen Behandlern. Um einen erfolgreichen Import zu gewährleisten, hätte dies bedeutet, dass für jeden neuen Datensatz zwingend auch das jeweils neu generierte ärztliche Personal im Lehre-KIS angelegt werden müsste. Da sich Synthea systemseitig nicht so konfigurieren lässt, dass es auf einen vordefinierten, fixen Pool an Ärzten zurückgreift, würde dieser Umstand bei fortlaufender Datengenerierung zu einem exponentiellen und unkontrollierbaren Anwachsen des Arztstamms in der Systemdatenbank führen. Um dieser Problematik zu begegnen, sollte ein dezidiertes Skript entwickelt werden. Dieses durchsucht das von Synthea erzeugte FHIR-Bundle automatisiert nach den verknüpften Behandlern und ersetzt diese systematisch durch eine vordefinierte Auswahl real existierender Ärzte aus dem Lehre-KIS. Um die Konsistenz zu wahren, wurden hierbei exakt dieselben fünf Fachabteilungen beziehungsweise ärztlichen Rollen verwendet (Kardiologie, Neurologie, Chest Pain Unit, Stroke Unit, Neurologische Reha), die zuvor auch dem Large Language Model (LLM) als Kontext für die Generierung der Patientendaten übergeben wurden. Aufgrund von zeitlichen Einschränkungen wurde das Skript final jedoch nicht entwickelt, es wurden lediglich Daten für einen Probepatienten erstellt, bei dem die meisten Einträge der Synthea-generierten Daten entfernt und bei den bleibenden Observations der Behandler ersetzt wurde.

- **Komplexe Synthea-Datenstruktur:** Selbst nach dem erfolgreichen Mapping fehlender klinischer Konzepte und der Ersetzung der Behandler bei den Daten des Probepatienten traten beim Import weiterhin gra-

vierende Kompatibilitätsprobleme auf. Die generierten Patientendaten enthielten spezifische FHIR-Ressourcen und Datenfelder, die von der Architektur des Lehre-KIS abgelehnt wurden. Um diese Inkompatibilitäten aufzulösen, wurde zunächst ein iterativer Ansatz verfolgt: Die Datensätze wurden eingelesen, der resultierende Error-Log analysiert, das Feld, welches den Fehler verursacht, im Skript entfernt und der Importprozess wiederholt. Obwohl das Skript dahingehend erweitert wurde, betroffene Felder sukzessive zu bereinigen, stieß dieser Ansatz auf weitere Hürden. Es traten kontinuierlich neue Fehler auf, beispielsweise durch klinische Konzepte, die selbst durch das zuvor implementierte CIEL-Mapping nicht abgedeckt waren. Eine theoretische Lösung hätte in der Erweiterung des Einleseskripts bestanden, sodass dieses fehlende Konzepte dynamisch im Lehre-KIS anlegt und mit den Codes aus dem Synthea-FHIR-Bundle verknüpft. Angesichts zeitlicher Restriktionen und der hohen technischen Komplexität wurde dieser Pfad jedoch verworfen, und der Fokus wurde vollständig auf den korrekten Import der durch das LLM generierten Patientendaten gelegt. Der Versuch, die Synthea-Datensätze durch kontinuierliches Entfernen von Feldern anzupassen, erwies sich letztlich als nicht ausreichend skalierbar: Bei jedem neuen Generierungslauf hätten potenziell neue, bislang unbekannte Felder auftreten können, die das Einleseskript noch nicht abfängt und die unweigerlich zu neuen Fehlimporten geführt hätten. Letztlich war dieses ressourcenintensive Vorgehen jedoch der unzureichenden Dokumentation des Bahmni-FHIR-Moduls geschuldet, da im Vorfeld nicht spezifiziert war, welche Felder in welcher Formatierung zwingend erwartet oder strikt abgelehnt werden.

- **Darstellung der Laborwerte:** Die Abbildung von Laborwerten in Bahmni erfordert regulär einen zweistufigen Prozess (Order-Anforderung und

LIS-Rückmeldung), um im Dashboard als dediziertes „Labor Ergebnis“ zu erscheinen. Die Nachbildung dieses spezifischen FHIR-Objekts für die generierten Patientendaten erwies sich als nicht zielführend. Die Replikation der Struktur eines manuell angelegten Laborwerts führte beim Import lediglich zur Erstellung einer allgemeinen Observation.

Ergebnisse

Aufgrund der anfänglich unzureichenden Dokumentation der FHIR-Schnittstelle wurde initial ein Skript entwickelt, das den manuellen Erstellungsprozess von Laborwerten über die grafische Benutzeroberfläche von Bahmni simuliert. Hierfür wurde systematisch analysiert, welche URLs bei den jeweiligen Interaktionen in der UI aufgerufen und welche Datenobjekte dabei übermittelt wurden. Diese Netzwerkinformationen wurden extrahiert und algorithmisch im Skript repliziert. Es gelang, nahezu den gesamten Prozess ausschließlich über diese nachgebildeten HTTP-Anfragen abzubilden. Eine Ausnahme bildete lediglich der Schritt, bei dem die ID des Laborauftrags mit der ID der Patientendaten verknüpft wird. Diese Verknüpfung wird bei der manuellen Eingabe nicht explizit als separates Datenpaket gesendet, sondern beim dynamischen Seitenaufbau des Laborinformationssystems direkt in die anklickbaren URLs eingebettet. Um diese zwingend notwendige Verbindung programmtechnisch zu extrahieren, musste als Workaround auf eine direkte SQL-Abfrage der zugrunde liegenden Datenbank zurückgegriffen werden. Mithilfe dieses Skripts war es anschließend möglich, automatisiert eine Vielzahl von Laborwerten zu den jeweiligen Patientendaten hinzuzufügen, ohne die zeitaufwendigen manuellen Schritte in der UI durchlaufen zu müssen. Ein entscheidender Nachteil dieser Methode war jedoch, dass sich die Laborwerte auf diese Weise nicht mit historischen Zeitstempeln versehen ließen.

Aus diesem Grund wurde der Fokus im darauffolgenden Schritt erneut auf die

Realisierung des Imports über die reguläre FHIR-Schnittstelle gelegt. Durch umfangreiche empirische Analysen und iteratives Testen konnte schließlich das exakte FHIR-Schema identifiziert werden, das vom Lehre-KIS fehlerfrei akzeptiert wird. Mit der Kenntnis dieser genauen strukturellen Anforderungen wurde das verwendete LLM mittels spezifischem Prompting angewiesen, die generierten Patientendaten exakt in diesem Format zu erzeugen. Diese Maßnahme führte dazu, dass sich die resultierenden Patientendaten problemlos in das System einlesen ließen.

Nachdem auch die korrekte FHIR-Datenstruktur für die Laborwerte ermittelt war, konnten diese ebenfalls über die Schnittstelle importiert werden. Einschränkend muss hierbei erwähnt werden, dass die Laborwerte systembedingt stets als Teil einer allgemeinen „Observation“ eingelesen wurden und nicht als dedizierte „Labor Ergebnisse“ in der nativen Kategorie des Systems persistiert werden konnten. Da sich diese Observationen jedoch problemlos und historisch korrekt im Trends-Dashboard visualisieren ließen, in welchem die Laborwerte im zeitlichen Verlauf dargestellt werden, wurde diese Art der Abspeicherung für die Zwecke des Projekts als ausreichend erachtet.

Beantwortung

Zusammenfassend lässt sich feststellen, dass der Import generierter Patientendaten über die FHIR-Schnittstelle in das Lehre-KIS erfolgreich realisierbar ist, sofern die Datensätze exakt der identifizierten, systemspezifischen FHIR-Struktur entsprechen.

Frage 3.4: Dynamischer Datenimport

Methodik: Import der Daten über einen zeitlichen Verlauf, um realitätsnahe zeitliche Verläufe abzublilden.

Ergebnisse: Die synthetisch generierten Daten werden mithilfe eines Python-Scripts in einzelne FHIR Transaktionen aufgeteilt, die in einer SQLite Datenbank abgespeichert werden (Link zum Python Script). Die Transaktionen besitzen neben den FHIR Daten eine Scheduled-Time zu der sie im KIS vorhanden sein sollen. Ein zweites Script, welches auf dem KIS-Server als Cronjob (Ein Cronjob wird zu einem definierten Zeitpunkt regelmäßig ausgeführt) läuft, importiert die Einträge aus der Datenbank in das KIS bei denen die planmäßige Zeit erreicht wurde (Link zum Cronjob Script).

Beantwortung: Mithilfe der entwickelten Skripte ist es möglich, ein „lebendiges" KIS zu simulieren. Durch die Aufteilung der FHIR Objekte in einzelne Transaktionen und den darauffolgenden sequentiellen Import in das KIS kann ein dynamischer Patientenverlauf, welcher das Krankenhaus zu verschiedenen Zeiten aufsucht, dargestellt werden.

A.3.4. Ziel 4: Bewertung der Ansätze

Frage 4.1: Auswahl der Datenquellen

Methodik

Für die Evaluierung werden Daten aus agentenbasierten und nicht-agentenbasierten Systemen sowie reale Patientendaten herangezogen. Diese Datensätze werden anschließend durch medizinisches Fachpersonal anhand der in Ziel 2 definierten Kriterien bewertet.

Ergebnisse

Als nicht-agentenbasiertes System wird das Framework Synthea zur Generierung synthetischer Patientendaten genutzt. Die agentenbasierten Daten werden durch Large Language Models erzeugt. Ergänzend werden reale, anonymisierte Patientendaten aus dem MIMIC-III-Datensatz integriert. Es

wurde eine gleichmäßige Verteilung (40/40/40-Split) gewählt, sodass sich eine Gesamtstichprobe von 120 Datensätzen für die Bewertung ergibt.

Beantwortung

Die Auswahl der Datenquellen deckt somit drei Dimensionen ab: agentenbasierte LLM-Generierung, regelbasierte Synthea-Simulation und reale klinische Daten aus MIMIC-III.

Frage 4.2: Einheitliches Darstellungsformat

Methodik

Um eine objektive und unverzerrte Evaluation aller Datensätze durch die Mediziner*innen zu gewährleisten, mussten die Patientendaten in einem einheitlichen und standardisierten Format präsentiert werden. Schließlich wurde entschieden, die Patientendaten im PDF-Format aufzubereiten, da dies die zugänglichste und universellste Option für alle Beteiligten darstellte.

Hürden

Der Prozess der Formatvereinheitlichung war mit mehreren technischen Herausforderungen verbunden. Das Framework Synthea verfügt über keinen nativen PDF-Export. Die integrierte Option eines HTML-Exports führte im lokalen Systemkontext zu Fehlern. Ebenso scheiterten Versuche, die von Synthea generierten FHIR-Dateien über webbasierte Plattformen wie clinFHIR oder den HAPI-FHIR-Server adäquat zu visualisieren. Als verbleibende Alternative wurde der CSV-Export von Synthea genutzt. Es wurde ein proprietäres Skript entwickelt, das die exportierten CSV-Strukturen einliest und daraus ein übersichtliches PDF-Dokument generiert, welches alle demografischen und klinischen Parameter abbildet.

Im nächsten Schritt mussten die übrigen Datensätze, bestehend aus den anony-

misierten MIMIC-III-Daten sowie den vom LLM generierten FHIR-Dateien, in exakt dieselbe CSV-Struktur überführt werden. Hierzu wurde ein zweites Skript implementiert, welches auf Basis der heterogenen FHIR-Eingangsdaten strukturell identische CSV-Tabellen erzeugt. Durch diese Standardisierung konnten im anschließenden Schritt mithilfe des primären Generierungsskripts für alle drei Datenquellen optisch und strukturell ununterscheidbare PDFs erzeugt werden.

Eine spezifische Hürde bei den MIMIC-III-Daten lag in deren vollständiger Anonymisierung: Alle Datensätze wiesen identische Personennamen auf und enthielten keinerlei Adressdaten. Um zu verhindern, dass die Evaluator*innen allein anhand dieser demografischen Leerstellen auf die Gruppenzugehörigkeit der Patientendaten schließen können, wurden die Datensätze algorithmisch mit pseudozufälligen Klarnamen und Adressen angereichert. Dabei wurde das im Originaldatensatz hinterlegte biologische Geschlecht berücksichtigt, um durch passende Vornamen logische Inkonsistenzen und damit verbundene Verzerrungen bei der Bewertung auszuschließen.

Zudem unterschieden sich die Datenquellen erheblich in ihrer Informationstiefe. Die Synthea-Datensätze enthielten komplexe Zusatzinformationen wie Versicherungsverhältnisse, Abrechnungsmodalitäten und detaillierte Pflegepläne, welche weder in den LLM-Daten noch in den Musterelementen der Datenspende existierten. Um strukturelle Diskrepanzen zu eliminieren, wurde der Anzeigeumfang in den PDF-Dokumenten vereinheitlicht. Neben den demografischen Stammdaten wurden restriktiv nur die folgenden fünf klinischen Kategorien aggregiert und dargestellt: „Allergien“, „Diagnosen“, „Medikamente“, „Behandlungen“ sowie „Labor- & Vitalwerte“.

Ergebnisse

Als zentrales Ergebnis resultieren zwei Skripte: Das erste Skript transformiert die LLM-generierten (Link) sowie die realen (Link) FHIR-Datensätze in standardisierte CSV-Strukturen, während das zweite Skript diese Tabellen in das finale PDF-Layout konvertiert (Link). Auf diese Weise wurden insgesamt 120 standardisierte PDF-Dokumente erzeugt, jeweils 40 Datensätze pro Kategorie. Diese einheitliche Aufbereitung ermöglicht den Mediziner*innen eine effiziente, vergleichende Analyse der klinischen Inhalte.

Beantwortung

Mithilfe der eigens entwickelten Transformationsskripte ist es gelungen, Patientendaten aus drei technologisch unterschiedlichen Quellen so zu harmonisieren, dass sie rein visuell nicht voneinander zu unterscheiden sind, was zu einer Verblindung der Expert*innen führt. Das finale PDF-Format ist zudem so reduziert und übersichtlich gestaltet, dass medizinische Expert*innen ohne FHIR-Anwenderkenntnisse eine valide qualitative Bewertung der synthetischen Krankengeschichten vornehmen können.

Frage 4.3: Effiziente Durchführung der Evaluierung

Methodik

Zur Durchführung der Evaluierung muss eine intuitive Infrastruktur geschaffen werden, die sowohl die Visualisierung der Patientendaten als auch die Datenerfassung effizient abbildet. Für die Kriterien wird eine 5-stufige Likert-Skala sowie ein optionales Freitextfeld für qualitatives Feedback festgelegt. Um die Validität der Ergebnisse zu gewährleisten, muss das System sicherstellen, dass jede bewertende Person jeden Datensatz exakt einmal evaluiert, um Mehrfachbewertungen auszuschließen. Zudem ist eine kontinuierliche Datenpersistenz notwendig, um Datenverlust zu jedem Zeitpunkt der Erhebung

auszuschließen. Als Evaluator*innen wurden die drei medizinischen Experten, die zuvor bei der Erstellung der Qualitätskriterien interviewt wurden, zu Rate gezogen.

Ergebnisse

Als Lösung für die definierten Anforderungen wurde eine passwortgeschützte, webbasierte Anwendung implementiert (Link: Evaluierungsplattform). Um Verzerrungen zu minimieren, wurden drei anonymisierte Bewertungsprofile angelegt, deren Zuweisung an die Mediziner*innen durch eine externe Person randomisiert und verblindet erfolgte. Zur Unterstützung der Evaluierenden und zur Gewährleistung eines standardisierten Ablaufs wurde eine schriftliche Anleitung erstellt und den Mediziner*innen bereitgestellt (Link: Anleitung Evaluation). Die Benutzeroberfläche wurde zweigeteilt gestaltet: Auf der linken Seite wird ein zufällig ausgewählter, noch unbewerteter Patientendatensatz als PDF dargestellt, während auf der rechten Seite die zugehörigen Likert-Skalen und Textfelder platziert sind. Nach dem Absenden einer Bewertung werden die Daten direkt in einer Cloud-Datenbank persistent gespeichert. Das System filtert nach jedem Durchgang die bereits bewerteten Fälle und lädt dynamisch den nächsten unbewerteten Datensatz, bis alle 120 Datensätze vollständig abgearbeitet sind. Der Prozess schließt mit einer Bestätigungsmeldung über den erfolgreichen Abschluss ab.

Beantwortung

Die Anforderungen an eine strukturierte, persistente und fehlerfreie Datenerhebung wurden durch die Entwicklung einer passwortgeschützten Webanwendung realisiert. Diese ermöglicht über ein zweigeteiltes Interface die Evaluierung mittels Likert-Skalen, schließt Mehrfachbewertungen systemseitig aus und sichert den Fortschritt der Bewertenden kontinuierlich in einer Cloud-

Datenbank. Eine begleitende Anleitung stellte zudem die Standardisierung des Bewertungsprozesses sicher.

Frage 4.4: Statistische Unterschiede zwischen den Patientengruppen

Methodik

Zur Überprüfung auf statistisch signifikante Unterschiede zwischen den Patientengruppen (Echte Daten, Synthea, LLM) hinsichtlich der vier Evaluierungskriterien (c1 bis c4) wurde der Kruskal-Wallis-Test angewandt. Da die Bewertungen auf einer ordinalen 5-stufigen Likert-Skala basieren und eine Normalverteilung nicht vorausgesetzt werden kann, ist dieses nicht-parametrische Verfahren gut geeignet. Für Kriterien, die im Gesamtvergleich ein signifikantes Ergebnis aufzeigten, wurde anschließend post-hoc der Mann-Whitney-U-Test für paarweise Vergleiche durchgeführt, um die genauen Gruppenunterschiede zu identifizieren. Das anfängliche Signifikanzniveau wurde auf $\alpha = 0{,}05$ festgelegt. Zur visuellen Analyse der relativen Verteilung wurde zudem ein Diverging Stacked Bar Chart (Abbildung A.3) erstellt.

Ergebnisse

Wie in Tabelle A.1 zu sehen, zeigten die Kruskal-Wallis-Tests für die Kriterien c1 ($p = 0{,}070$), c3 ($p = 0{,}112$) und c4 ($p = 0{,}165$) keine statistisch signifikanten Unterschiede zwischen den drei Datenquellen. Für das Kriterium c2 ergab sich jedoch ein signifikantes Ergebnis ($p = 0{,}002$).

Diese statistischen Befunde spiegeln sich deutlich in der Verteilung der Bewertungen wider (siehe Abbildung A.3). Das Balkendiagramm verdeutlicht, dass bei Kriterium c2 (und partiell c3) die Synthea-Daten den mit Abstand größten Anteil an positiven Bewertungen (Scores 4 und 5) aufweisen, während extrem negative Bewertungen (Scores 1 und 2) nahezu ausbleiben. Bei den

Tabelle A.1.: Deskriptive Statistiken und Kruskal-Wallis-Testergebnisse zur Überprüfung auf Gruppenunterschiede zwischen den Datenquellen.

Kriterium	**Datenquelle**	**M**	**SD**	**Mdn**	**p-Wert**
c1	Echte Daten	3,02	1,24	3,0	0,070
	Synthea	3,15	0,93	3,0	
	LLM	2,80	0,95	3,0	
c2	Echte Daten	3,07	1,42	3,0	**0,002****
	Synthea	3,49	1,00	3,0	
	LLM	2,93	1,13	3,0	
c3	Echte Daten	3,16	1,45	3,0	0,112
	Synthea	3,57	1,00	4,0	
	LLM	3,26	1,15	3,0	
c4	Echte Daten	2,97	1,26	3,0	0,165
	Synthea	3,11	0,95	3,0	
	LLM	2,82	0,90	3,0	

Anmerkung: M = Mittelwert, SD = Standardabweichung, Mdn = Median.
** $p < 0{,}01$. Signifikante Unterschiede bei c2 (Kruskal-Wallis-Test).

LLM-generierten Daten zeigt sich hingegen eine klare Verschiebung in den negativen Bereich.

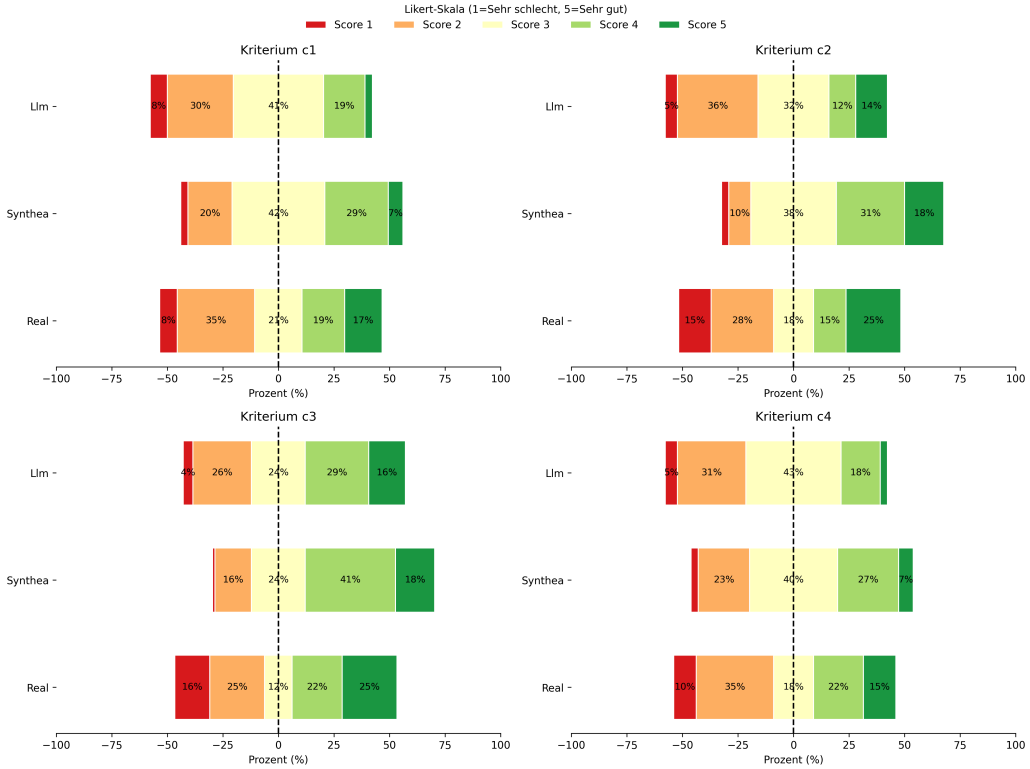


Abbildung A.3.: Prozentuale Verteilung der Bewertungen (1 = dunkelrot, 5 = dunkelgrün) aufgeschlüsselt nach Kriterien und Datenquellen. Die neutrale Bewertung 3 (gelb) ist um die Nulllinie zentriert. Anteile auf der rechten Seite spiegeln eine tendenziell positive, Anteile auf der linken Seite eine negative Bewertung wider.

Die anschließenden paarweisen Mann-Whitney-U-Tests für c2 bestätigten dies rechnerisch: Die durch Synthea generierten Fälle (Mittelwert 3,49) wurden signifikant besser bewertet als die durch das LLM generierten Fälle (Mittelwert 2,93; $p < 0{,}001$). Im Vergleich zu den echten Fällen (Mittelwert 3,07) zeigten die Synthea-Fälle ebenfalls höhere Bewertungen ($p = 0{,}027$), was jedoch unter Berücksichtigung einer konservativen Bonferroni-Korrektur für multiples Testen (angepasstes Signifikanzniveau $\alpha = 0{,}0167$) knapp das Signifikanzniveau verfehlt. Zwischen den echten Fällen und den LLM-generierten Fällen konnte bei c2 kein signifikanter Unterschied festgestellt werden ($p = 0{,}634$).

Allerdings muss einschränkend festgehalten werden, dass ein dritter Prüfer

die Daten von 90 Patienten aus Zeitgründen nicht mehr auswerten konnte. Insgesamt wurden somit 120 Datensätze jeweils zweimal vollständig evaluiert, während eine Teilmenge von 30 Datensätzen zusätzlich durch den dritten Prüfer beurteilt wurde. Die verbleibenden 90 Datensätze liegen demnach nur in zweifacher Ausfertigung vor. Obwohl diese fehlende dritte Evaluation das Risiko einer Verzerrung birgt, wurde die statistische Analyse wie geplant durchgeführt. Ausschlaggebend hierfür war, dass bereits zwei vollständig evaluierte Datensätze vorlagen und eine Fertigstellung der ausstehenden Bewertungen aus zeitlichen Gründen nicht mehr abgewartet werden konnte.

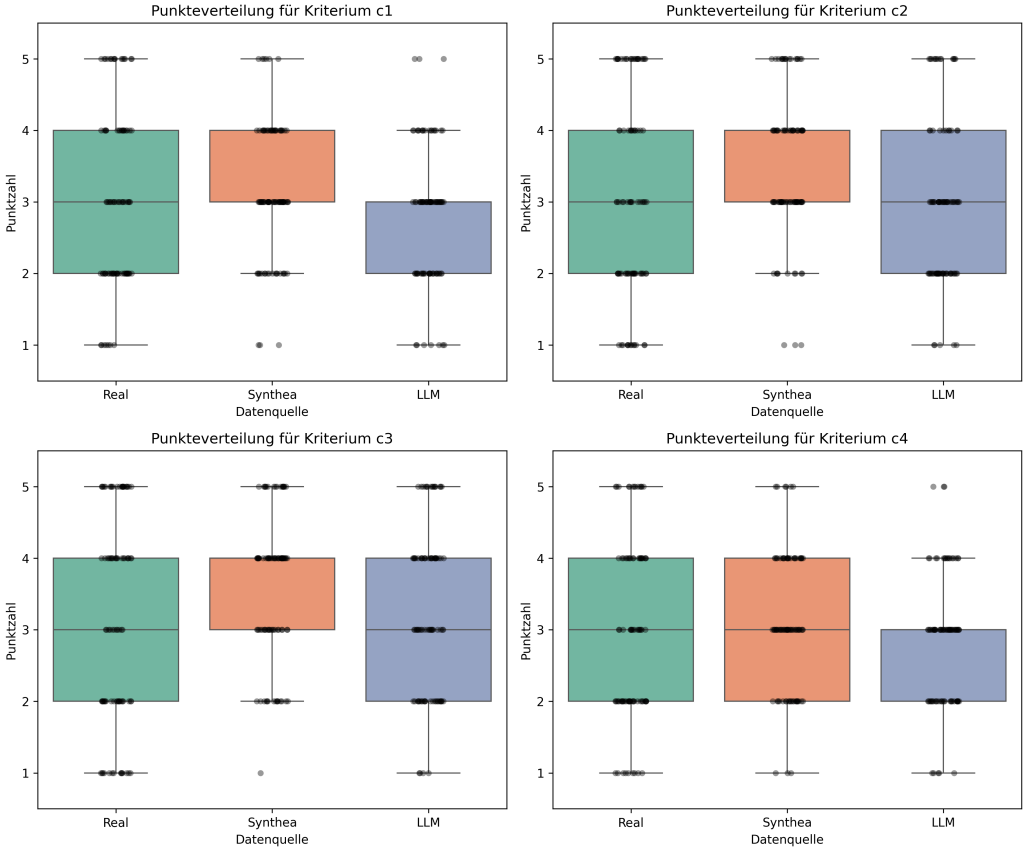


Abbildung A.4.: Verteilung der quantitativen Bewertungen (Scores 1 bis 5) für die vier Kriterien (c1–c4), aufgeschlüsselt nach den Datenquellen (Real, Synthea, LLM).

Beantwortung

Es gibt statistisch signifikante Unterschiede zwischen den generierten Patientendaten, allerdings beschränken sich diese auf das Kriterium c2. Wie sowohl der statistische Test als auch die grafische Darstellung (A.3) aufzeigen, schneiden hierbei die regelbasiert generierten Patientendaten signifikant besser ab als die LLM-generierten Daten (sowie tendenziell besser als die echten Patientendaten). Bei den restlichen Kriterien (c1, c3, c4) ließen sich keine statistisch signifikanten Abweichungen zwischen den Datenquellen nachweisen. Dies deutet darauf hin, dass die generierten Daten in diesen Aspekten statistisch vergleichbar zu den echten Daten bewertet wurden, wenngleich die in der Grafik ersichtliche, breite Verteilung von positiven und negativen Scores innerhalb der jeweiligen Gruppen auf eine generelle Uneinigkeit der Bewerter*innen hindeutet (niedrige Inter-Rater-Reliabilität).

Frage 4.5: Qualitative Auffälligkeiten zwischen den Patientengruppen

Methodik

Ergänzend zur rein quantitativen Auswertung wurden die optionalen qualitativen Freitext-Kommentare der drei medizinischen Evaluator*innen systematisch analysiert. Ziel war es, wiederkehrende inhaltliche Muster, klinische Inkonsistenzen sowie spezifische Schwächen der jeweiligen Datenquellen zu identifizieren. Hierfür wurden die Rückmeldungen manuell gesichtet, thematisch gruppiert und den drei Patientengruppen zugeordnet, um die Ursachen für die quantitativen Bewertungen und die stark variierende Punktevergabe fachlich fundiert zu begründen.

Ergebnisse

Die Auswertung der Kommentare hat gezeigt, dass alle drei Datenquellen mit sehr spezifischen, systemischen Problemen behaftet sind, die von den Expert*innen stark kritisiert wurden:

- **Echte Daten (Real):** Trotz ihrer Authentizität haben die echten Akten massive Defizite in der Lesbarkeit. Ein Hauptkritikpunkt war die zeitliche Kompression (z. B. Datierungen sämtlicher Befunde, Labore und Diagnosen auf denselben Tag), wodurch klinische Verläufe kaum nachvollziehbar waren. Zudem erschwerten unspezifische Diagnoseschlüssel (z. B. „NOS“) und klinikinterne Abkürzungen ohne Kontext die Bewertung erheblich.
- **Synthea:** Obwohl diese Akten strukturell sehr sauber und konsistent wirkten, fielen sie durch algorithmische Stereotypen auf (z. B. eine unnatürliche Überrepräsentation zahnärztlicher Diagnosen wie Gingivitiden). Ein noch größeres Problem stellte das Fehlen klinischer Logik dar: Beispielsweise wurden in Schwangerschaften streng kontraindizierte Medikamente verordnet, oder chronische Diagnosen wie eine Anämie blieben völlig unbehandelt im Raum stehen.
- **LLM-generierte Daten (LLM):** Diese Datensätze schnitten qualitativ am schlechtesten ab. Kritisiert wurde primär die klinische Unplausibilität der Behandlungsverläufe (z. B. Einleitung einer Metformin-Therapie erst drei Jahre nach einer Diabetes-Diagnose oder die unglaubwürdige Monotherapie eines Schlaganfalls mit Aspirin). Zudem neigten die Sprachmodelle zu Halluzinationen bei Laborwerten (falsche Einheiten) und bauten, ähnlich wie bei den echten Daten, oft keinen plausiblen zeitlichen Behandlungsstrahl auf.

Beantwortung

Die qualitative Analyse beantwortet die Frage, warum die Datensätze unterschiedlich bewertet wurden und erklärt zeitgleich die sehr niedrige Inter-Rater-Reliabilität. Jede Datenquelle weist spezifische Schwächen auf: Echte

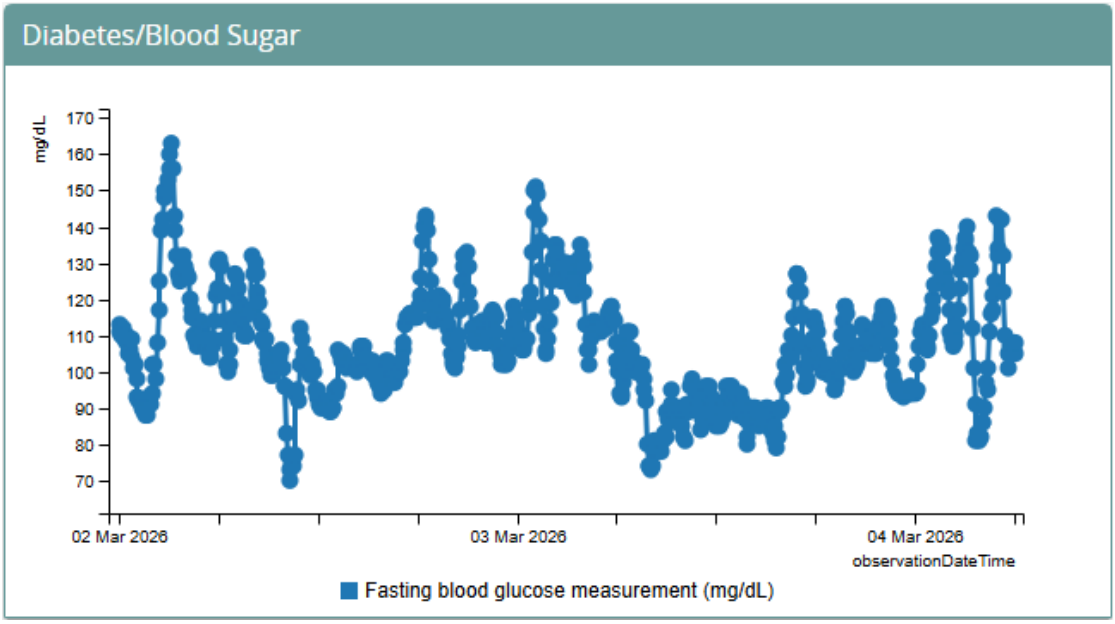


Abbildung A.5.: Die Visualisierung der Sensordaten im Patientendashboard von Bahmni.

Patientendaten leiden an einer chaotischen, oft abrechnungsgetriebenen Struktur, Synthea-Daten zeigen trotz struktureller Brillanz gravierende Defizite im medizinischen Kontextwissen und LLM-generierte Patientendaten versagen bei harten klinischen Fakten (wie Maßeinheiten und Leitlinien). Die Bewerter*innen legten beim Evaluieren subjektiv unterschiedliche Schwerpunkte, was dazu führte, dass je nach ärztlichem Fokus strukturelles Chaos (Real), fehlender klinischer Sachverstand (Synthea) oder fachliche Halluzinationen (LLM) unterschiedlich hart abgestraft wurden.

Frage 4.6: Qualität der Visualisierung der Sensordaten

Methodik

Die Qualität der Visualisierung der Sensordaten im KIS wurde anhand der vier definierten Qualitätskriterien bewertet. Es wurde analysiert, welche der Kriterien bereits nativ im Patientendashboard von Bahmni vorhanden sind

und welche noch nachgebessert werden müssen.

Ergebnisse

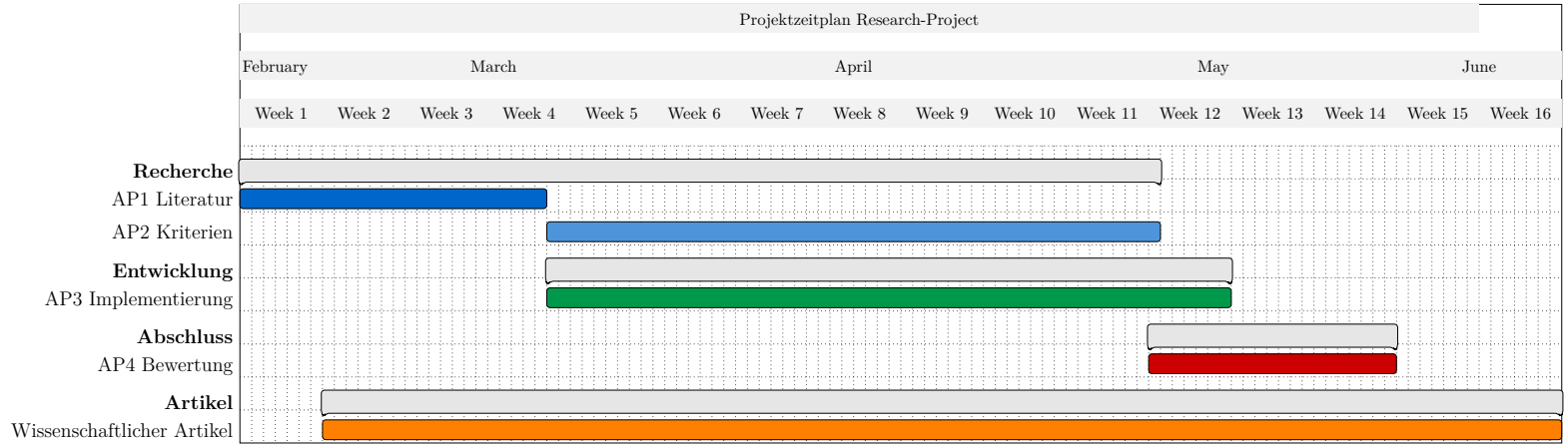
In Abbildung A.5 ist die Visualisierung der Glucose-Daten über drei Tage dargestellt. Anhand dieser Visualisierung wurden für die Qualitätskriterien folgende Ergebnisse festgestellt:

- **Grafische Trendvisualisierung:** Die Daten werden von Bahmni nativ als Graph angezeigt. Damit wird dieses Kriterium erfüllt.
- **Adaptive Granularität:** In der Oberfläche an sich kann die Granularität nicht verändert werden. Man muss vor dem Import festlegen, in welcher Granularität die Daten in der Datenbank abgespeichert werden sollen. Zudem kann man die Anzahl der angezeigten Tage, in diesem Fall drei, nur in den Konfigurationsdateien und nicht in der Benutzeroberfläche verändern. Dieses Kriterium ist demnach nicht erfüllt.
- **Hervorhebung von Anomalien:** Bei korrekt eingelesenen Laborwerten zeigt Bahmni Anomalien an, indem es diese farblich hervorhebt. Allerdings scheint die Bewertung, ob ein Laborwert als abnormal gilt, in der Datenbank festgelegt zu werden und wird nicht erst bei der Visualisierung ermittelt. Denn die über FHIR eingelesenen Glucosewerte haben Einträge, die außerhalb der spezifizierten Normen liegen und sie werden dennoch nicht markiert. Damit ist dieses Kriterium teilweise erfüllt, da die generelle Infrastruktur zur Hervorhebung von Anomalien vorhanden ist, aber es sich mit der aktuellen Import-Strategie nicht auf die importierten Werte anwenden lässt.
- **Zeitliche Ereignisannotation:** Dieses Kriterium wird nicht nativ in Bahmni unterstützt.

Beantwortung

Bahmni erfüllt nativ die Anforderung, dass die Daten als Graph und nicht nur als Tabelle angezeigt werden. Zudem besteht die Möglichkeit, Anomalien darzustellen, allerdings nur, wenn es möglich ist, die Werte als abnormal in der Datenbank mittels FHIR Import zu markieren. Die adaptive Granularität in der Benutzeroberfläche und die zeitliche Ereignisannotation werden aktuell nicht von Bahmni nativ unterstützt und müssten noch entwickelt werden.

A.4. Zeitplan


Projektzeitplan Research-Project
February
March
April
May
June
Week 1
Week 2
Week 3
Week 4
Week 5
Week 6
Week 7
Week 8
Week 9
Week 10
Week 11
Week 12
Week 13
Week 14
Week 15
Week 16
Recherche
AP1 Literatur
AP2 Kriterien
Entwicklung
AP3 Implementierung
Abschluss
AP4 Bewertung
Artikel
Wissenschaftlicher Artikel