

MIDOCEAN UNIVERSITY
College of Management
Master of Health Administration



"A Research Proposal Titled"
Analyzing Data Silos Challenges in Public Hospitals in Saudi Arabia and
Proposing AI-Based Solutions Using Open Health Data"

Prepared by the Student:

Osama Jehad Alostaz

Supervisor

Dr. AL Faisal Abdelhamid Mohamed

Dr. Fayez Ali Jrad

Academic Year: 2025/2026

All printing rights are reserved by the university.

Introduction

Saudi Arabia has taken significant strides toward digital transformation in healthcare under Vision 2030, with large-scale investments in electronic health records (EHRs), telemedicine platforms, and national health data infrastructures. The Ministry of Health (MOH) manages over 287 public hospitals with approximately 45,180 beds, in addition to 167 private hospitals with 19,427 beds—demonstrating the scope and complexity of the Kingdom's healthcare system (JMIR, 2024).

Despite these developments, public hospitals continue to face critical challenges related to health information interoperability. Interoperability readiness scores remain uneven, with government hospitals averaging 77 out of 100, compared to 83 in private facilities (PMC, 2022). These gaps reflect the persistent issue of data silos, where fragmented health information systems prevent efficient data exchange between departments and across institutions. As a result, healthcare providers often encounter duplicated diagnostic tests, delayed clinical decisions, and difficulty accessing comprehensive patient histories (Wang et al., 2018; Kankanhalli et al., 2016).

Studies have identified several root causes of these silos in Saudi Arabia, including the lack of a unified health information exchange framework, insufficient standardization of data protocols, and technical barriers to system integration (Alkraiiji et al., 2016). These challenges have constrained the effective use of digital tools for care coordination, population health monitoring, and real-time decision support.

However, emerging artificial intelligence (AI) technologies offer a promising avenue for addressing data fragmentation. By leveraging advanced techniques such as federated learning and natural language processing (NLP), AI systems can

analyze distributed datasets, extract clinical insights, and enable predictive analytics without compromising patient privacy (Topol, 2019; Esteva et al., 2021). In this context, national digital health initiatives—such as Seha Virtual Hospital (SVH)—have already demonstrated the effectiveness of AI-powered tools in reducing hospital admissions for chronic conditions by up to 33% (Healthpoint, 2024).

Moreover, Saudi Arabia possesses a growing ecosystem of open health data made available by agencies like MOH, SFDA, CHI, and the Saudi Health Council. Initiatives such as the Unified Health File (Nafis), covering over 21.5 million patients and supporting more than 40 data-sharing agreements, provide a valuable yet underutilized resource for building interoperable, AI-driven health systems (SHC, 2023).

Problem Statement

Despite extensive efforts to digitize the healthcare sector in Saudi Arabia, data silos remain a persistent obstacle to achieving integrated, patient-centered care in public hospitals. These silos stem from fragmented and isolated information systems, limited interoperability across departments, and the absence of a unified, standardized health data exchange framework (Wang et al., 2018).

Recent national evaluations show that approximately 62% of public hospitals in Saudi Arabia lack full integration between their EHR systems and other core platforms such as laboratory, radiology, and pharmacy systems (PMC, 2022). Furthermore, only 29% report having centralized infrastructure for real-time data sharing and clinical decision-making (JMIR, 2024). This fragmentation contributes to increased medical errors, duplicated diagnostics, higher operational costs, and diminished quality of care.

Complicating the issue further is the inconsistent implementation of national initiatives such as the Nafis program and Seha Virtual Hospital. While these platforms represent key steps toward system-wide integration, their adoption across the public hospital network remains limited and uneven (SHC, 2023).

Additionally, the Kingdom's valuable repositories of open health data—such as MOH registries, SFDA product databases, and CHI insurance datasets—are not fully leveraged to build intelligent health systems capable of predictive modeling and automated support. Despite their availability, there is a clear gap in both research and practical application in utilizing these resources to overcome data silos through AI-powered strategies (Statista, 2025; Healthpoint, 2024).

This study therefore aims to investigate the structural and organizational causes of health data silos in public hospitals and to propose AI-based integration solutions using publicly available health data. By doing so, the research seeks to contribute to the national vision of a more connected, efficient, and intelligent healthcare ecosystem in Saudi Arabia.

Significance of the Study

This study holds substantial theoretical and practical significance in the intersecting fields of health informatics, artificial intelligence, and digital health transformation. Theoretically, it contributes to the limited body of literature addressing data silos within the public healthcare sector in Saudi Arabia—a context that remains underexplored despite increasing national investment in health digitization. While global research has addressed information fragmentation in healthcare, few studies have examined how such challenges manifest in Saudi Arabia's unique regulatory, infrastructural, and institutional environment. By applying frameworks such as Systems Theory, Information Integration Theory, and Diffusion of Innovation

Theory, the study provides a contextualized understanding of how data silos emerge, persist, and obstruct efficient healthcare delivery in public hospitals.

Practically, the study responds to a critical and timely national priority: improving data-driven decision-making in public healthcare institutions under Saudi Arabia's Vision 2030. As the Ministry of Health (MOH), the Saudi Food and Drug Authority (SFDA), and the Saudi Health Council (CHI) continue to release open health data, there exists an untapped opportunity to harness these resources through AI-driven models—particularly federated and privacy-preserving approaches. The study offers actionable insights and scalable AI-based solutions to address interoperability gaps, enhance clinical outcomes, and optimize operational workflows. By focusing on public hospitals, which serve a large portion of the population and often operate under the greatest integration constraints, the research directly aligns with national health equity objectives and the Kingdom's long-term sustainability goals. Ultimately, the study aims to support the development of a more intelligent, connected, and patient-centered healthcare ecosystem in Saudi Arabia.

Research Objectives

Main Objective: To analyze the causes of data silos in Saudi public hospitals and propose AI-based solutions using open health data to enhance data integration and information flow.

Sub-Objectives:

1. To examine the nature and distribution of data silos within public hospitals, focusing on technological, administrative, and organizational dimensions.
2. To assess the availability, structure, and accessibility of open health data in Saudi Arabia for research and healthcare integration purposes.

3. To identify global best practices and AI methodologies for integrating fragmented healthcare data and improving clinical decision-making.
4. To develop an AI-based framework or model leveraging Saudi Arabia's open health datasets to address data silos and support digital health transformation.
5. To evaluate the alignment of the proposed solutions with Saudi Arabia's Vision 2030 objectives, particularly regarding healthcare efficiency, digital innovation, and equitable access to health services.

Research Questions

This study seeks to address the following key research questions:

1. **What are the main causes and consequences of data silos** in public hospitals in Saudi Arabia?
2. **To what extent do technological, administrative, and regulatory factors** contribute to the persistence of fragmented healthcare information systems?
3. **What types of open health data are available in Saudi Arabia**, and how accessible and usable are they for research and integration purposes?
4. **How can artificial intelligence (AI) be leveraged to analyze and integrate fragmented health data** across public healthcare institutions?
5. **What are the global best practices in using AI to address data silos** in healthcare systems, and how can these be adapted to the Saudi context?
6. **How well do AI-based integration solutions align with Saudi Arabia's Vision 2030 objectives**, particularly in terms of healthcare efficiency, innovation, and digital transformation?

Research Hypotheses

Based on the study's objectives and questions, the following hypotheses are proposed:

1. **H1:** There is a statistically significant relationship between the presence of data silos and the inefficiency of clinical and administrative operations in public hospitals in Saudi Arabia.
2. **H2:** The lack of standardized interoperability protocols significantly contributes to the persistence of fragmented health information systems.
3. **H3:** The availability and quality of open health data in Saudi Arabia are positively associated with the feasibility of implementing AI-based integration models.
4. **H4:** AI-powered solutions can significantly reduce the impact of data silos by enhancing data accessibility, accuracy, and integration across public healthcare institutions.
5. **H5:** AI-based integration frameworks that utilize open health data contribute positively to the achievement of Vision 2030 objectives, particularly in terms of healthcare efficiency and digital innovation.

Previous Studies

1. Al-Kahtani et al. (2022) – Digital health transformation in Saudi Arabia: A cross-sectional analysis using HIMSS digital health indicators

This study aimed to evaluate the level of digital health transformation in healthcare facilities across the Eastern Province of Saudi Arabia. The authors applied a cross-sectional methodology using the Healthcare Information and Management Systems Society (HIMSS) digital health indicator framework, which measures four

dimensions: interoperability, governance and workforce, person-enabled health, and predictive analytics. Data were collected from 10 healthcare institutions, including both public and private hospitals, through an online structured survey. The findings revealed that governance and workforce scored the highest (median = 80), while interoperability received a median score of 78.5. Notably, predictive analytics had the lowest readiness score across institutions. Moreover, public hospitals showed lower digital readiness compared to their private counterparts (median score of 71 vs. 77). The study concluded that interoperability remains a significant challenge for Saudi hospitals, hindering effective data integration across the healthcare system.

2. Al-Shamrani et al. (2024) – A Secure Federated Learning Approach for Detecting COVID-19 from Medical CT Images

This study proposed a federated learning-based artificial intelligence model designed to detect COVID-19 infections using computed tomography (CT) images while preserving data privacy across participating Saudi hospitals. The researchers adopted a decentralized model training process, where each hospital locally trained its machine learning model without sharing actual patient data. Only encrypted model updates were aggregated centrally to form a robust global model. The aim was to reduce the risk of data breaches and prevent the formation of data silos. The model demonstrated high classification accuracy (around 97.3%) in diagnosing COVID-19 cases, surpassing conventional centralized machine learning models. The study's outcomes emphasized the applicability of federated learning in enhancing collaborative medical diagnostics across isolated healthcare systems without compromising patient confidentiality.

3. Alasmari et al. (2025) – Federated learning-based multimodal approach for early detection and personalized care in cardiac disease

This research introduced an innovative multimodal federated learning framework for early cardiac disease detection and tailored care delivery in Saudi Arabia. The authors integrated diverse data sources including cardiac imaging, electrocardiograms (ECG), clinical history, and dietary patterns into a unified machine learning model using an attention-based fusion approach. Each hospital trained its own local deep learning model (SGD-DNN), and model parameters were securely exchanged and aggregated, ensuring data privacy. The proposed system achieved remarkable classification accuracies: 97.76% on dataset 1, 98.43% on dataset 2, and 99.12% on dataset 3. The research highlighted the framework's effectiveness in maintaining data confidentiality while providing precise diagnostic predictions. The study concluded that such AI-powered, privacy-preserving approaches could significantly transform clinical decision-making in cardiac care, especially in fragmented health systems like those in Saudi Arabia.

4. Nguyen et al. (2021) – Federated learning in healthcare: A survey

Nguyen and colleagues conducted a comprehensive survey on the application of federated learning (FL) in the healthcare domain. The study aimed to understand how FL can address critical data privacy challenges and break down data silos across hospitals and research institutions. The authors reviewed over 120 articles and real-world FL implementations in medical imaging, electronic health records (EHRs), genomics, and mobile health. Methodologically, the study adopted a qualitative synthesis approach. It concluded that FL offers substantial promise in preserving patient data confidentiality while enabling collaborative AI model development. However, challenges remain, including data heterogeneity, limited communication infrastructure, and regulatory barriers. The study emphasized that achieving

trustworthy and scalable FL systems in healthcare requires joint efforts in algorithm design, policy frameworks, and technical standardization.

5. Mohsen et al. (2022) – AI integration of medical imaging and EHR for clinical decision support

Mohsen and co-authors investigated the integration of artificial intelligence (AI) systems that combine medical imaging data with electronic health records to enhance clinical decision-making processes. The study employed a multi-layered convolutional neural network trained on a hybrid dataset combining radiology reports and structured EHR fields. The objective was to improve diagnosis accuracy and reduce diagnostic delays. The study demonstrated that combining data sources significantly improved model performance—achieving an AUC of 0.93 compared to 0.81 using imaging alone. It also reduced false positives in disease detection. The authors concluded that data integration, when enabled by AI, can overcome the limitations of siloed healthcare data and lead to safer and faster diagnostic support.

6. Liu et al. (2018) – FADL: Federated-Autonomous Deep Learning for distributed EHR data analysis

Liu and colleagues developed the Federated-Autonomous Deep Learning (FADL) framework to analyze intensive care unit (ICU) data collected from 58 hospitals without centralizing patient records. The study adopted a two-phase training approach where the initial layers of the deep learning model were trained collaboratively, while the final layers were fine-tuned locally at each hospital. The primary goal was to preserve data privacy while ensuring model performance. Results indicated that the FADL model achieved an AUC of 0.78 in mortality prediction, outperforming both isolated and naive federated models. The study

confirmed the value of balancing shared and local learning to handle distributed healthcare data and mitigate institutional data silos.

7. Yang et al. (2022) – Data harmonization in health systems: Techniques and challenges

Yang et al. explored the importance of data harmonization as a foundational step toward breaking down data silos in health information systems. The study categorized harmonization approaches into three types: syntactic, semantic, and process-level. Through case studies from multiple health systems, the study illustrated how inconsistent coding, incompatible formats, and unaligned metadata hinder interoperability. It used mixed-method analysis combining system architecture evaluation and expert interviews. The authors concluded that standardization initiatives such as HL7 FHIR and openEHR can significantly enhance data sharing but require institutional commitment and technical capacity. Data harmonization was identified as the “missing bridge” between healthcare silos and integrated analytics.

8. Wunker (2023) – Why AI fails without integration: Breaking data silos in healthcare

In this thought leadership article, Wunker highlighted why many AI initiatives in healthcare fail due to fragmented data infrastructures. Drawing on real-world case studies and organizational research, the author argued that without resolving interoperability issues, even the most advanced AI models remain underutilized. The article identified three major sources of silos: lack of common data standards, poor governance policies, and institutional resistance. It further emphasized the role of leadership in driving integration strategies. The key conclusion was that AI is only

as effective as the ecosystem that supports it—thus, breaking silos must be a foundational priority for digital health leaders.

9. Juneja (2025) – AI bias and the hidden dangers of fragmented health data

Juneja’s study focused on how fragmented datasets in healthcare settings can introduce bias into AI models. The paper reviewed over 30 AI diagnostic models and identified that those trained on non-representative or siloed datasets produced skewed predictions, particularly across minority groups. The methodology included bias testing metrics such as disparate impact and equalized odds. The findings revealed that data fragmentation was directly correlated with reduced model fairness. The study recommended building integrated and diverse health data ecosystems to ensure equitable AI outcomes and to comply with ethical AI guidelines, especially in developing health systems.

10. SS&C Blue Prism HealthTech Survey (2025) – The state of AI adoption in global healthcare systems

This global survey collected responses from 297 healthcare organizations across 16 countries to assess readiness for AI adoption in clinical and administrative tasks. One of the core insights was that although 94% of respondents recognized AI as a strategic priority, only 52% had systems capable of cross-departmental data sharing. The study used structured questionnaires and statistical analysis to correlate data integration maturity with successful AI use cases. Organizations with strong interoperability frameworks were 3.4 times more likely to report improved patient outcomes. The survey underscored that overcoming data silos is not merely a technical challenge but a strategic enabler for digital health transformation.

Commentary on the Reviewed Literature

The reviewed body of literature reflects a growing academic and professional interest in the challenges and opportunities associated with data silos and digital integration in healthcare systems, both globally and within the Saudi Arabian context. Several themes emerge across the studies.

Firstly, a significant number of international studies—such as those by Nguyen et al. (2021), Liu et al. (2018), and Yang et al. (2022)—have emphasized the technical and infrastructural barriers to health data interoperability, often proposing federated learning and AI-based architectures as feasible solutions. These works provide a solid foundation for understanding the role of advanced technologies in overcoming fragmentation, especially in environments where patient privacy and institutional autonomy are paramount.

Secondly, the Saudi-specific studies—such as Al-Kahtani et al. (2022), Alkraihi et al. (2016), and SHC (2023)—highlight unique challenges within the Kingdom’s healthcare ecosystem. These include the lack of standardized data protocols across institutions, delayed digital transformation in public hospitals, and underutilization of available open data resources provided by national authorities like the MOH, SFDA, and CHI. This context-specific insight is crucial, as it underlines how global solutions must be localized to be effective in Saudi Arabia.

Thirdly, a recurring gap observed across several studies is the limited attention to the practical implementation of AI in real-time hospital settings, particularly in the public sector. While many studies provide theoretical frameworks or simulations, there remains a shortage of applied research that combines stakeholder insights, policy review, and real-world data to develop scalable solutions. Additionally, the social and ethical dimensions of AI—such as algorithmic bias and the digital

divide—are acknowledged but not thoroughly investigated in many of the sources reviewed.

Another important observation is that while federated learning and decentralized data models are frequently suggested, few studies demonstrate how these technologies can be effectively integrated with open government health datasets in Saudi Arabia. This represents a major research gap that the current study seeks to address, by applying a mixed-methods approach grounded in national open data and aligned with Vision 2030's goals for digital health transformation.

In conclusion, the existing literature provides a rich but fragmented base of knowledge. This study builds upon and extends previous work by focusing specifically on public hospitals in Saudi Arabia, using both qualitative and quantitative methods to propose AI-driven solutions for data silos. It distinguishes itself by grounding the proposed solutions in real, openly available datasets and policy frameworks unique to the Saudi health system.

Theoretical Foundations of the Study

This study is based on several interrelated theoretical frameworks that explain both the persistence of data silos in healthcare systems and the potential of artificial intelligence (AI) and open government health data to address them.

1. Systems Theory Systems Theory views hospitals as complex and interdependent organizations, where fragmentation across departments, units, and information systems creates inefficiencies and communication breakdowns. In this context, data silos act as structural barriers that hinder coordination across the system. This perspective is grounded in the work of von Bertalanffy (1968), emphasizing that understanding hospitals as systems helps identify how structural fragmentation affects overall performance and care delivery.

2. Information Integration Theory Information Integration Theory emphasizes that effective decision-making in healthcare depends on combining data from multiple sources, both structured (e.g., electronic health records) and unstructured (e.g., clinical notes, radiology images). When data is isolated or inconsistently formatted, its analytical value decreases, increasing the likelihood of clinical errors. This theory justifies the focus on interoperability and the consolidation of fragmented health information sources.

3. Diffusion of Innovation Theory Diffusion of Innovation Theory (Rogers, 2003) explains differences in the adoption of AI technologies across healthcare institutions. It highlights that adoption is influenced by factors such as perceived relative advantage, complexity, compatibility with existing workflows, and organizational readiness. This framework helps understand why hospitals may embrace or resist AI-based integration tools like federated learning or NLP systems.

4. Data Governance Theory Data Governance Theory emphasizes the need for clear policies, defined responsibilities, and ethical frameworks in managing digital health assets. This theory is particularly relevant when studying the use of open health data from entities such as the Ministry of Health (MOH), the Saudi Food and Drug Authority (SFDA), and the Saudi Health Council (CHI), as these organizations require strict compliance with data privacy and usage regulations.

5. AI Ethics and Explainable AI (XAI) Finally, principles from AI Ethics and the emerging field of Explainable AI (XAI) stress the importance of transparency, fairness, and accountability in AI deployment. These frameworks ensure that AI-based solutions for breaking down data silos not only improve efficiency but also maintain public trust and adhere to regulatory standards.

Table 1: Theoretical Foundations of the Study – Theories, Founders, Core Hypotheses, and Scientific Justification

Theory	Founder	Core Hypothesis	Scientific Justification for the Study
Systems Theory	von Bertalanffy (1968)	Organizations are interconnected systems whose components interact with each other	Helps understand hospitals as complex systems and explains how fragmentation across departments and units leads to weak coordination and operational inefficiency.
Information Integration Theory	–	Effective decision-making depends on integrating data from multiple sources	Justifies focusing on the integration of structured and unstructured data and illustrates how data silos affect analytical quality and clinical decision-making.
Diffusion of Innovation Theory	Rogers (2003)	The rate of innovation adoption depends on the characteristics of the innovation and its fit with the organizational context	Helps explain hospital adoption of modern technologies like AI and clarifies reasons for slow uptake or resistance to digital solutions.
Data Governance Theory	Khatri & Brown (2010)	Data management requires clear policies and defined responsibilities	Justifies examining frameworks and policies needed to handle open data while ensuring privacy and regulatory compliance.
AI Ethics and Explainable AI (XAI)	Gunning (2017)	AI applications must be transparent, fair, and accountable	Supports ensuring that proposed AI solutions align with ethical and legal

			standards when used in the public healthcare sector.
--	--	--	--

Theoretical Framework of the Study

The theoretical framework provides the structure upon which the researcher builds their understanding of the study topic. It clarifies key concepts and related variables and is based on knowledge frameworks that support the study hypotheses. This study aims to analyze the causes of health data fragmentation in public hospitals and explore the role of open government health data and artificial intelligence (AI) in addressing this fragmentation, enhancing data integration, and improving decision-making.

1. Data Silos

Definition: Isolated data repositories within healthcare institutions that are not easily accessible or shareable across departments or entities.

Importance: They slow down decision-making processes, weaken coordination among departments, and reduce the effectiveness of healthcare delivery.

Types of Data Silos:

- Organizational Silos: Data divided according to departments or units.
- Technical Silos: Use of incompatible information systems or different data standards.
- Legal/Regulatory Silos: Restrictions on data sharing or privacy policies.

Steps to Address Data Silos:

1. Assess the level of data fragmentation in the institution.
2. Identify the causes of isolation (structural, technical, regulatory).

3. Propose integration mechanisms, such as adopting standardized data formats or utilizing AI technologies.

2. Open Government Health Data

Definition: Publicly available datasets issued by government agencies, such as the Ministry of Health, including statistics on health facilities, medical records, and public health surveillance.

Importance: They provide opportunities for transparent data analysis and evidence-based decision-making.

Types of Open Data:

- Health facility and capacity data.
- Medical records and clinical statistics.
- Regulatory and policy-related data.

Steps for Utilization:

1. Access data through government portals.
2. Verify data accuracy and completeness.
3. Integrate it with internal hospital data using analytical tools.

3. Artificial Intelligence in Healthcare

Definition: Computational techniques capable of analyzing data, learning from it, and making intelligent decisions similar to human capabilities.

Importance of AI: It addresses the challenges of data fragmentation without compromising privacy and facilitates integration across multiple data sources.

AI Tools Used in the Study:

- **Federated Learning:** Training machine learning models on distributed data without transferring it, thus preserving data privacy.
- **Natural Language Processing (NLP):** Extracting and analyzing unstructured data such as medical notes and regulatory documents.

Linking the Theoretical Framework to Research Variables

- **Independent Variable:** Health data fragmentation in public hospitals.
- **Mediating Variable:** Utilization of open government health data.
- **Dependent Variable:** Data integration and improved decision-making using AI.

The theoretical framework illustrates how AI can mitigate the effects of data fragmentation by integrating distributed and open data while adhering to governance and ethical standards, thereby enhancing the effectiveness of decision-making within public hospitals.

Research Methodology and Methods

This study adopts a mixed-methods research design, combining both quantitative and qualitative approaches to provide a comprehensive and nuanced understanding of data silos in Saudi Arabia's public hospitals. It also aims to explore AI-based strategies for integrating health information systems using open datasets published by the Saudi Ministry of Health (MOH).

The quantitative phase involves the distribution of a structured questionnaire to IT professionals, clinical data officers, and administrative personnel working in public hospitals across different regions. The survey is designed to assess several dimensions, including the degree of data fragmentation, perceptions of interoperability barriers, familiarity with MOH open data resources, and

organizational readiness for AI integration. Responses will be collected using Likert-scale and closed-ended questions. Statistical analyses—including descriptive statistics, correlation analysis, and regression modeling—will be performed using SPSS to identify significant patterns and relationships.

The qualitative phase consists of semi-structured interviews with key stakeholders such as policymakers from the Ministry of Health, health informatics specialists, and hospital CIOs. These interviews aim to explore institutional, technical, and regulatory factors that contribute to data silos, as well as perspectives on the applicability and limitations of AI in public healthcare settings. Interview transcripts will be thematically analyzed using NVivo or a similar tool to extract recurring themes and contextual insights.

In parallel, a secondary data analysis will be conducted using publicly available datasets from the MOH Open Data Portal. These datasets will serve as the foundation for modeling a proof-of-concept for AI-driven integration solutions. Specifically, the study will explore the feasibility of implementing federated learning algorithms and natural language processing techniques on anonymized health data, while maintaining compliance with privacy and security standards.

By triangulating evidence from surveys, interviews, and real-world public data, this mixed-methods approach enhances the validity and depth of the findings. It is particularly well-suited to address the complex technological and policy-related challenges associated with health information fragmentation in Saudi Arabia's public healthcare sector.

Research Instruments

This study utilizes a combination of **three primary research instruments**, each serving a distinct purpose aligned with the mixed-methods approach:

1. Structured Questionnaire

A structured questionnaire will be administered to IT personnel, clinical data officers, and administrative staff working in public hospitals across different regions of Saudi Arabia. The instrument is designed to assess:

- The existence and extent of data silos within hospital departments
- Challenges related to data interoperability, system integration, and access restrictions
- Organizational attitudes toward the adoption of artificial intelligence
- Awareness and perceptions of national health data strategies (e.g., MOH digital initiatives)
- The questionnaire will consist of Likert-scale and multiple-choice items. A pilot version will be tested to ensure reliability and internal consistency using Cronbach's Alpha.

2. Semi-Structured Interview Guide

A semi-structured interview guide will be used to conduct in-depth interviews with a selected group of stakeholders, including Ministry of Health officials, hospital CIOs, and data scientists involved in digital transformation. The guide will cover topics such as:

- Structural and policy-related causes of data silos
 - Readiness for AI deployment in public hospital systems
 - Perceptions of existing digital infrastructure and challenges in data sharing
 - Expectations for future AI-based integration efforts
- Thematic analysis will be conducted on the interview transcripts to extract key insights.

3. **Secondary Data Analysis Tool (MOH Public Datasets)** The third instrument involves the use of **official datasets published by the Saudi Ministry of Health**, including:

- Hospital performance reports
- Bed capacity and staffing statistics
- Regional distribution of healthcare services
- Health informatics readiness indicators

These datasets will be used for descriptive and comparative analysis to identify patterns of fragmentation and disparities in data availability across institutions. The findings will also serve as the foundation for simulating hypothetical AI-based integration models using anonymized data structures.

By triangulating data from surveys, interviews, and official datasets, the study aims to produce a well-rounded, evidence-based understanding of data silo challenges and practical AI-driven solutions within Saudi Arabia's public healthcare system.

Delimitations and Limitations of the Study

This study, while comprehensive in its scope and methodological design, is subject to several **delimitations** and **limitations** that must be acknowledged.

Delimitations refer to the boundaries intentionally set by the researcher to narrow the focus of the study. This research is delimited to:

- **Public hospitals** under the administration of the Saudi Ministry of Health (MOH), thereby excluding private healthcare institutions, university hospitals, and military medical facilities.

- **Key stakeholders** such as hospital IT staff, clinical data officers, and policy decision-makers, rather than frontline healthcare providers or patients.
- **National-level data sources**, particularly those published by the MOH, which limits the analysis to datasets officially released and available at the time of the study.

Limitations, on the other hand, are factors beyond the researcher's control that may impact the generalizability or interpretation of findings:

- The **availability and completeness of secondary data** from MOH may restrict the depth of certain analyses, especially if some datasets are outdated or lack granularity at the departmental level.
- **Access constraints** to participants in high-level administrative or policy roles could limit the diversity of qualitative insights.
- The **simulation of AI-based models** will be conducted on publicly available or representative data structures, not on real-time hospital data, which may reduce the precision of technical feasibility evaluations.
- Additionally, given the rapidly evolving nature of Saudi Arabia's digital health transformation under Vision 2030, some findings may be time-sensitive and subject to change as new technologies or regulations are introduced.

Despite these limitations, every effort will be made to ensure methodological rigor and data triangulation to strengthen the credibility and relevance of the research outcomes.

Proposed Structure of the Study

This research will be structured into **four main chapters**, each contributing to a systematic exploration of the problem of data silos in Saudi public hospitals and the development of AI-based solutions using official health datasets.

Chapter One: Introduction

This chapter provides the foundation of the study. It begins by presenting the background of the research problem, focusing on the persistence of data silos in public hospitals and the growing national interest in digital health transformation under Saudi Arabia's Vision 2030. The chapter clearly states the problem, defines the research objectives, and outlines key questions and hypotheses. It also discusses the theoretical and practical significance of the study, clarifies the delimitations and limitations, and defines relevant concepts. The chapter concludes with a brief overview of the study structure.

Chapter Two: Literature Review and Theoretical Framework

The second chapter integrates a comprehensive review of prior research and relevant theories. It explores global and local studies that have addressed data fragmentation, health information system challenges, and the role of artificial intelligence in overcoming interoperability issues. Special emphasis is placed on the Saudi context and the availability of national health data. In addition, this chapter introduces the theoretical foundations guiding the study, including Systems Theory and Diffusion of Innovation Theory. A conceptual framework is developed to illustrate the relationships between the study variables—data silos, organizational readiness, AI strategies, and expected outcomes.

Chapter Three: Methodology and Tools of the Study

This chapter explains the methodological approach used in the study, which follows a **mixed-methods design**. It details the quantitative component (structured survey), the qualitative component (semi-structured interviews), and the use of official datasets from the Ministry of Health for simulation or analysis. The chapter describes the study population, sampling methods, research instruments, data collection procedures, and data analysis techniques (both statistical and thematic). Ethical considerations and the reliability of instruments are also addressed to ensure the credibility and validity of the study.

Chapter Four: Results, Discussion, and Recommendations

The final chapter presents the findings from the data collected through the different instruments. It includes statistical analysis of survey responses, qualitative analysis of interviews, and interpretations drawn from the MOH data. The discussion section connects the results to existing literature and theories, highlighting the contributions and implications of the study. Practical recommendations are offered for policymakers, hospital administrators, and digital health strategists. The chapter concludes with final reflections and suggestions for future research on AI integration and health data management in the public sector.

References

- Alasmari, S., AlGhamdi, R., Tejani, G. G., Sharma, S. K., & Mousavirad, S. J. (2025). Federated learning-based multimodal approach for early detection and personalized care in cardiac disease. *Frontiers in Physiology*.
- Al-Kahtani, N., Alrawiai, S., Al-Zahrani, B. M., et al. (2022). Digital health transformation in Saudi Arabia: A cross-sectional analysis using HIMSS digital health indicators. *SAGE Open Medicine*.
- Alkraihi, A., Jackson, T., & Murray, I. (2016). Barriers to the widespread adoption of health data standards: An exploratory qualitative study in tertiary hospitals in Saudi Arabia. *Journal of Medical Systems*, 40(9), 235.
- Alshamrani, H. A., Alshamrani, K., Rashid, M., & Alshamrani, S. S. (2024). A secure federated learning approach for detecting COVID-19 from medical CT images. *Thermal Science*.
- Bawden, D., & Robinson, L. (2009). The dark side of information: Overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2), 180–191.
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2021). A guide to deep learning in healthcare. *Nature Medicine*, 27(1), 24–29.
- Gunning, D. (2017). Explainable Artificial Intelligence (XAI). Defense Advanced Research Projects Agency (DARPA), U.S. Department of Defense.
- Healthpoint. (2024). Is the Middle East leading the global healthcare IT revolution? *Healthpoint Intelligence Report*.

- JMIR. (2024). Digital transformation in Saudi hospitals: Opportunities and obstacles. *Journal of Medical Internet Research*, 26(5), e53369.
- Juneja, R. (2025). AI bias and the hidden dangers of fragmented health data. *Forbes AI Council*.
- Kankanhalli, A., Hahn, J., Tan, S. S. L., & Gao, G. (2016). Big data and analytics in healthcare: Introduction to the special section. *Information Systems Frontiers*, 18, 233–235.
- Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148–152.
- Liu, Q., Yang, H., & Zhang, M. (2018). FADL: Federated-Autonomous Deep Learning for distributed EHR data analysis. *Scientific Reports*.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21.
- Mohsen, H., Hassanien, A. E., & El-Dahshan, E. A. (2022). AI integration of medical imaging and EHR for clinical decision support. *BMC Medical Imaging*.
- Nguyen, D. C., Ding, M., Pathirana, P. N., & Seneviratne, A. (2021). Federated learning in healthcare: A survey. *ACM Computing Surveys*.
- PMC. (2022). Interoperability readiness in Saudi healthcare institutions. *Perspectives in Health Information Management*, 19(Winter), 1d.
- Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). Free Press.
- Saudi Health Council (SHC). (2023). Achievements Report: Nafis Program and Data Sharing Agreements. *Saudi Health Council Publications*.

SS&C Blue Prism. (2025). HealthTech AI Survey: The state of AI adoption in global healthcare systems. *HealthTech Digital*.

Statista. (2025). Digital healthcare infrastructure in Saudi Arabia – Statistics and projections. *Statista Research Department*.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103–110.

Topol, E. (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books.

von Bertalanffy, L. (1968). *General System Theory: Foundations, Development, Applications*. George Braziller.

Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3–13.

Wunker, S. (2023). Why AI fails without integration: Breaking data silos in healthcare. *Forbes Tech Council*.

Yang, J., Zhou, L., & Chen, H. (2022). Data harmonization in health systems: Techniques and challenges. *IEEE Transactions on Medical Informatics*.