Sensors and Materials, Vol. 37, No. 3 (2025) 1099–1138
MYU Tokyo

S & M 3976

Integrating Sensor Data with Large Language Models for Enhanced Elderly Care: A Methodological Framework

Ziaullah Momand,[1*] Pornchai Mongkolnam,[1,2] Jonathan H. Chan,[1,2] and Nipon Charoenkitkarn[1,2]

[1]School of Information Technology, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand
[2]Innovative Cognitive Computing Research Centre (IC2), King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

(Received October 17, 2024; accepted March 10, 2025)

Keywords: elderly care, sensor technologies, sensor data, large language models, older adults

The global aged population is expected to exceed 2.1 billion, representing 21.65% of the total population by 2050. This demographic shift underscores an urgent need for efficient elderly care, particularly in home settings. AI advancements have made sensor technology, including wearable biosensors, environmental monitors, and biochemical sensors, essential for elderly care by enabling the collection of physiological and activity data. Current systems overwhelm caregivers with complex data analysis and personalized recommendations. Large language models (LLMs) address this by offering insights through natural language interfaces, using extensive medical data. While some studies have integrated sensor data with LLMs for health monitoring applications, a comprehensive framework for seamlessly combining diverse sensor data with LLMs in elderly care is still missing. In this study, we propose a novel methodological framework that addresses the challenges of integrating heterogeneous sensor data with LLMs to provide real-time healthcare insights for caregivers of the elderly using sensor technologies. Our framework employs few-shot learning on Generative Pre-trained Transformer (GPT-4) and GPT-3.5 to process structured sensor data from wearable and environmental devices. The LLM-powered application then generates insightful responses based on the user's input, providing actionable and personalized recommendations. The GPT-4 model outperformed GPT-3.5 in Structured Query Language (SQL) query generation for sensor data retrieval and processing, achieving a semantic similarity score of 0.95, precision of 88.5%, recall of 98.92%, and an F1-score of 93.40%. In this study, we explore how integrating sensor data with LLMs enhances usability and reduces complexity in health monitoring systems. Our framework sets a new benchmark for advancing elderly care through innovative LLM-powered applications and sensor technology.

[*]Corresponding author: e-mail: mommand.csf@gmail.com
https://doi.org/10.18494/SAM5387

ISSN 0914-4935
https://myukk.org/

1. Introduction

The world's population is aging, and the growing social and economic consequences of a globally aging population have propelled research associated with aging into the limelight. In fact, significant strides in medical science over the recent decades have greatly expanded global life expectancy.(1) As a result, it is anticipated that by 2050, the number of individuals aged 60 and above will surpass 2.1 billion (approximately 21.65% of the total population).(2) This demographic transition signals emerging social, economic, and healthcare challenges, necessitating a greater emphasis on promoting healthy aging as a strategy to mitigate the effects of an aging population.(1,2)

In this regard, the emergence of Industry 4.0 has led to considerable progress in the advancement of health monitoring, particularly through the incorporation of sensor technology for elderly care.(3,4) These technologies are applied in various areas, including healthcare, personal health management, and physical activities, to provide convenience and real-time services.(5) This technology has revolutionized health monitoring, extending from traditional hospital settings to long-term residential care. Examples include wearable blood glucose monitors,(6) sport performance trackers,(7,8) and respiratory function measuring systems.(9)

Sensor technology enables the efficient, low-cost capture of large data volumes. However, processing and interpreting this data to derive valuable insights is a major challenge.(10) For example, wearable healthcare sensors generate vast amounts of data on heart rate, blood pressure, movements, and more, monitoring individual health and activity levels. Traditional sensors operate autonomously,(11) lack collaborative or integrated functionality, and do not have the ability to receive feedback or control inputs from the monitoring system.(12) Consequently, while they collect data, they cannot directly affect system operations. These constraints hinder the development of standard sensors into intelligent, interactive technologies integrated with AI. However, advancements in cloud computing, big data, deep learning, and generative AI have enhanced sensor applications. They enable the processing of large data volumes and high-precision predictions through multilayer neural networks.(13) They provide powerful tools and algorithms for data processing and analysis, which provide solutions for development limitations faced by sensor technologies in elderly health monitoring.(14–18)

The emergence of generative AI, particularly large language models (LLMs), is observed to provide strong evidence supporting the personalization of health services.(19) LLMs are pretrained on vast text datasets, then fine-tuned to make novel, responsible predictions based on the foundation of their knowledge.(20,21) These models can analyze new healthcare data and offer recommendations through detailed predictive findings presented in textual format.

Apart from healthcare, LLMs can be used for other specific contexts such as finance,(22,23) business,(24,25) education,(26–28) and manufacturing, particularly in enhancing knowledge management, operational efficiency, decision-making processes,(29) and design modification suggestions.(30) LLMs need specific context-based data to provide services for particular purposes. In the healthcare domain, data can be collected through various means, including clinical notes, electronic health records, medical documentations and reports, patient portal correspondences, drug and medical device data, and sensor-generated data.

Recent advances in healthcare sensor technology enable real-time diagnostics and various functions, including the continuous monitoring of physiological parameters. AI evolution has made sensor technologies crucial in elderly care, generating signal-based time series data. However, current health monitoring systems[31–43] often overwhelm caregivers for lack of personalized insights owing to the complexity of data captured by sensor technologies. LLM services can bridge this gap by providing personalized, actionable recommendations in natural language, making sensor data interpretation more accessible and effective for elderly care. Various studies[44–49] have integrated LLMs with specific sensor data to provide health insights and personalized recommendations for patients and healthcare professionals. However, no methodological approach that seamlessly combines LLM capabilities with heterogeneous sensor data in elderly care has been proposed. Such integration would provide actionable insights and personalized recommendations for caregivers while expanding sensor technology applications by conveying health insights in natural language.

To address this gap, in this study, we propose a comprehensive methodological framework for integrating diverse sensor data types into LLMs. This unified approach aims to enhance elderly care by leveraging sensor technologies to provide caregivers with actionable insights and personalized recommendations in real time. By demonstrating how sensor data such as physiological parameters (e.g., heart rate, blood oxygen saturation, steps, heart rate variability) can be effectively processed and interpreted using LLM, in this framework, we establish a scalable solution for personalized healthcare. Furthermore, we set the foundation for future studies on LLM integration with sensor technologies in diverse health monitoring applications for older adults, taking a significant step forward in the evolution of sensor technologies in healthcare. This study is carried out with the following specific objectives.

a. To identify sensor types, their data, and processing techniques in elderly care and to define data sources and processing requirements for LLM integration
b. To propose a framework for integrating diverse sensor data with LLMs for personalized elderly care and to evaluate LLMs' potentials in sensor data interpretation

The rest of the paper is organized as follows. In Sect. 2, we review existing work, and the materials and methods are described in Sect. 3. In Sect. 4, we present the results, whereas in Sect. 5, we introduce the proposed framework. Experiment details are described in Sect. 6, the discussion in Sect. 7, and challenges and opportunities in Sect. 8. Finally, in Sect. 9, we conclude this study and outline future work.

2. Related Work

Several LLMs such as Med-PaLM2,[19] HuatuoGPT,[50] DISC-MedLLM,[51] ChatDoctor,[52] and Baize-HealthCare[53] have been developed for medical question answering, using large medical datasets for diagnostic dialogues. With supervised fine-tuning, these models outperform Generative Pre-trained Transformer (GPT-4) and Llama2 on benchmark datasets such as USMLE questions, MedQuAD, and online medical consultation datasets.[54]

Wang *et al.*[55] suggested models using multiple LLMs, each specialized in a medical area, for automated diagnosis. Li *et al.*[56] enhanced LLM diagnostic abilities by incorporating clinical decision tree, whereas Li *et al.*[57] fine-tuned LLaVA using biomedical figure captions for tasks such as interpreting computed tomography scans. Tu *et al.*[58] refined LLMs through simulated doctor–patient dialogues generated in a self-play environment, combined with public medical datasets and real hospital interactions.

Several groups have explored enhancing the LLM capabilities in medical knowledge through retrieval augmented generation (RAG), few-shot learning, and knowledge base. For instance, an LLM based on fast healthcare interoperability resources (FHIRs) was developed using GPT-4 and integrated into a mobile application that provides patients with health information on the basis of their health records.[59] Models such as LLM augmented with medical textbooks[60] and Health-LLM[61] have created databases of medical textbooks for LLMs access. In another approach, LLM was combined with human expertise to annotate unstructured clinical data, creating ground truth labels.[62] A framework has also been developed where LLMs use health-related features to predict diseases, with LLM scoring the potential outcome.[63] LLMs can match patients with suitable clinical trials to provide optimal care. In one study, health records were used and four data augmentation techniques were applied to enhance the LLM performance, thereby generating additional data while preserving the original trial information meanings.[64]

Biomedical LLMs such as Meditron-70B,[65] BioMistral-7B, and BioMistral-7B-DARE[66] are widely used for medical purposes, with their performance depending on the quality and type of healthcare datasets. Researchers developed evaluation benchmarks using MeDiSumQA,[67] MedNLI,[68] and MedQsum[69] with few-shot learning.[70] In another study, Li *et al.* integrated the quantized low-rank adaptation (QLoRA) algorithm with the ChatGLM2-6B and Llama2-6B models, fine-tuning them with a medical Structured Query Language (SQL)-based dataset and using prompt engineering on ChatGPT-3 to elicit patient health data through SQL queries.[71] Advances in the conversational abilities of LLMs have driven healthcare innovation. Instruction prompt tuning improved the pathways language models (PaLMs), leading to the creation of Med-PaLMs for answering medical queries.[72] These developments promise further innovation in healthcare conversational models.

2.1 Recent advances in sensor data integration with LLMs

Recent advances in modeling wearable sensor data have enabled novel applications of LLMs in health monitoring, leading to significant improvements in personalized recommendations. In a number of studies,[44,73,74] models such as Llama, Phi3, and GPT-4, all proposing a hybrid model comprising Llama combined with LIMU-BERT, which processed Inertial Measurement Unit (IMU) sensor data to integrate with the Llama model, were utilized.

The integration of accelerometer, gyroscope, and biomechanical sensor data such as joint movement, into GPT-4 and Llama has advanced their application in multiturn medical consultations and physiotherapy.[44] GPT-4 also incorporates glucose sensor data for real-time diabetes monitoring.[75] LLMs have made significant progress in analyzing sleep quality,[45] physical activities,[46–49] Parkinson's disease monitoring, and fitness,[76] offering personalized recommendations and aiding healthcare professionals. Hybrid models, such as combinations of convolutional neural networks and transformers, enhance the accuracy of LLMs in analyzing sensor data.[77] Task-specific models such as PhysioLLM[48] and HealthLLM[61] further improved prediction accuracy by leveraging contextual data obtained from wearable sensors. Yu *et al.*[78] evaluated GPT-3.5 and Llama2-70B in diagnosing sleep apnea and arrhythmia from electrocardiogram (ECG) data by the few-shot retrieval method, converting numerical ECG data into text for analysis. Table 1 summerizes the comparative analysis of existing work with our approach.

3. Materials and Methods

In this study, we aim to enhance the integration of sensor data with LLMs for personalized elderly care solutions for caregivers. We also propose a methodological framework that lays the foundation for LLM-powered healthcare research in the elderly care domain and advance the

Table 1
Comparative analysis of existing LLMs for sensor data analysis.

Approach	LLM	Data type	Evaluation method	Sensor data integration	Purpose
LLaSA[73]	Llama2	IMU data	F1-score, recall, precision	LMUT-BERT encoder	Health monitoring
Physiotherapy LLM[74]	Phi3, Llama3, Llama3:70B	Biomechanical sensor data	Accuracy, response time	Text-to-SQL method	Physical activity monitoring
DrHouse[44]	Llama3:70B, GPT-3.5 Turbo, GPT-4, and LLaMA-3B-Instruct	Data of physiological indicators	Accuracy	Knowledge base and fusion approach, which integrates sensor data as text format	Medical consultation
Glucose monitoring with LLM[75]	GPT-4	Glucose sensor	Subjective evaluation	Linguistic summarization	Real-time monitoring of diabetics
LLM for sleep quality forecasting[45]	GPT-3.5	Motion, physical activity data	Pearson correlation coefficient	Random decision tree decision path to create text data for LLM	Sleep quality forecasting
LLM clinical Insight[46]	GPT-4, GPT-3.5, PaLM 2	Motion data	Accuracy	Tabular and markdown formatting methods	Support mental health professionals
Automated health coaching[47]	GPT-4, GPT-3.5, Claude-2	Data of physiological indicators	Human in loop	Integrated sensor data into prompts	Personalized health insights
PhysioLLM[48]	GPT-4	Data of physiological indicators	Manual validation	Statistical summarization	Improving sleep quality
AutoHealth[76]	GPT-4, GPT-3.5, Claude-2	Data of motion sensors	Accuracy	Labeled data by ML technique.	Parkinson's disease real-time monitoring
HARGPT[49]	GPT-4, Llama2-70B	IMU data	F1-score	Input raw IMU data is tokenized	Real-time human activity recognition
Our proposed framework	GTP-4, GPT-3.5 Turbo	Motion, physiological, body thermal, biochemical, and environmental data	F1-score, precision, recall, semantic similarity score	Text-to-SQL and natural language string methods	Comprehensive elderly care (real-time monitoring, personalized support, long-term health tracking).

applications of sensor technologies for health monitoring through a natural language interface. Given the framework's focus on sensor-based healthcare data, it is crucial to explore and identify the current sensor technologies and the types of data they capture for elderly care. To achieve this, we conducted a systematic literature review of existing sensor technologies for elderly care to identify the types of sensor, the specific data they generate, and their associated processing requirements. These findings directly guide the design of the data identification and processing layers of our proposed framework, as shown in Sect. 5. Figure 1 illustrates the methodological flow of this study, which begins with a systematic exploration of sensor technologies and culminates in the development of a comprehensive framework. The systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta Analysis (PRISMA) guidelines(79) approach represented in Fig. 2.

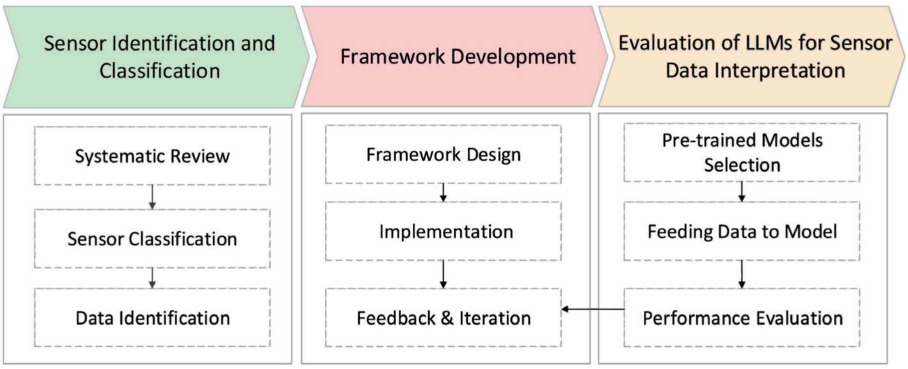


Fig. 1. (Color online) Diagram illustrating the methodological flow of the study.

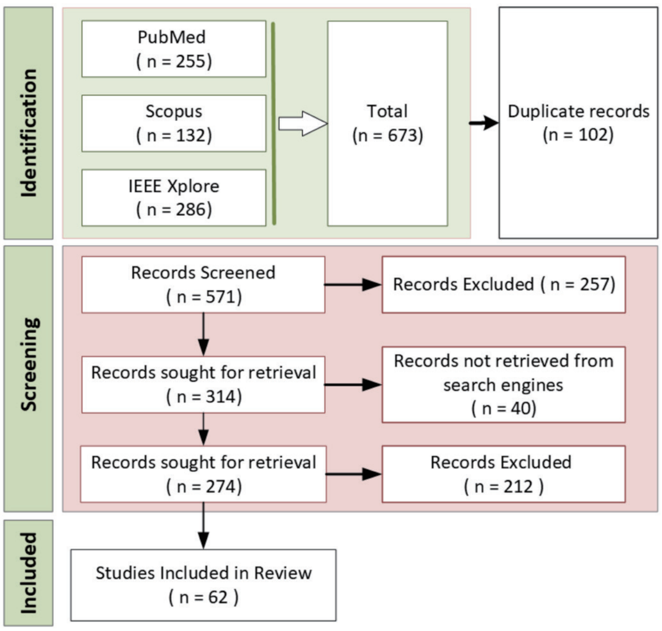


Fig. 2. (Color online) Article screening process as per the PRISMA standard.

The articles considered for this review were published between 2018 and March 2024 and were retrieved from PubMed, Scopus, and IEEE Xplore databases. This timeframe ensures the coverage of latest innovations in sensor technology applications, including wearable devices and integration of IoT and AI systems, which address contemporary elderly care challenges. These advancements provide actionable insights and a strong foundation for designing state-of-the-art elderly care systems, such as improved health monitoring and personalized care solutions.

While earlier research, including the 37 articles identified in the initial screening, which were published before 2018, contributed to foundational knowledge, the selected timeframe highlights significant technological advancements and methodological shifts. From 2018 onward, sensor technologies saw transformative progress, including the rise of IoT devices, widespread adoption of smart wearables, and the integration of AI and machine learning in healthcare. Recent studies emphasize multimodal data integration, combining multiple sensor data with other inputs to enhance decision-making and care personalization. This focus ensures that the findings remain relevant and aligned with contemporary developments, particularly in frameworks involving LLMs. To maintain a comprehensive and targeted search, predefined keywords and their combinations were carefully crafted and applied across these databases.

3.1 Eligibility criteria

- The studies considered must be articles published between January 2018 and March 2024, focusing on recent advancements in sensor applications for elderly care.
- Articles must focus on the use of sensors in the context of elderly care.
- Studies should clearly specify the types of sensor used, such as motion sensors, heart rate monitors, and temperature sensors.
- Articles should describe how the sensor data is applied in elderly care.

3.2 Study selection and screening process

Following the initial search, we identified a total of 673 results. Figure 2 shows the journey of identifying, screening, and including articles, guided by the PRISMA(79) flowchart. After identifying the articles, we utilized the Zotero citation tool(80) to manage the deduplication process. A total of 102 duplicate articles were automatically flagged and removed by Zotero, leaving us with 571 unique records to be screened. In the initial screening, 257 records were excluded owing to our predefined exclusion criteria. We identified 314 promising records for further retrieval and analysis. Despite our efforts, 40 of these records could not be retrieved from search engines and remained out of reach. We then conducted a detailed assessment of 274 records, ultimately setting aside 212 articles for the following reasons: 99 did not focus on the nuances of elderly care, 20 discussed IoT without mentioning specific sensors, 35 provided extensive technical details on sensors, and 46 explored nonsensor-based interventions. A quality assessment using an eight-criteria checklist was conducted on the remaining articles to evaluate their relevance. Twelve articles did not meet the quality standards, leaving 62 studies as the most relevant and insightful, forming the core of our review.

4. Results of the Review

Our analysis of the included articles identified various sensor categories used in elderly care, such as motion, environmental, physiological, vision, body thermal, and biochemical sensors. These sensors support vital functions such as monitoring vital signs, fall detection, cognitive impairment, daily activities, physical activity, Parkinson's disease, and pneumonia detection. Sensors were classified on the basis of the data they collect and their healthcare applications, as shown in Table 2. The review highlights the most feasible sensors and data types that can be integrated with LLMs to enhance elderly care. By focusing on motion, physiological, environmental, and vision sensors, which provide critical data on physical activity, fall prevention, and vital signs, these sensors offer the best potential for improving safety and personalized care through advanced predictive models. Integrating this sensor data with LLMs can further optimize elderly care strategies, addressing gaps in current practices and enabling more comprehensive health management solutions.

The distribution of sensor types in elderly care, as shown in Fig. 3, reveals that motion and physiological sensors are the most prevalent, reflecting a strong focus on monitoring physical activity, fall prevention, and vital signs, which are critical for ensuring the safety and health of the elderly. Vision-based sensors are moderately used, indicating a growing interest in noninvasive monitoring for elderly care. Environmental and body thermal sensors are emerging technologies, suggesting areas for future research, while the limited use of biochemical sensors highlights potential but underexplored avenues in elderly care. This distribution underscores both the current priorities in elderly care and the opportunities for innovation in sensor types for LLM integration to enhance personalized care and expand the scope of elderly care practices.

Table 2
Sensor types with associated context of use in elderly care.

	Physiological parameters	Fall dectection	Cognitive impairment	Activity of daily living	Gait analysis	Physical activity recognition	Parkinson's disease monitoring	Kinematic analysis	In-bed movement recognition	Pneumonia early detection	Diagnosis of colorectal cancer	Ambient assisted living	Loneliness
Motion sensors	✓	✓	✓	✓	✓	✓	✓	✓	✓				
Environmental sensors	✓	✓	✓	✓	✓							✓	✓
Physiological sensors	✓	✓	✓	✓		✓				✓	✓		
Vision-based sensors	✓	✓	✓	✓	✓		✓						
Body thermal sensors	✓	✓	✓	✓									
Biochemical sensors	✓												

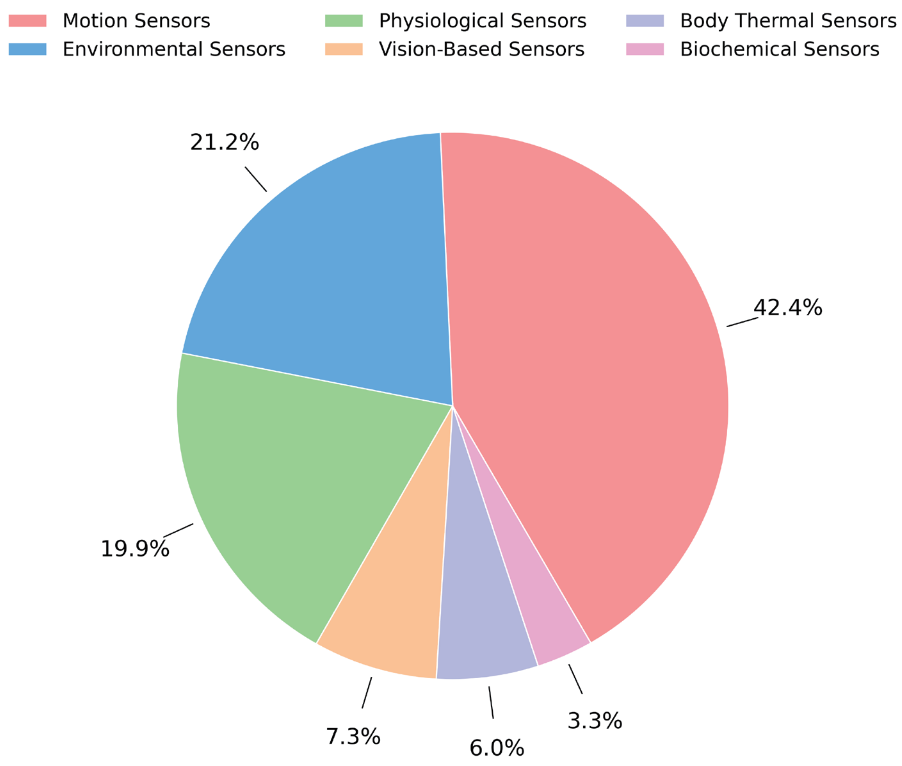


Fig. 3. (Color online) Distribution of sensor types identified across 62 included articles.

5. Proposed Framework

In this section, we propose a novel methodological framework that integrates the identified sensor data and their associated processing techniques into LLMs for elderly care. LLMs are most efficient with text-based data formats, ranging from purely natural language to various structured and semistructured textual representation. Recent advancements in LLMs have demonstrated their remarkable capabilities across a range of tasks, from medical tasks(19,50–53) and financial data analysis(22,23) to identifying patterns in manufacturing.(29,30) Recent research explored the capabilities of LLMs(44,45,73–78) for processing sensor data to gain predictive insight for various healthcare tasks. Sensor data plays a crucial role in patients' health outcomes, as it is collected in real time.

Integrating sensor data into LLMs is challenging owing to the complexity of processing raw numerical sequences essential for healthcare. Our novel framework incorporates these sequences into LLMs, enhancing predictive insights for elderly healthcare. As shown in Fig. 4, the framework has five layers: identifying sensor data sources; collecting, integrating, and processing data; preparing data for LLM processing; analyzing the data with LLMs; and interpreting the output for caregivers, as will be elaborated in the following subsections.

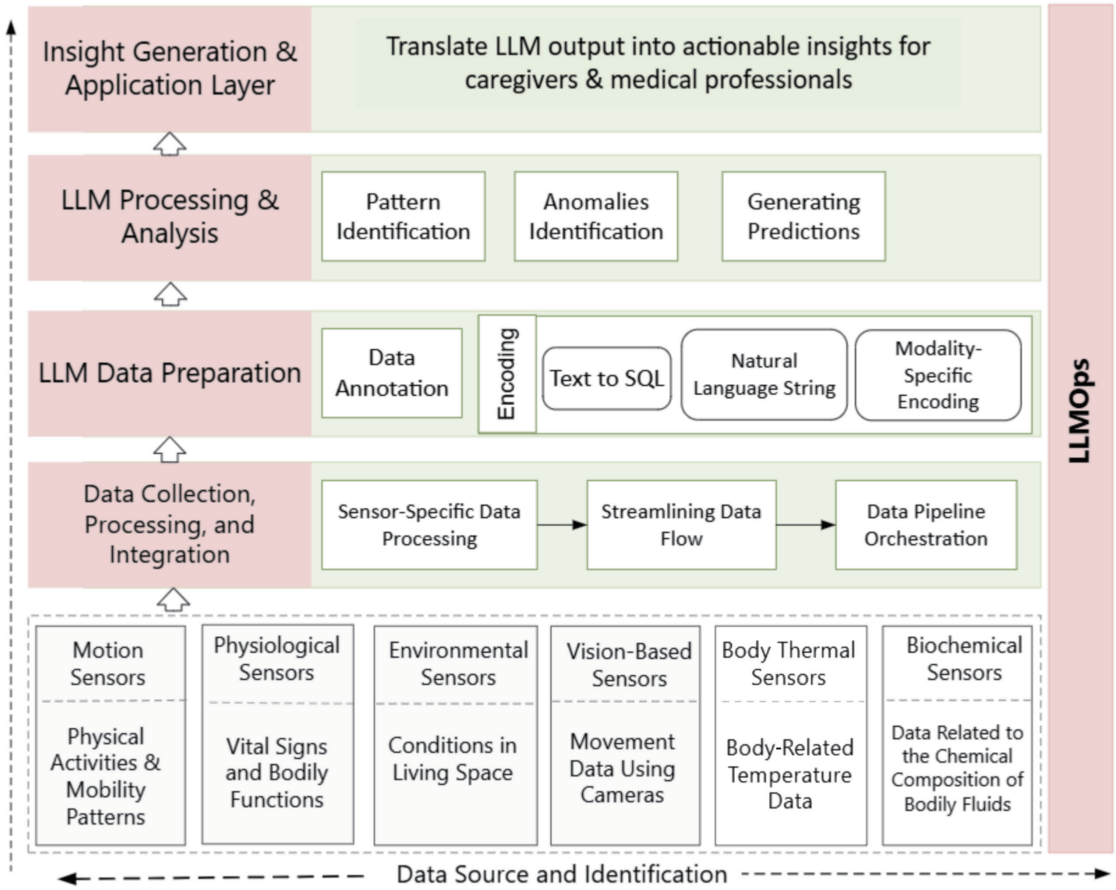


Fig. 4. (Color online) Proposed framework for integrating sensor data in LLM for enhanced elderly care.

5.1 Data source and identification

Identifying the right data is essential for LLMs to generate healthcare insights. Data can come from the elderly's environment, their bodies, and other sources, depending on the type of care needed. For example, healthcare professionals may require physiological data for health assessments, while caregivers may need information on physical activity, social interaction, and cognitive function. We have identified key sensor data sources that enrich elderly care.

5.1.1 Motion sensors

Motion sensors detect physical movements and are widely used in elderly care for movement tracking patterns, which are essential for kinematic analysis,(81,82) fall detection,(34–38,41,43) and monitoring physical activity,(83) and daily living activities.(32,84–86) Common sensors used in elderly care include accelerometers,(34–40,83) gyroscopes,(37,41,87) PAMsysTM,(88) posture sensors,(89) Pablo,(90) and magnetometers.(41)

These are embedded into wearable and nonwearable devices such smartwatches and smartphones for easy use. Motion sensors are key data sources for LLMs such as Llama2-70B,(73,74) GPT-4,(47,76) GPT-3.5,(45,47,76) and Claude-2,(47,76) which enable personalized recommendations and health monitoring. These models use sensor data to monitor physical activities, provide physiotherapy consultations,(74) predict sleep quality,(45) and offer real-time monitoring of Parkinson's disease,(76) which are vital for the elderly population.

5.1.2 Physiological sensors

Physiological sensors play a key role in monitoring elderly well-being by detecting early signs of health deterioration and preventing crises. They measure vital signs such as heart rate, blood pressure, body temperature, and sleep patterns. Various sensors, including those of heart rate,(88–97) blood pressure,(94,98) body temperature,(94,99,100) and sleep pattern,(101) collect essential physiological data that LLMs use to analyze trends in vital signs and assess potential indicators of mental health and other physiological conditions.(102) By integrating contextual information, such as user demographics and health history, with sensor data on heart rate and sleep patterns, LLMs can generate clinically relevant insights and explain data trends related to mental health.(46,50)

5.1.3 Environmental sensors

Environmental sensors continuously monitor elderly living spaces, providing valuable data for LLMs. Key types include temperature sensors,(103) humidity sensors,(104) light sensors,(105) pressure sensors,(106) gas sensors,(104) air quality sensors,(107) magnetic contacts(108), ultra-wideband radar,(109) bed sensors,(84) and door sensors.(110) Humidity sensors impact respiratory health and comfort,(111) whereas light sensors enhance independence.(105) Pressure sensors monitor body position and detect pressure changes.(112) Passive infrared sensors, placed throughout the home, detect movements, enabling LLMs to monitor dementia, safety, and cognitive impairment.(110–116) Environmental sensors enhance LLMs' enabling the analysis of long-term data, air temperature, air quality, and humidity, aiding dementia diagnosis and occupancy tracking for models such as GPT-3.5 and Llama2-70B.(117) These sensors provide a crucial knowledge base to LLMs for ambient sensor-based human activity recognition, the accuracy of which improves when contextual information such as sensor location, function, time, and environmental conditions are integrated.(118)

5.1.4 Vision-based sensors

Vision-based sensors play a significant role in elderly care, enabling real-time monitoring,(119,120) activity recognition,(121,122) action recognition, and fall prevention.(39) These sensors incorporate computer vision technologies and visual sensors, such as RGB cameras, depth cameras,(123,124) and Kinect version 2,(125,126) to monitor the activities and movements of elderly adults in living environments. These sensors support independent living by detecting normal and abnormal activity patterns. Nonintrusive data collection allows LLMs to perform complex tasks such as reminding seniors to wash their hands if unsanitary eating is detected.[127] These sensors also assist LLMs in motion reasoning, providing insights into acceleration and deceleration through geometrical and temporal analysis[128] and emotion recognition using vision sensors.[77]

5.1.5 Body thermal sensors

Body thermal sensors collect data regarding changes in sweat gland activity. Data can be collected by placing sensors on the skin, such as on the hands, feet, or fingers.[129] For elderly care, these sensors can be used to monitor changes in skin conductance,[130] detect mental disorder,[131,132] and assess stress levels,[98] which may be indicative of hydration levels, various health conditions, and emotional and physical states.[133] These sensors equip LLMs with detailed data to track long-term temperature trends, offering insights into elderly health patterns and predicting potential health issues before they become severe.[48,50]

5.1.6 Biochemical sensors

Biochemical sensors for elderly care are devices that detect and measure different biochemical markers. These sensors can be either invasive or noninvasive. Noninvasive sensors are attached to the skin or clothing to detect biomolecules such as glucose, lactate, alcohol, and ions.[134] Invasive biochemical sensors are implanted in the body to monitor biomolecules related to diabetes in real time.[135] Glucometers are devices equipped with glucose sensors and monitor diabetes levels in the elderly.[15,136] Data from wearable continuous glucose monitors can be presented in both numerical and graphical formats for easy interpretation.[137] This data serves as a knowledge base for LLMs such as GPT-4, enabling personalized insights into glucose levels,[75,138] linguistic summarization of glucose data,[139,140] and personalized recommendations.

5.2 Data collection, processing, and integration

After data sources and types are identified, deploying the appropriate sensors is essential for data collection in areas such as living environments and wearables. Wearable devices such as smartwatches and fitness trackers[32,99,100,141] can be used for data collection owing to their affordability, ease of use, and seamless integration into daily life. They transmit data via Bluetooth, Wi-Fi, or cellular networks, allowing real-time uploads to cloud or local storage. Ambient sensors in living environments, including vision-based[123,124] and environmental monitoring devices,[104–109] can collect real-time data while considering privacy concerns. To maintain consistency across diverse sensor data, establishing data standards is crucial. These standards should cover naming conventions, data types, formats, units, ranges, and validation criteria for all sensor data.[142]

5.2.1 Sensor-specific data processing

The data collected from the identified sensors requires diverse processing steps before being fed into the LLM for analysis. Each sensor type demands specific preprocessing techniques tailored to the nature of data it captures. Motion sensors capture data related to physical activities and mobility patterns such as walking, running, sitting, and fall events.[32,84–86] The raw data collected from these sensors are often noisy, requiring preprocessing steps such as noise filtering, signal smoothing, and normalization to ensure data quality. Feature extraction is necessary for identifying steps, posture changes, or fall events.[37,38–40,41,83,87] Temporal analysis helps in segmenting data into meaningful activity states (e.g., walking, running, and resting). This processed data helps LLMs provide accurate insights into the physical activity patterns of older adults. For instance, LLMs can identify irregular activity patterns or prolonged periods of inactivity that may signal a decline in mobility.[74,76]

Physiological sensors measure vital signs, including heart rate,[88–97] respiratory rate, oxygen saturation (SpO2), and blood pressure.[94,98] These sensors generate time-series data, which must undergo artifact removal and validation to ensure accurate use in LLM applications. To enable LLMs to better interpret physiological data, it is crucial to preprocess the data to ensure consistency checking, physiological range validation, and noise removal.[143] This includes validating timestamps to ensure that each data entry corresponds to a valid time interval. Methods such as linear interpolation (commonly used for heart rate data) and exponential interpolation (particularly effective for respiratory rate) can be applied to fill the gaps caused by missing time intervals. Other techniques, such as association rule mining, clustering, and single value decomposition, can also be used. Each physiological parameter must be checked against established clinical ranges to detect anomalies, sensor errors, or outliers. These are measurements that deviate significantly from surrounding data points and must be addressed.[144–149] For example, SpO2 must be between 90% and 100%, and values below 90% should be flagged for clinical concern or sensor error.

Environmental sensor data must be standardized to consistent time intervals to enable LLMs to predict timely trends for parameters such as temperature and humidity. Resampling methods including mean, median, and interpolation are commonly used to handle time interval consistencies.[104,107] Beyond resampling, the data should also be checked for “stuck-at-zero” or “dead sensor” faults, where values remain zero for extended periods, indicating potential malfunctions. Removing noise is essential, as it refers to random data points that distort the true signal.[150,151] Common methods for noise removal include low-pass filters, which remove high-frequency noise from temperature or humidity signals; wavelet transform methods, which decompose signals into frequency components to eliminate noise; and subspace methods, which identify and filter out noise on the basis of signal patterns.[152,153] Once processed, environmental data is refined for accuracy and becomes actionable. It can then be fed into LLMs for advanced analysis and personalized recommendations. For example, LLMs can suggest adjustment of the heating or cooling system to maintain optimal indoor conditions tailored to the health and comfort of elderly individuals.

Using LLMs to analyze vision-based sensor data provides crucial health recommendations. Computer vision models, such as pose estimation and object detection algorithms, process this data to extract meaningful features.[121,122] Since data from these sensors is typically captured in image or video formats, standardizing image dimensions is essential for efficient processing. Additional steps such as normalization and data augmentation are also critical to improve the efficiency of LLMs in analyzing this data. Vision-based data is inherently high-dimensional and complex. By focusing on relevant features and using dimensionality reduction techniques such as principal component analysis, data becomes easier to process,[154] enabling real-time integration with LLMs for actionable healthcare recommendations. Key steps such as noise reduction[155] and image correction[156,157] are crucial for refining the raw data.

Body thermal sensors provide clinical data for identifying conditions such as fever, hypothermia, and other temperature-related conditions. To ensure accuracy, sensor reading should be calibrated to account for external environmental influences, and noise removal techniques should be applied. Time-series analysis of body thermal sensor data enables LLMs to detect gradual changes that may indicate emerging health issues.[48,50] For instance, if an elevated body temperature is detected alongside poor air quality, the LLM might suggest hydration and rest, as well as adjustments to the room environment to improve comfort and health.

Biochemical sensors provide highly specific health data, but it must be processed to remove noise and ensure accuracy and consistency. Feature extraction methods, such as time-window-based feature extraction[158] and multitask self-supervised learning,[159] are used to identify significant biomarkers, such as glucose spikes, which may indicate issues with diabetes management. When integrated with LLMs, biochemical sensor data enables personalized and precise health recommendations. For example, LLMs can analyze glucose levels to suggest dietary adjustments and alert users about potential insulin requirements.[75,138]

Sensor data fusion, which integrates data from physiological, motion, environmental, biochemical, and body thermal sensors, is crucial for LLM-based applications to enhance contextual understanding, provide personalized insights, and generate tailored recommendations. Various techniques can be employed for data fusion, including probabilistic, statistical, knowledge-based, and inference methods.[152,160] Probabilistic methods encompass techniques such as Bayesian networks[161,162] and maximum likelihood estimation.[163]

To seamlessly integrate data from the six sensor types, a robust data pipeline orchestration is essential. This pipeline ensures the efficient collection, preprocessing, validation, and fusion of data streams from motion, physiological, environmental, vision-based, body thermal, and biochemical sensors. It manages data flow across diverse formats and time intervals, standardizes units, handles missing or noisy data, and performs feature extraction tailored to each sensor type. The orchestrated pipeline combines these multimodal data streams into a unified format ready for analysis by LLMs.

5.3 Data preparation for LLMs

Leveraging sensor data in LLM applications is challenging since LLMs are primarily trained on text. To enable LLMs to work with raw sensor data, signal transformation and data preprocessing are essential, allowing models to better understand and interact with the physical world.(164) Sensor data annotation is crucial for LLMs, especially in classification tasks such as labeling physical activities like standing, sitting, walking, and lying down. Annotation can be performed manually or automatically. Manual annotation is accurate but time-consuming and requires experts. Automatic methods are faster and include supervised,(165,166) semisupervised, and unsupervised(167) learning approaches. Supervised learning relies on prelabeled data, which is costly, whereas semisupervised methods have reduced labeling needs.(168) Annotated sensor data enhances LLMs' understanding by linking raw data to specific conditions.

Further encoding is necessary to make temporal sensor data readable for LLMs and thus enhance next-token prediction. Methods including (1) natural language string,(169) (2) modality-specific encoding, (3) statistical summary,(170) and (4) the Text-to-SQL method(171) can be used. The natural language string approach represents time series as a string of numerical digits, where forecasting is treated as next-token prediction, as shown in Fig. 5. Physiological and biochemical sensor data can be converted into text using linguistic summarization, by crafting specific templates for each parameter.(75) For example, "the patient exhibits an elevated heart rate of 120 bpm and a high blood pressure of 150/95 mmHg." Machine learning techniques such as random decision tree can also translate sensor data into text, particularly for physical activities.(45) In this method, the decision path can be converted into a textual format. These methods have been evaluated using models such as GPT-3.5,(45) GPT-4o,(75) and Llama-3 70B.(44)

The modality-specific encoding technique, as shown in Fig. 6, uses pretrained encoders for each data type to convert time-series numerical data into embeddings that share a latent space with language tokens. Although this approach provides detailed data representations, it

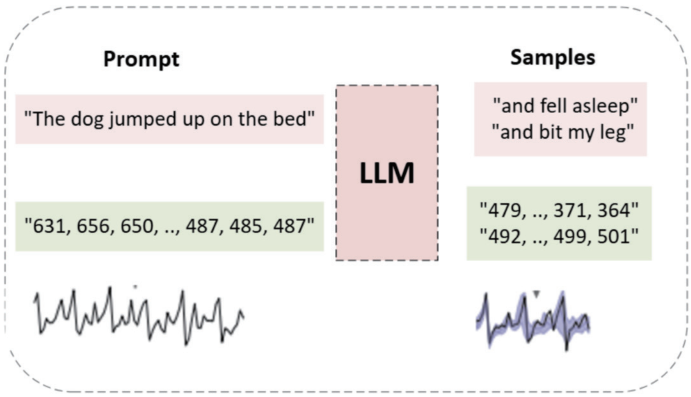


Fig. 5. (Color online) Natural language string encoding. example.

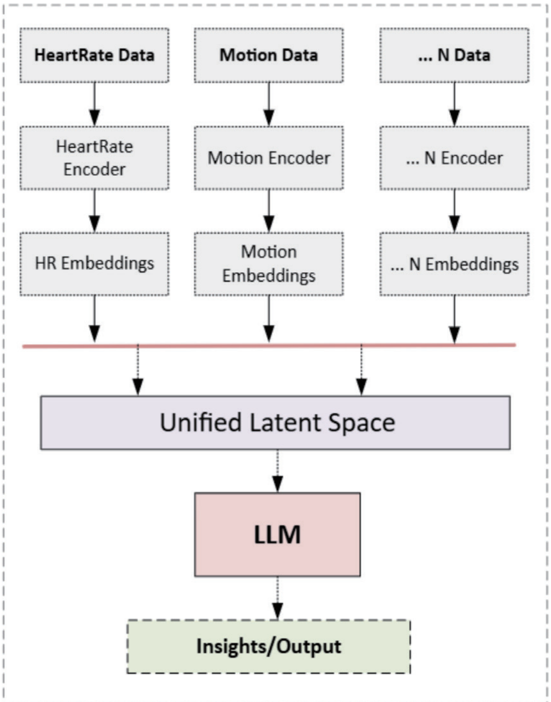


Fig. 6. (Color online) Modality-specific encoding.

introduced complexity and computational demands that may not always lead to proportional performance gains.(63) For instance, LIME-BERT, a modified BERT-based pretrained encoder, extracts features from sensor data and converts them into embeddings that share the same latent space as language tokens, enabling integration with text-based data. This method has been used with Llama-2 for IMU sensor data.(73)

The statistical summarization method presents data through key statistical measures such as mean, standard deviation, and median. For instance, physiological data can be integrated into LLMs by providing statistical summaries and trends. This includes metrics such as average weekly steps, average weekly heart rate, and hourly steps.(48) Although this approach significantly reduces the volume of data to be analyzed, it risks overlooking important temporal patterns present in the raw data.(170) In the Text-to-SQL method, sensor data is stored in a SQL database with structured tables and logical relationships. LLMs are guided to generate SQL queries to extract and analyze sensor data.(171) For example, the Text-to-SQL method has been used to process biochemical sensor data analyzed with Phi3 and Llama-3 70B LLMs.(74)

5.4 LLM processing and analysis

The LLM is the core analytical engine in the framework, and processes transformed data to identify patterns, make predictions, and generate insights. Using historical sensor data, it predicts risks such as falls or health deterioration. To ensure accuracy, LLMs should be trained with domain-specific sensor data by fine-tuning pretrained models on targeted datasets for specialized tasks.(172) Choosing the right pretrained model for the target task is essential. Various models are available, from healthcare-specific to general purpose. For instance, Med-Alpaca-70B(173) is fine-tuned for medical question-answering (Q&A) using extensive medical texts. ClinicalCamel-70B,(174) based on the Llama-2 70B architecture and optimized with QLoRA, is designed for clinical applications. Palmyra-Med-20B,(175) fine-tuned on custom medical datasets, excels in tasks such as PubMedQA and MedQA, whereas PMC-Llama-13B(176) integrates knowledge from 4.8 million biomedical papers. Additionally, models such as GPT-3.5, GPT-4, and Llama-3 can serve as base models.

Fine-tuning methods can be selected on the basis of the nature of the task and data. One approach is instruction fine-tuning, which is particularly effective for improving model performance on small tasks.(177) The fine-tuning dataset consists of prompts that provide instructions for the desired task, followed by the expected outcomes, as shown in Fig. 7. However, careful optimization is required to achieve the best results. This involves designing the instruction dataset thoughtfully and tuning hyperparameters such as learning rate, batch size, and the number of epochs. Additionally, combing instruction fine-tuning with prompting techniques such as chain-of-thought prompting can further enhance model performance.(46)

Full-parameter and parameter-efficient fine-tuning are the key methods for adapting models to specific tasks. Full fine-tuning updates all the model's parameters but is resource-intensive, requiring significant memory and computer power, often using techniques such as low-memory optimization(178) careful hyperparameter tuning, and regularization strategies to optimize performance. In contrast, parameter-efficient fine-tuning modifies only a subset of parameters, offering a more resource-efficient alternative.(179)

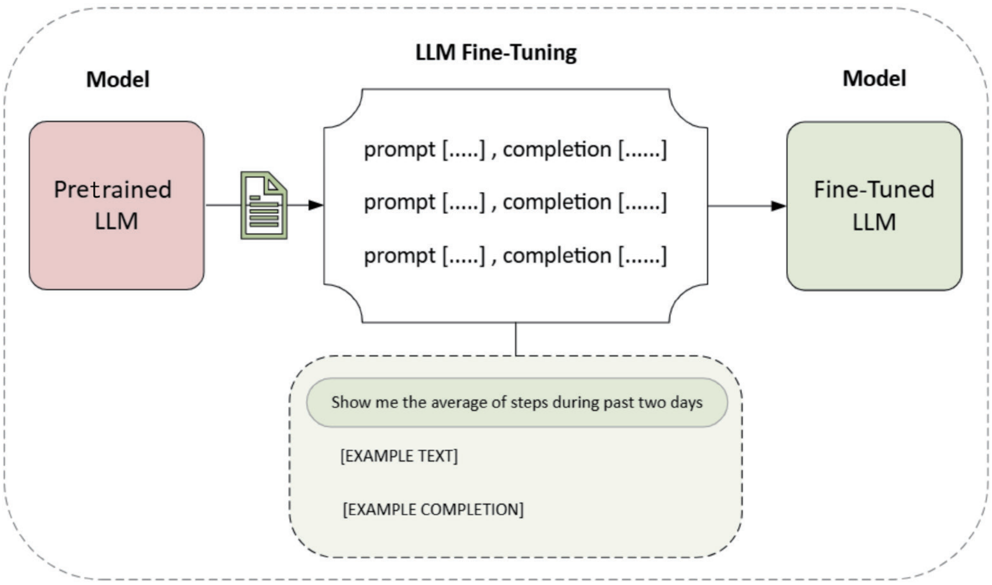


Fig. 7. (Color online) LLM instruction fine-tuning for sensor data.

RAG, an alternative to fine-tuning, is ideal for LLM-based healthcare applications such as chatbots, as it combines natural language generation with information retrieval, adapting to evolving facts.(180) Its main advantage is the continuous updating of training data, keeping the model accurate.(177) Additionally, methods such as zero-shot learning, one-shot learning, and few-shot learning can also guide a model's responses without modifying its parameters. These approaches do not involve training the network's weights; instead, they rely on crafting appropriate inputs to achieve the desired output. These methods are particularly suitable for sensor data applications when used with RAG(49) and are useful for simpler data or small datasets.(181) In few-shot learning, the model is given a few task examples without weight updates. One-shot learning provides one example with a task description, whereas zero-shot learning relies solely on natural language instructions without any examples.

5.5 Insight generation and application interface

At this stage of the framework, the goal is to ensure that the insights generated from sensor data analysis via LLM are effectively communicated to the end users who need them. The output of the LLMs should be translated into clear, understandable, actionable insights, recommendations, or alerts for caregivers. This can include natural language generation to create textual summaries of findings or visualization for easier interpretation. The proposed framework is designed with flexibility and adaptability in mind, allowing it to be incorporated into current healthcare infrastructures with minimal disruption. The framework can utilize IoT-enabled devices and sensors already in use within current electronic health record systems by leveraging application program interfaces (APIs) and interoperability standards. For caregivers, the framework can provide user-friendly interfaces, such as mobile apps or web-based dashboards, that offer actionable insights derived from sensor data and LLM analysis. These tools are designed to simplify complex data into easily understandable recommendations, helping caregivers make timely and appropriate interventions.

5.6 LLM operations (LLMOps)

The framework is concluded with the concept of LLMOps. LLMOps encompass the operational processes for handling, analyzing, and optimizing LLMs to generate insights from sensor data. It is essential for managing, deploying, monitoring, and optimizing LLMs, ensuring the framework's functionality and effectiveness. LLMOps ensures that data from various sensors is validated, accurately processed, and seamlessly integrated into the model, enhancing prediction accuracy and anomaly detection. LLMOps automates the retraining of the model when necessary, ensuring it remains effective over time. For instance, it can automatically trigger model updates if sensor data patterns change, maintaining the reliability and relevance of the insights provided to caregivers. This can be further improved by developing an automatic orchestration pipeline that involves continuous integration and continuous deployment. This framework can serve as an abstract baseline for LLMOps in handling diverse sensor data, providing the necessary tools and processes to manage the LLM and ensuring that it operates effectively and reliably in elderly care.

6. Case Study

To validate our framework, we developed a NodeJS-based healthcare chatbot powered by an LLM. This chatbot provides elderly caregivers with insights and recommendations based on sensor data. Leveraging LLM capabilities, it interprets elderly health data, allowing caregivers to inquire about health status through text or voice commands. The chatbot then generates responses from sensor-collected data, offering timely guidance. Figure 8 shows the flow of the LLM-based healthcare chatbot for elderly care.

6.1 Dataset

For the experiment, the PMData[182] dataset was used, which comprises lifelogging and sports activity data from the Fitbit Versa 2 smartwatch and the PMSys sport logging app. The dataset spans five months and includes records from 16 individuals, enabling analyses such as predicting weight, sleep patterns, and sports activities. It contains comprehensive metrics such as calories burned, distance moved, heart rate, steps, time in heart rate zones, and sleep patterns. Additionally, detailed exercise data is provided, including start and end times, duration, activity type, and specific performance metrics such as distance, time, steps, speed, and other activities.

6.2 Preprocessing

Although the dataset was preprocessed, concerns about data dimensionality remained. Dimensionality reduction is crucial for LLM processing, especially in healthcare monitoring, as it reduces computational complexity, accelerates processing, and improves performance. It also enables cost-effective storage and transmission by feeding fewer tokens to the LLM. Imputation techniques were employed to address missing data, and the data was preannotated for tasks such as tracking calories, steps, and sleep patterns.

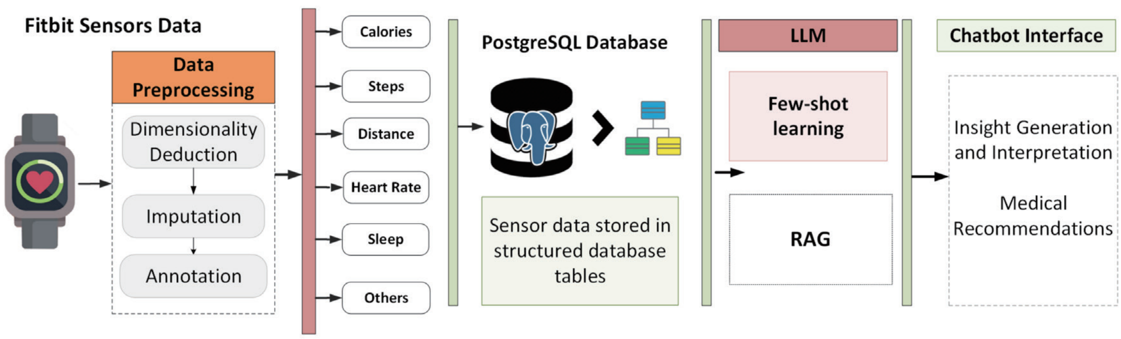


Fig. 8. (Color online) System flow of an LLM-based healthcare chatbot for elderly care using sensor data.

6.3 Data preparation for LLM processing

The crucial task involves preparing the sensor data in a format that is understandable for LLMs to process and analyze. As previously discussed, we explored various methods for transforming numerical sensor data to textual format for LLM processing. At this stage, data was transformed into tabular format and stored in a PostgreSQL database. We utilized the Text-to-SQL technique to retrieve and interpret the numerical sensor data for LLM processing. The Text-to-SQL technique leverages natural language processing (NLP) to convert natural language queries into a corresponding SQL statement, enabling database interaction through human-like queries.(74) We employed OpenAI's GPT-4 and GPT-3.5 to analyze sensor data stored in structured database tables, tailored for healthcare settings. Data such as calories, heart rate, heart rate zones, activity types, activity levels, and sleep and other patterns are organized in separate tables. We meticulously established logical relationships between each table to facilitate comprehensive data analysis.

6.4 Few-shot learning

To guide the LLMs, we performed few-shot learning. At this stage, we not only guided the model on the expected output but also helped it understand the specific schema of the database it would query. We provided the model with numerous examples of data to enable it to handle different types of query related to various sensor outputs, such as calories burned, heart rate, and sleep patterns. This approach allowed the LLM to efficiently manage these queries with specific training on sensor data. To enhance the model's performance in generating accurate queries, we focused on refining our prompting strategies. Initially, we implemented simple prompts, asking straightforward questions without specifying any parameters. Next, we applied dialect-specific prompting, tailoring LLM to effectively handle the specific SQL dialect of our PostgreSQL database. Additionally, we incorporated elements of the database schema, such as table names, into our prompts to provide the model with relevant context. The most effective strategy involved dynamic few-shot example prompt engineering, as shown in Fig. 9. With a robust set of examples at our disposal, we selectively incorporated only the most pertinent examples into the prompts. This approach helps avoid overwhelming the model with less relevant data that could fall outside its context window or distract from the primary task. To facilitate this process, all the examples were stored in a vector database. During runtime, a similarity search was conducted to match the input with examples, selecting those that were most semantically similar for use in the models' prompts.

The semantic similarity was calculated for a set of prompt examples $E = \{e_1, e_2, ..., e_m\}$ and a set of input text $I = \{i_1, i_2, ..., i_n\}$. Each set was converted to vector representation $V_E = \{V_{e1}, V_{e2}, ..., V_{em}\}$ and $V_I = \{V_{i1}, V_{i2}, ..., V_{in}\}$. The prompt set E was systematically crafted using a hybrid approach that combined automated generation with manual validation for thoroughness and precision. A python script was employed to programmatically generate diverse prompts based on predefined templates. To reduce omissions, a coverage-based approach ensured that all relevant query types were systematically included. This approach begins by identifying broad categories

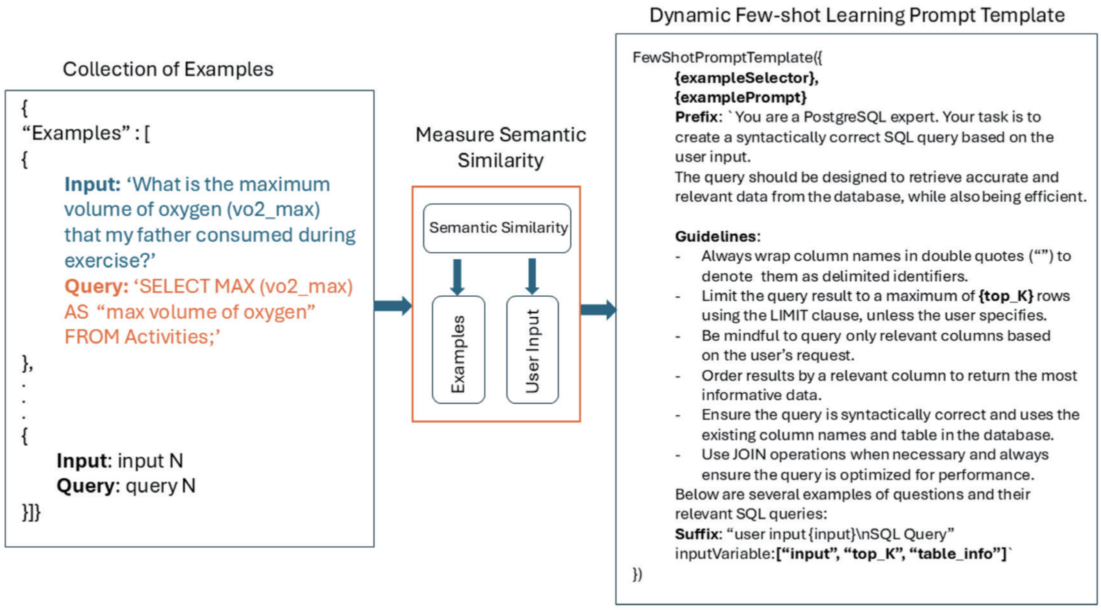


Fig. 9. (Color online) Dynamic few-shot learning prompt design.

of expected queries. The prompt space P was defined as a combination of categories C (e.g., data retrieval, aggregation, anomaly detection) and their associated parameters P_i. For each category $c_i \in C$, a template T_i with placeholders was defined as

$$T_i = f(c_i, P_i)\,. \tag{1}$$

Here, P_i represents the parameters for each category. The systematic variation of these parameters generated a diverse set of prompts as follows.

$$P_i = \{T_i(p_{i1}, p_{i2}, \ldots, p_{ik}) \mid p_{ij} \in P_i\} \tag{2}$$

This ensures that different scenarios and data types are represented. By combing these parameters, a large and diverse set of prompts is generated. Parameterized templates are used to automate the prompt generation process. For example, templates such as '*Retrieve the average [parameters] for [user_id] over the last days [day]*' *and* '*Find all instances where [parameter] exceeded [threshold] on [date]*' were used to produce a diverse set of prompts by varying parameters such as physiological metrics (e.g., SpO2, steps, calories), thresholds, and time periods.

To validate coverage, we mapped generated prompt examples to their categories to ensure that all query types and parameter combinations were represented. Semantic coverage was assessed using the diversity of the generated prompts within the embedding space. The complete prompt set was defined as shown in Eq. (3). Additionally, we conducted a manual review to check for logical errors.

$$P_{final} = \bigcup_{i=1}^{m} p_i \tag{3}$$

Once all the examples and inputs were represented as vectors in a high-dimensional space, semantic similarity was computed for each v_{ij} corresponding to an input text i in I represented by their vectors V_{ij} and V_{ek}, as defined below.

$$Similarity\left(V_{i_j}, V_{e_k}\right) = \frac{V_{i_j} . V_{e_k}}{\left\|V_{i_j}\right\| . \left\|V_{e_k}\right\|} \tag{4}$$

Here, "·" denotes the dot product, and $\|V\|$ denotes the norm of the vector. A similarity matrix S was constructed where each element S_{jk} represents the similarity between the j-th input text and the k-th prompt examples, as follows.

$$S = \begin{pmatrix} Sim(V_{i_1}, V_{e_1}) & \dots & Sim(V_{i_1}, V_{e_m}) \\ \vdots & \ddots & \vdots \\ Sim(V_{i_n}, V_{e_1}) & \cdots & Sim(V_{i_n}, V_{e_m}) \end{pmatrix} \tag{5}$$

For each row in matrix S, which corresponds to an input text, the column indices with the highest similarity score were identified. This process determines which prompt examples are most semantically similar to each input text.

We implemented RAG to enhance the LLM's accuracy in Q&A tasks and recommendations for elderly care. RAG retrieves relevant information and integrates it into the model's prompt, enriching its domain-specific knowledge. By combining retrieval and generation approaches, RAG involves indexing documents such as SQL templates, database schemas, and sensor data contextual information. The LLM processes queries using this retrieved context, which helps it to understand natural language more effectively and generate accurate SQL queries. In this system, a natural query is processed by searching the indexed documents for relevant content. The LLM generates an SQL query based on this retrieved context and executes it on the PostgreSQL database. For instance, when asked, "What is the average heart rate of my father today?" the system retrieved relevant schema information and generated the SQL query: 'SELECT AVG (heart_rate) FROM hr_table WHERE date = CURRENT_DATE;'. The execution returned the correct average heart rate with a personalized recommendation, demonstrating the system's effectiveness. Beyond simple queries, the framework also handles complex scenarios such as retrieving trends, identifying anomalies, and aggregating metrics for broader analysis. Examples are shown in Table 3.

As we used OpenAI's chat models such as GPT-4 and GPT-3.5, we optimized them by adjusting parameters as outlined in Table 4. Temperature controls response randomness, whereas topP limits token selection to a specified probability mass for nucleus sampling. FrequencyPenalty reduces repetition by penalizing tokens that have already appeared. Since the

Table 3
Examples of user input with associated generated queries and LLM response.

User input	Generated query	LLM analysis/response
Tell me the maximum heart rate and blood oxygen saturation during sleep of my father during last night.	SELECT MAX (hr.heart_rate) AS max_heart_rate, MAX (vs.spo2) AS max_spo2 FROM heart_rate hr JOIN vital_signs vs ON hr.user_id = vs.user_id AND hr.timestamp = vs.timestamp JOIN sleep_table st ON hr.user_id = st.user_id WHERE hr.timestamp BETWEEN st.sleep_start AND st.sleep_end AND DATE (st.sleep_start) = CURDATE() - INTERVAL 1 DAY;	Your father's maximum heart rate during sleep is 105 bpm, slightly elevated. Consider checking if your father had a restless night or if there are underlying factors such as stress or sleep apnea. The SpO2 level (98%) is within normal range (95–100%). Encourage maintaining a healthy environment for sleeping such as optimal room temperature and ensuring no interruptions.
Are there any abnormalities in my father's heart rate during sleep last night?	SELECT COUNT (*) AS hr_count FROM heart_rate hr JOIN sleep_table st ON hr.user_id = st.user_id WHERE hr.timestamp BETWEEN st.sleep_start AND st.sleep_end AND hr.heart_rate > 100 AND DATE (st.sleep_start) = CURDATE() -INTERVAL 1 DAY;	There were three instances of elevated heart rate during sleep. Consider monitoring for stress or other potential health issues.
How was my father's sleep quality last night? Can you analyze the sleep levels?	SELECT SUM(light) AS total_light_sleep, SUM(deep) AS total_deep_sleep, SUM(rem) AS total_rem_sleep, SUM(wake) AS total_wake_time, TIMESTAMPDIFF(MINUTE, sleep_start, sleep_end) AS total_sleep_duration FROM sleep_table WHERE st.user_id = user_id AND DATE(sleep_start) = CURDATE() - INTERVAL 1 DAY;	Your father slept for 6 hours last night, with light sleep for 180 min. Sleep quality is good, with sufficient deep sleep (90 min). REM sleep (60 min) is slightly below ideal. Try stress-free activities before bed to improve mental recovery. Minimize wake time by optimizing his sleep environment and bedtime routine to reduce disturbances.

Table 4
Parameter optimization for GPT-4 and GPT-3.5 models.

Parameter	Value	Purpose
Temperature	0.0	To choose the most probable next token.
topP	0.9	To consider only the top 90% of most probable tokens.
FrequencyPenalty	0.5	To reduce the repetitive responses.

data was stored in a database, we implemented several feasible security measures to ensure safety. First, we presanitized user inputs to remove potential risky characters. We also used few-shot learning with examples of both safe and unsafe queries, validated the structure of LLM-generated queries to detect suspicious patterns, and converted queries into parameterized formats where feasible. Additionally, we filtered LLM outputs using regular expressions to eliminate unsafe query patterns before execution.

6.5 Mechanism for query processing and LLM-driven healthcare recommendations

The framework efficiently processes user queries and utilizes LLMs to provide personalized healthcare recommendations based on sensor data. As shown in Fig. 10, the process begins with Query Input, where users provide a natural language request or select predefined templates (e.g., "Show the average heart rate for the past week"). These inputs undergo preprocessing, where intent detection and parameter validation ensure accuracy and relevance. Next, the SQL query generation stage dynamically constructs optimized SQL queries tailored to the input, such as retrieving specific health metrics or aggregating trends over a defined period. The query

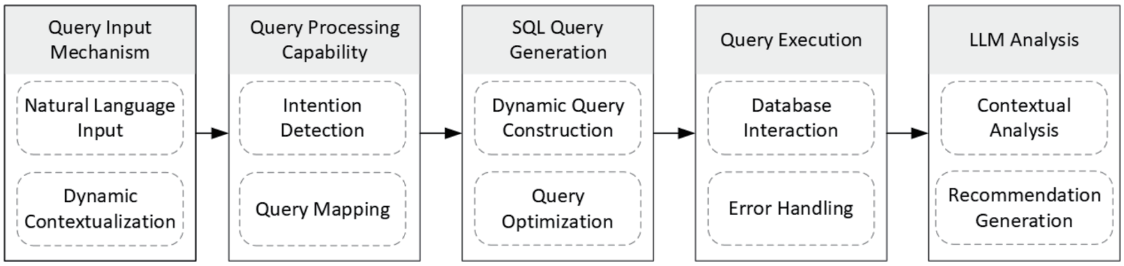


Fig. 10. (Color online) Query processing workflow using LLMs for personalized recommendation for sensor data.

execution step interacts with the PostgreSQL database to retrieve the required data while handling potential errors gracefully. The retrieved sensor data is then passed to the LLM in the data analysis stage. Here, the LLM analyzes patterns, detects anomalies, and synthesizes personalized recommendations in natural language, such as highlighting a consistent increase in heart rate or confirming no significant anomalies; examples are shown in Table 3. By focusing on LLM-driven analysis, the framework provides actionable insights directly from the processed data. This mechanism highlights the framework's versatility in handling diverse queries and its ability to deliver tailored, context-aware recommendations for effective decision-making. Each query is customized for user needs and follows the same workflow within the framework.

6.6 Performance evaluation

We evaluated LLMs in a few-shot learning scenario using sensor data by testing their ability to generate accurate SQL queries in response to caregiver inquiries about elderly health. The focus was on retrieving relevant information and providing reliable insights, with accuracy as the key metric. To facilitate this, we created a ground truth using a diverse test set of prompts that reflected the types of SQL queries expected to be executed. This test set contained a range of queries with varying complexities, including selections, aggregations, joins, and queries involving multiple clauses such as WHERE, GROUP BY, ORDER BY, and HAVING. The ground truth queries were then executed in the PostgreSQL database for validation.

The semantic similarity between generated SQL queries and ground truth was assessed to determine their alignment. This involved analyzing the structure of queries, including table names, referenced fields, and conditions used. Figure 11 shows the performance of GPT-4 and GPT-3.5 in generating SQL queries for sensor data retrieval and analysis, evaluated using query similarity scores across prompts of varying complexity, which is determined by the number of conditions, joins, and nested queries. In simple prompts, the initial difference showed that GPT-4 performed better than GPT-3.5. Both models showed improvement with dialect-specific prompts; however, the gap between the two models remained consistent, with GPT-4 maintaining higher similarity. Further refinements by adding schema information in the prompt led to another increase in similarity, where GPT-3.5 improved significantly, narrowing the gap between the two models. With few-shot examples, both models experienced a substantial jump. However, the gap widened slightly again in favor of GPT-4.

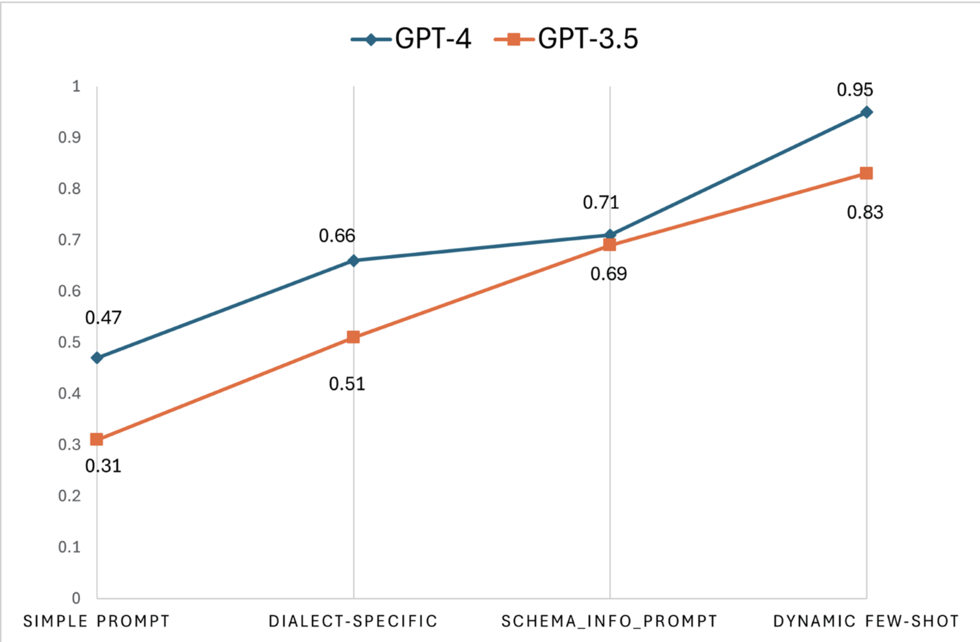


Fig. 11. (Color online) Results of GPT-4 and GPT-3.5 performance evaluation based on sematic similarity.

In the dynamic few-shot prompts, GPT-4 reached its highest similarity value of 0.95. GPT-4 maintained a lead throughout, with the largest difference observed in this final prompt type. Overall, GPT-4 consistently outperformed GPT-3.5 across all prompt types, with increasing similarity as the prompts became more complex or provided more contexts. Furthermore, as shown in Tables 5 and 6, we employed evaluation metrics such as precision, recall, and F1-score. Precision is the proportion of correctly generated queries among all generated queries, as shown in Eq. (6), recall is the proportion of correctly generated queries among all relevant queries, as shown in Eq. (7), and F1-score is the harmonic mean of precision and recall, as shown in Eq. (8).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

We conducted a detailed confusion matrix analysis to evaluate the performance of the models. This analysis comprises key components: True Positive (TP) cases are reviewed to understand patterns of success, such as how correct queries are generated. False Positive (FP) cases are analyzed to understand instances where the model incorrectly generated a query. False Negative (FN) cases are identified to investigate why the model failed to recognize a correct query. True Negative (TN) cases are examined to confirm instances where the model correctly

Table 5
F1-score, precision, and recall evaluation for GPT-4 in (%).

Prompt strategy	Precision	Recall	F1-score
Simple prompt	38.01	44.00	40.02
Dialect-specific	64.55	71.62	68.21
Few-shot examples	71.02	55.59	62.50
Dynamic few-shot	88.55	98.92	93.40

Table 6
F1-score, precision, and recall evaluation for GPT-3.5 in (%).

Prompt strategy	Precision	Recall	F1-score
Simple prompt	31.33	35.21	33.14
Dialect-specific	49.51	53.11	51.28
Few-shot examples	61.44	54.00	57.96
Dynamic few-shot	66.67	60.00	62.80

determined that no query was needed.Zero or low TN can indicate issues in detecting irrelevant cases. The confusion matrix served as a tool to identify basic error patterns and assess the model's behavior in various scenarios, providing useful metrics to evaluate how many query components were correctly generated, ultimately leading to better results.

Tables 5 and 6 present results for F1-score, precision, and recall, corresponding to the GPT-4 and GPT-3.5 models, respectively. GPT-4 consistently outperformed GPT-3.5 across all prompt strategies, with the most significant differences observed in the dynamic few-shot scenario, where GPT-4 achieved a much higher F1-score of 93.40% compared with GPT-3.5's 62.80%. The superior performance of GPT-4 can be attributed to its advanced architecture, which allows for better contextual understanding and more accurate generation of response. Both models improved as prompt complexity increased, but GPT-4 showed more substantial gains, particularly in balancing precision and recall.

While few-shot examples improved precision in both models, recall gains were less significant, particularly for GPT-3.5. The dynamic few-shot strategy proved to be most effective for both, with GPT-4 excelling in nearly all metrics, showcasing its superior ability to generate contextually accurate responses. In contrast, although GPT-3.5 benefited from these strategies, it fell short of GPT-4's performance level owing to its relatively less sophisticated model design. The findings show that the accuracy of the sensor data interpretation by the LLM depends on the precision of SQL queries in retrieving data based on user input. Generating valid queries to extract sensor data from the database is crucial, and this process heavily relies on effective prompt engineering strategies.

As shown in Table 7, the results demonstrate a clear improvement in performance from the simple prompt to the dynamic few-shot approach when considering successful and failed cases. The simple prompt had a high number of failed cases (FN = 25, FP = 33), indicating limited contextual understanding and frequent mistakes in query generation. For instance, it often failed to include essential conditions such as "WHERE date = 2019-06-25" or added irrelevant joins, resulting in fewer successful cases (TP = 22, TN = 13). The dialect-specific prompt improved the number of successful cases (TP = 48) by better handling queries within a specific dialect, reducing FN (19). However, it failed to recognize when no query was required, generating irrelevant outputs (e.g., retrieving temperature instead of heart rate), leading to no TN (0) and a significant number of FP (26).

The few-shot examples strategy further increased the numbers of TP (35) and TN (16) by retrieving more specific queries, such as "SELECT calories, heart rate, steps FROM Activities WHERE date = '2019-07-20'". However, it still struggled with edge cases FN (28) and had a moderate number of failed cases FP (14). The dynamic few-shot strategy delivered the best

Table 7
GPT-4 successful and failed cases for SQL query generation based on confusion matrix analysis.

Prompt strategy	Successful cases (TP + TN)	Failed cases (FP + FN)
Simple prompt	22 + 13 = 35	33 + 25 = 58
Dialect-specific	48 + 0 = 48	26 + 19 = 45
Few-shot examples	35 + 16 = 51	14 + 28 = 42
Dynamic few-shot	62 + 22 = 84	8 + 1 = 9

performance, with the highest number of successful cases, which includes TP (62) and TN (22), and the lowest number of failed cases FN (1) and FP (8). It effectively handled complex queries, such as aggregating summaries, and trends over time, demonstrating remarkable adaptability and precision. Overall, these results highlight a significant reduction in the number of failed cases and an increase in the number of successful cases as the prompt strategies became more sophisticated, with the dynamic few-shot approach achieving the most accurate and reliable outcomes.

Table 8 provides analysis results of GPT-3.5's performance across prompt strategies for SQL query generation in terms of successful and failed cases. The simple prompt strategy had a lower number of successful cases with TP (14) and TN (22) and a high number of failed cases (57), struggling with high FN (26) and FP (31). These errors were often due to missing conditions or unnecessary columns in the generated queries. The dialect-specific strategy showed moderate improvement, with 42 successful cases and 51 failed cases. While it improved recall (TP = 27), it frequently added irrelevant filters (FP = 28), reducing its overall effectiveness. The few-shot examples strategy further increased the number of successful cases to 53 and reduced the number of failed cases to 40. It performed well with simple trend analysis queries such as "SELECT MAX (heart_rate) FROM hr_table WHERE date = '2019-06-10';", but FP (23) persisted owing to inconsistent handling of edge cases. The dynamic few-shot strategy achieved the high number of successful cases (58) and the lowest number of failed cases (35). It effectively handled basic aggregation queries such as "SELECT COUNT (*) FROM Activities WHERE heart_rate > 100;", demonstrating significant improvements in both precision and recall. While GPT-3.5 performed worse than GPT-4, the dynamic few-shot strategy showed the greatest effectiveness by substantially reducing errors and increasing successful outcomes.

Moreover, the findings suggest that refining the few-shot examples provided to the model or adjusting prompts iteratively enhances the performance, helping to fine-tune the few-shot learning setup for better SQL query generation. The semantic similarity evaluation method used in this experiment is a flexible approach that examines the intent and structure of the query rather than exact syntax, performing a deep analysis of the model's understanding and handling of SQL queries. However, this method requires advanced NLP efforts, which may not be consistently replicable and might not perform well with large-scale databases where more complex queries are needed.

The precision and recall metrics are more component-focused, and indicate the effectiveness of the specific parts of an SQL query, such as how well certain clauses are generated. However, these methods are complex because they require breaking down the query into components and determining the relevance of each. Without a clear understanding of what constitutes TP, FP, and FN in the context of SQL query components, applying these metrics correctly can be challenging.

Table 8
GPT-3.5 successful and failed cases for SQL query generation based on confusion matrix analysis.

Prompt strategy	Successful cases (TP + TN)	Failed cases (FP + FN)
Simple prompt	14 + 22 = 36	31 + 26 = 57
Dialect-specific	27 + 15 = 42	28 + 23 = 51
Few-shot examples	27 + 26 = 53	17 + 23 = 40
Dynamic few-shot	30 + 28 = 58	15 + 20 = 35

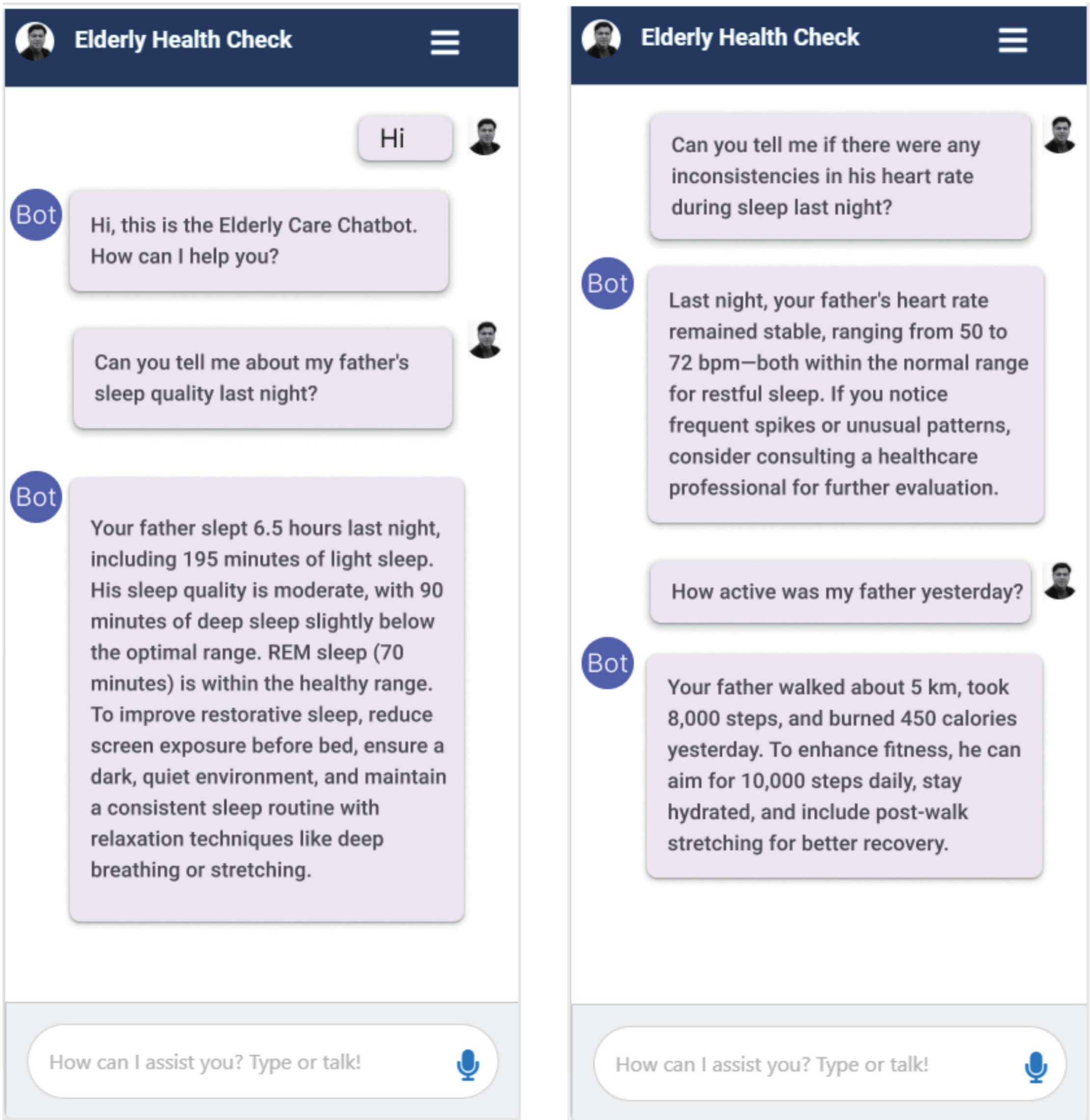
6.7 Chatbot interface

A user-friendly web-based chatbot application, as shown in Figs. 12(a) and 12(b), was developed in NodeJS for caregivers to check the health status of the elderly. The caregiver can interact with the system using text or voice commands. Before the user input is passed to LLM, it undergoes validation to remove any unexpected irrelevant content. This process ensures that only the essential parts of the input remain, transforming it into a clear, standalone question that is easy for LLM to understand. The chatbot generates responses from sensor data collected from the elderly individual. It gracefully handles unexpected inputs by providing meaningful messages to guide the user back on track. It also has a fallback response mechanism. For example, if the chatbot cannot understand a query, it gives a generic response such as "I am not sure I understood that, could you please rephrase?". For managing LLMOps, we used the LangChainJs(183) framework, which is specifically tailored for developing LLM-based applications within a NodeJS environment. It provides a well-orchestrated pipeline for LLMOps. The LLM delivers well-formatted and human-readable insights in response to the user's inquiry via the chatbot interface. Additionally, the chatbot offers recommendations based on insights derived from sensor data and retains this information in its memory for future interactions.

7. Discussion

In this study, we proposed a methodological framework to integrate diverse types of sensor data with LLMs for elderly care. The types of sensors and their specific processing techniques for elderly care were identified through a comprehensive systematic review. In the review, we found six major categories of sensors used for diverse applications in elderly care. These categories include motion, physiological, environmental, vision-based, body thermal, and biochemical sensors. This identification of sensor technologies and their associated data types served as a foundational understanding of data sources, data collection, preprocessing, and integration, all of which are key components of the proposed framework. By incorporating such diverse sensor data, the proposed framework can generalize the adaption of LLMs for various healthcare applications in home settings, such as vital sign monitoring,(91–99,184) fall detection,(34–40,83) cognitive impairment,(110,113) daily life activity monitoring,(32,84–86,104,185) physical activity recognition,(91) kinematic analysis,(81,82) ambient assisted living,(100) and social isolation.(186)

The proposed framework demonstrates that LLMs have the ability to analyze physiological sensor data to derive insights about vital sign parameters such as heart rate, caloric expenditure,



(a) (b)

Fig. 12. (Color online) Chatbot Interface: (a) initial user input and (b) continuation of conversation.

heart variability, sleep patterns, oxygen situation, active zone minutes, and physical activity patterns such as steps, walking, and running. These sensors serve as crucial data sources for LLM-based applications in monitoring the vital signs of elderly adults. Given this data, LLMs function as few-shot learners, effectively learning from physiological sensor data to provide valuable health insights.[170,187] Each sensor class provides a unique benefit for elderly care. Integrating this data into LLMs can considerably enhance elderly care through continuous monitoring and quick response. However, addressing privacy and data accuracy is crucial to maximizing benefits while maintaining the dignity and trust of the elderly adults.[188,189]

Integrating LLMs with sensor-generated data is challenging owing to various data formats. Transforming signal-based and numerical data into a processable format is essential. Encoding methods such as natural string encoding,[169] statistical summary encoding,[170,190] modality-based encoding,[63] and Text-to-SQL[74,171] enhance the capabilities of LLMs. However, modality-based encoding, although powerful, is complex and may lead to loss of detail, while natural string encoding is simpler but may not handle complex data adequately. In this study, we found that despite the encoding methods mentioned, well-structured sensor data stored in databases can serve as a knowledge store for LLMs. LLMs can access and process this data by interacting with database tables via the Text-to-SQL method. In the experimental results, we found an improvement in efficiency in data retrieval. Text-to-SQL can quickly extract only the relevant data needed for a specific context, reducing the processing burden. However, converting all sensor data into textual format is resource intensive.[44,47]

This process is enabled by few-shot learning, rather than extensive fine-tuning of the model's parameters.[170] Few-shot learning significantly enhances an LLM's capabilities to generate accurate SQL queries for retrieving and analyzing sensor data. This approach is particularly useful in LLM-based Q&A applications, such as healthcare chatbots. However, it requires more prompt engineering efforts because the output of the LLM depends on the robustness and the accuracy of the generated query to retrieve the sensor data from the database tables.

Validating queries is essential and can be done by evaluating the LLM's performance in generating them by methods such as few-shot learning. In this study, we assessed query accuracy by comparing their semantic structure with the ground truth, assigning a similarity score between 0 and 1, with higher values indicating greater alignment. An alternative approach measures LLM accuracy by calculating the ratio of successful matches to total queries, including both successes and failures.[191] While the semantic similarity approach matches all the semantic components of the queries (generated and ground truth), leading to identical results from the database, the accuracy of this approach was found to be more dependent on how comprehensive the ground truth is.

Our proposed framework can be integrated into existing health monitoring systems as a multimodal solution, combining sensor data, temporal information, and LLM-driven health insights. This integration is technically feasible through API services, which enhance interoperability. However, additional improvements are required to increase generalizability. For example, it is necessary to implement the automation of a data preprocessing pipeline that directly handles health monitoring data to ensure expected and reliable performance.[61,192] Prompt optimization would also be needed to improve the consistency and uncertainty for LLM health predictions with existing systems.[117] To further enhance the scalability of the proposed framework for seamless integration with existing systems, implementing a robust edge-cloud architecture would be essential to reduce the latency and bandwidth usage.[117]

Despite these advantages of the proposed framework, there are certain limitations that necessitate further exploration. In this study, we aimed to improve elderly care in home settings; however, some geographical limitations may arise. The framework relies on IoT infrastructure for collecting sensor data, so regions with limited IoT infrastructure or connectivity may face challenges in deploying the system effectively. Additionally, some regions may have stricter data privacy laws for AI in healthcare, making regulatory compliance a critical consideration when deploying the system. The success of the framework also heavily relies on the ability of caregivers and elderly individuals to effectively use the technology. In regions with low digital literacy, this could pose a significant barrier.

Additionally, to make sensor data interpretable for LLMs, data is stored in an SQL database, which may raise security concerns, particularly for Q&A systems that execute model-generated SQL queries. Robust security practices such as strict connection permission and private deployment of LLM can help address this issue. The current evaluation primarily focused on the semantic structure and the accuracy of query building components, rather than holistic results of the queries or entire system performance. While outcomes are affected by the nature of the queries, discrepancies may arise owing to the LLM misinterpreting user input or errors in schema understanding; these can potentially lead to unexpected results. To address these limitations, future work will include a comprehensive performance evaluation of the entire system in real-world settings. In this evaluation, we will assess key aspects such as accuracy, usability, reliability, and security when integrated with live sensor data to provide caregivers with actionable health insights and personalized recommendations for elderly care. Additionally, we aim to conduct controlled experiments and real-time case studies to systematically evaluate the system's workflow and its ability to provide caregivers with actionable health insights and personalized recommendations for elderly care. This holistic approach will help identify potential bottlenecks, improve system performance, and enhance practical applicability in elderly care scenarios.

8. Challenges and Opportunities

The integration of LLMs with various sensor data presents unique challenges and opportunities. Understanding these aspects is crucial for effectively leveraging LLMs in processing and interpreting sensor data. LLMs function as black-box models, making their decision-making processes difficult to understand. This lack of interoperability and transparency poses significant challenges when analyzing sensor data with LLMs in healthcare settings.(61) Developing explainable AI techniques for LLMs can help stakeholders understand the factors influencing the model's predictions, enabling more informed decision-making based on the outputs.(48) Attention mechanisms can be used to identify the most relevant parts of the input data, providing transparency in the model's decision-making process.(193) These approaches can offer insights into how LLMs process and analyze data.

Integrating physiological and motion sensor data with LLMs is complex, requiring data harmonization, noise filtering for IMU data, and strict privacy for ECG and photoplethysmography data. This demands robust data fusion techniques and precise synchronization of multimodal data.(194,195) LLMs, being computationally intensive, may face challenges in adapting sensor data in real time, which is critical for emergency response. Optimizing models and training algorithms can help mitigate this issue. For instance, incorporating machine learning techniques with LLMs can provide a balanced solution for accurately integrating real-time data.(196) Additionally, edge computing can be a promising approach for distributing the computational load.(197)

9. Conclusions

In this study, we addressed a critical gap in elderly care by proposing and implementing a framework that integrates diverse sensor technology for collecting data, including physiological, environmental, and biochemical data, with LLMs to enhance health monitoring and support for caregivers. Our review of existing literature identified key sensor types and their data outputs such as physiological, motion, environmental, and biochemical sensor and data, alongside processing techniques that enable real-time health monitoring, forming a solid foundation for the proposed framework. These sensor-driven data streams provide timely and actionable insights, ensuring proactive elderly care. Through few-shot learning, GPT-4 demonstrated that properly structured, real-time sensor data from heterogeneous sources can be effectively interpreted by LLMs to generate clinically relevant insights, outperforming GPT-3.5.

The results reveal significant benefits in enhancing caregiver support, underscoring the potential of this approach to transform elderly care. However, the study's generalizability is currently limited by the exclusive use of the GPT-4 and GPT-3.5 models. In future work, we will expand the framework by integrating other LLMs such as Llama3, DeepSeek, and domain-specific models such as PubMedQA, BioGPT, and ClinicalGPT, assessing their accuracy and response time, and the relevance of LLM-generated queries to users' queries. Key challenges include adaptability to diverse data formats, data privacy, scalability, and bias reduction. Addressing these challenges will strengthen the framework's robustness and applicability. A user study will also be conducted to evaluate the chatbot interface and gather caregiver feedback, ensuring that the system meets practical needs. Despite its limitations, this study sets a new benchmark for advancing AI-driven elderly care through the innovative integration of sensor technology and LLMs. The synergy between sensor-driven health monitoring and AI-powered analytics holds immense promise for significantly improving the quality of life for the elderly, marking an important step forward in the evolution of personalized, sensor-assisted healthcare solutions.

Acknowledgments

This work was supported by the Petchra Pra Jom Klao Research Scholarship from King Mongkut's University of Technology Thonburi (KMUTT). The authors extend their gratitude to the members of the Innovative Cognitive Computing Research Centre (IC2) at KMUTT for their invaluable support.

References

1 United Nations Department of Economic and Social Affairs, World Population Ageing 2020 Highlights (United Nations, 2021) pp. 44–45. https://doi.org/10.18356/9789210051934
2 World Health Organization (WHO): https://www.who.int/europe/news/item/11-10-2023-by-2024--the-65-and-over-age-group-will-outnumber-the-youth-group--new-who-report-on-healthy-ageing (accessed April 2024).
3 Y. Yang, X. Guo, M. Zhu, Z. Sun, Z. Zhang, T. He, and C. Lee: Adv. Energy Mater. **13** (2023) 2203040. https://doi.org/10.1002/aenm.202203040

4 Y. Cho, J. Park, C. Lee, and S. Lee: Bioelectron. Med. **6** (2020) 23. https://doi.org/10.1186/s42234-020-00059-z
5 L. Zhang, K. S. Kumar, H. He, C. J. Cai, X. He, H. Gao, S. Yue, C. Li, R. C. Seet, H. Ren, and J. Ouyang: Nat. Commun. **11** (2020) 4683. https://doi.org/10.1038/s41467-020-18503-8
6 J. Hanna, M. Bteich, Y. Tawk, A. H. Ramadan, B. Dia, F. A. Asadallah, and A. A. Eid: Sci. Adv. **6** (2020) eaba5320. https://doi.org/10.1126/sciadv.aba5320
7 J. Dai, J. Meng, X. Zhao, W. Zhang, Y. Fan, B. Shi, and Z. Li: Adv. Mater. Technol. **8** (2023) 2201535. https://doi.org/10.1002/admt.202201535
8 S. Gao, T. He, Z. Zhang, H. Ao, H. Jiang, and C. Lee: Adv. Sci. **8** (2021) 2101834. https://doi.org/10.1002/advs.202101834
9 M. Hämäläinen, L. Mucchi, S. Caputo, L. Biotti, L. Ciani, D. Marabissi, and G. Patrizi: Sensors **21** (2021) 3158. https://doi.org/10.3390/s21093158
10 K. Dong and Z. L. Wang: J. Semicond. **42** (2021) 101601. https://doi.org/10.1088/1674-4926/42/10/101601
11 I. Poitras, F. Dupuis, M. Bielmann, A. Campeau-Lecours, C. Mercier, L. J. Bouyer, and J.-S. Roy: Sensors **19** (2019) 1555. https://doi.org/10.3390/s19071555
12 S. M. Ellison, S. R. Mghabghab, and J. A. Nanzer: IEEE Sens. J. **22** (2022) 1629. https://doi.org/10.1109/JSEN.2021.3130793
13 Y. LeCun, Y. Bengio, and G. Hinton: Nature **521** (2015) 436. https://doi.org/10.1038/nature14539
14 F. Luna-Perejón, L. Muñoz-Saavedra, J. M. Castellano-Domínguez, and M. Domínguez-Morales: Biomed. Signal Process. Control **69** (2021) 102848. https://doi.org/10.1016/j.bspc.2021.102848
15 G. Pires, A. Lopes, P. Correia, L. Almeida, L. Oliveira, R. Panda, D. Jorge, D. Mendes, P. Dias, N. Gomes, and T. Pereira: Univ. Access Inf. Soc. **22** (2023) 525. https://doi.org/10.1007/s10209-021-00859-3
16 E. M. Aldahasi, M. M. Alhejazi, and E. A. Alyahayan: Proc. 2020 8th Int. Conf. Wireless Networks and Mobile Communications (IEEE, 2020) 1–6. https://doi.org/10.1109/WINCOM50532.2020.9272412
17 A. S. Kumar, D. Gobinath, N. Nithiyanandam, P. Vijayakarthik, S. Dhanasekaran, and G. V. Rajkumar: Proc. 2022 Int. Conf. Inventive Computation Technologies Conf. (IEEE, 2022) 1016–1023. https://doi.org/10.1109/ICICT54344.2022.9850896
18 S. Campanella, A. Alnasef, L. Falaschetti, A. Belli, P. Pierleoni, and L. Palma: IEEE Sens. J. **24** (2024) 15219. https://doi.org/10.1109/JSEN.2024.3375603
19 K. Singhal, T. Tu, J. Gottweis, R. Sayres, E. Wulczyn, M. Amin, L. Hou, K. Clark, S. Pfohl, H. Cole-Lewis, D. Neal, Q. M. Rashid, M. Schaekermann, A. Wang, D. Dash, J. H. Chen, N. H. Shah, S. Lachgar, P. A. Mansfield, S. Prakash, B. Green, E. Dominowska, B. Agüera y Arcas, N. Tomašev, Y. Liu, R. Wong, C. Semturs, S. S. Mahdavi, J. K. Barral, D. R. Webster, G. S. Corrado, Y. Matias, S. Azizi, A. Karthikesalingam, and V. Natarajan: Nat. Med. **31** (2025) 1. https://doi.org/10.1038/s41591-024-03423-7
20 B. Meskó and E. J. Topol: NPJ Digit. Med. **6** (2023) 873. https://doi.org/10.1038/s41746-023-00873-0
21 Z. Ren, Y. Zhan, B. Yu, L. Ding, and D. Tao: arXiv (2024). https://doi.org/10.48550/arXiv.2402.13408
22 H. Zhao, Z. Liu, Z. Wu, Y. Li, T. Yang, P. Shu, S. Xu, H. Dai, L. Zhao, G. Mai, N. Liu, and T. Liu: arXiv (2024). https://arxiv.org/abs/2401.11641 (accessed April 2024).
23 Y. Li, S. Wang, H. Ding, and H. Chen: Proc. 4th ACM Int. Conf. AI Finance (ICAIF 2023) (ACM, 2023) 374–382. https://doi.org/10.1145/3604237.3626869
24 M. Nasseri, P. Brandtner, R. Zimmermann, T. Falatouri, F. Darbanian, and T. Obinwanne: Proc. HCI Int. 2023 – Late Breaking Papers (Springer, 2023) 14059. https://doi.org/10.1007/978-3-031-48057-7_12
25 M. Vidgof, S. Bachhofner, and J. Mendling: Proc. BPM Forum 2023 Conf. (Springer, 2023) 490. https://doi.org/10.1007/978-3-031-41623-1_7
26 C. Xiao, S. X. Xu, K. Zhang, Y. Wang, and L. Xia: Proc. 18th workshop on Innovative Use of NLP for Building Educational Applications Conf. (BEA 2023) 610–625. https://doi.org/10.18653/v1/2023.bea-1.52
27 M. M. Rahman and Y. Watanobe: Appl. Sci. **13** (2023) 5783. https://doi.org/10.3390/app13095783
28 S. Moore, N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, and O. C. Santos: Proc. Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky, AIED 2023 Conf. (Springer, 2023) 1831. https://doi.org/10.1007/978-3-031-36336-8_4
29 S. Kernan Freire, C. Wang, M. Foosherian, S. Wellsandt, S. Ruiz-Arenas, and E. Niforatos: Front. Artif. Intell. **7** (2024) 1293084. https://doi.org/10.3389/frai.2024.1293084
30 L. Makatura, M. Foshey, B. Wang, F. Hähnlein, P. Ma, B. Deng, M. Tjandrasuwita, A. Spielberg, C. E. Owens, P. Y. Chen, A. Zhao, A. Zhu, W. J. Norton, E. Gu, J. Jacob, Y. Li, A. Schulz, and W. Matusik: Large Language Models for Design and Manufacturing. An MIT Exploration of Generative AI (March 2024). https://mit-genai.pubpub.org/pub/nmypmnhs/release/2 (accessed May 2024).

31 R. Aburukba, A. Sagahyroon, L. T. Kamel, A. M. Al-Shamsi, H. Surti, and E. Sajwani: Proc. 2020 IEEE Int. Conf. E-health Networking, Application & Services (IEEE, 2021) 1–6. https://doi.org/10.1109/HEALTHCOM49281.2020.9231276
32 K. Swathi, S. B. Latha, M. N. Vadlamudi, M. A. Alkhafaji, K. S. Mohsen, and K. Saikumar: Proc. 2023 IEEE 8th Int. Conf. Convergence in Technology (IEEE, 2023) 1–6. https://doi.org/10.1109/I2CT45611.2023.9074223
33 S. G. Yoo and D. Oh: Int. J. Eng. Bus. Manag. **10** (2018) 1. https://doi.org/10.1177/1847979018787905
34 D. Yacchirema, J. S. de Puga, C. Palau, and M. Esteve: Pers. Ubiquit. Comput. **23** (2019) 801. https://doi.org/10.1007/s00779-018-01196-8
35 S. Vimal, Y. H. Robinson, S. Kadry, H. V. Long, and Y. Nam: J. Internet Technol. **22** (2021) 173. https://doi.org/10.3966/160792642021012201017
36 G. L. Santos, P. T. Endo, K. H. C. Monteiro, E. S. Rocha, I. Silva, and T. Lynn: Sensors **19** (2019) 1644. https://doi.org/10.3390/S19071644
37 D. Mrozek, A. Koczur, and B. Małysiak-Mrozek: Inf. Sci. **537** (2020) 132. https://doi.org/10.1016/j.ins.2020.05.070
38 S. B. Khojasteh, J. R. Villar, C. Chira, V. M. González, and E. de la Cal: Sensors **18** (2018) 1350. https://doi.org/10.3390/S18051350
39 M. Kepski and B. Kwolek: IET Comput. Vis. **12** (2018) 48. https://doi.org/10.1049/IET-CVI.2017.0119
40 A. Godfrey: Maturitas **100** (2017) 16. https://doi.org/10.1016/j.maturitas.2017.03.317
41 E. Casilari-Pérez and F. García-Lagos: Expert Syst. Appl. **138** (2019) 112811. https://doi.org/10.1016/j.eswa.2019.07.028
42 M. Ahmed, N. Mehmood, A. Nadeem, A. Mehmood, and K. Rizwan: Healthc. Inform. Res. **23** (2017) 147. https://doi.org/10.4258/HIR.2017.23.3.147
43 P. Bellagente, C. Crema, A. Depari, P. Ferrari, A. Flammini, G. Lanfranchi, G. Lenzi, M. Maddiona, S. Rinaldi, E. Sisinni, and G. Ziliani: Proc. 2018 IEEE Sensors Applications Symposium Conf. (IEEE, 2018) 1–6. https://doi.org/10.1109/SAS.2018.8336753
44 B. Yang, S. Jiang, L. Xu, K. Liu, H. Li, G. Xing, H. Chen, X. Jiang, and Z. Yan: arXiv (2024). https://arxiv.org/abs/2405.12541 (accessed May 2024).
45 E. Corda, S. M. Massa, and D. Riboni: Future Internet **16** (2024) 46. https://doi.org/10.3390/fi16020046
46 Z. Englhardt, C. Ma, M. E. Morris, C.-C. Chang, X. Xu, L. Qin, D. McDuff, X. Liu, S. Patel, and V. Iyer: Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies Conf. (ACM, 2024) 1–25. https://doi.org/10.1145/3659604
47 V. Ragavan: Automated Health Coaching: A Study on Integrating Wearable Data with Large Language Models (master's thesis, University of Illinois at Chicago, 2024). https://www.proquest.com/openview/acca4d6a2a07677a7a69bd8b15e1990b/1?cbl=18750&diss=y&pq-origsite=gscholar (accessed August 2024).
48 C. M. Fang, V. Danry, N. Whitmore, A. Bao, A. Hutchison, C. Pierce, and P. Maes: arXiv (2024). https://arxiv.org/abs/2406.19283 (accessed May 2024).
49 S. Ji, X. Zheng, and C. Wu: Proc. 2024 IEEE Int. Workshop on Foundation Models foy Cyber-Physical System & Internet of Things Conf. (IEEE, 2024) 38–43. https://doi.org/10.1109/fmsys62467.2024.00011
50 H. Zhang, J. Chen, F. Jiang, F. Yu, Z. Chen, G. Chen, J. Li, X. Wu, Z. Zhiyi, Q. Xiao, X. Wan, B. Wang, and H. Li: Proc. Findings of the Association for Computational Linguistics: EMNLP 2023 Conf. (Association for Computational Linguistics, 2023) 10859–10885. https://doi.org/10.18653/v1/2023.findings-emnlp.725
51 Z. Bao, W. Chen, S. Xiao, K. Ren, J. Wu, C. Zhong, J. Peng, X. Huang, and Z. Wei: arXiv (2023). https://arxiv.org/abs/2308.14346 (accessed May 2024).
52 Y. Li, Z. Li, K. Zhang, R. Dan, S. Jiang, and Y. Zhang: Cureus **15** (2023) e40895. https://doi.org/10.7759/cureus.40895
53 C. Xu, D. Guo, N. Duan, and J. McAuley: Proc. 2023 Conf. Empirical Methods in Natural Language Process (Association for Computational Linguistics, 2023) 6268–6278. https://doi.org/10.18653/v1/2023.emnlp-main.385
54 A. Ben Abacha and D. Demner-Fushman: BMC Bioinf. **20** (2019) 511. https://doi.org/10.1186/s12859-019-3119-4
55 H. Wang, S. Zhao, Z. Qiang, N. Xi, B. Qin, and T. Liu: arXiv (2024). https://arxiv.org/abs/2401.16107 (accessed May 2024).
56 B. Li, T. Meng, X. Shi, J. Zhai, and T. Ruan: arXiv (2023). https://arxiv.org/abs/2312.02441 (accessed May 2024).
57 C. Li, C. Wong, S. Zhang, N. Usuyama, H. Liu, J. Yang, T. Naumann, H. Poon, and J. Gao: Proc. Advance Neural Information Processing Systems **36** (2024) 28541–28564.
58 T. Tu, A. Palepu, M. Schaekermann, K. Saab, J. Freyberg, R. Tanno, A. Wang, and B. Li: arXiv (2024). https://arxiv.org/abs/2401.05654 (accessed May 2024).

59 P. Schmiedmayer, A. Rao, P. Zagar, V. Ravi, A. Zahedivash, A. Fereydooni, and O. Aalami: arXiv (2024). https://arxiv.org/abs/2402.01711 (accessed May 2024).
60 Y. Wang, X. Ma, and W. Chen: Proc. Findings of the Association for Computational Linguistics Conf. (Association for Computational Linguistics, 2024) 1754–1770. https://doi.org/10.18653/v1/2024.findings-emnlp.95
61 M. Jin, Q. Yu, D. Shu, C. Zhang, L. Fan, W. Hua, S. Zhu, Y. Meng, Z. Wang, M. Du, and Y. Zhang: arXiv (2024). https://arxiv.org/abs/2402.00746 (accessed May 2024).
62 A. Goel, A. Gueta, O. Gilon, C. Liu, S. Erell, L. H. Nguyen, X. Hao, B. Jaber, S. Reddy, R. Kartha, J. Steiner, I. Laish, and A. Feder: Proc. Machine Learning Research Conf. (ML4H, 2023) 82–100. https://proceedings.mlr.press/v225/goel23a
63 A. Belyaeva, J. Cosentino, F. Hormozdiari, K. Eswaran, S. Shetty, G. Corrado, A. Carroll, C. Y. McLean, and N. A. Furlotte: Proc. Machine Learning for Multimodal Healthcare Data Conf. (Springer, 2024) 86–102. https://doi.org/10.1007/978-3-031-47679-2_7
64 J. Yuan, R. Tang, X. Jiang, and X. Hu: Proc. American Medical Informatics Association Annual Symposium Conf. (AMIA, 2023) 1324–1333.
65 Z. Chen, A. Hernández Cano, A. Romanou, A. Bonnet, K. Matoba, F. Salvi, M. Pagliardini, and S. Fan: arXiv (2023). https://arxiv.org/abs/2311.16079 (accessed May 2024).
66 Y. Labrak, A. Bazoge, E. Morin, P.-A. Gourraud, M. Rouvier, and R. Dufour: arXiv (2024). https://arxiv.org/abs/2402.10373 (accessed May 2024).
67 A. Q. Jiang, A. Sablayrolles, A. Roux, A. Mensch, B. Savary, C. Bamford, D. S. Chaplot, and D. de las Casas: arXiv (2024). https://arxiv.org/abs/2401.04088 (accessed May 2024).
68 A. Romanov and C. Shivade: Proc. 2018 Conf. Empirical Methods in Natural Language Processing (EMNLP, 2018) 1586–1596. https://doi.org/10.18653/v1/d18-1187
69 A. Ben Abacha and D. Demner-Fushman: Proc. 57th Annual Meeting of the Association of Computational Linguistics (ACL 2019) 2228–2234. https://doi.org/10.18653/v1/p19-1215
70 A. Dada, M. Bauer, A. Butler Contreras, O. A. Koraş, C. M. Seibold, K. E. Smith, and J. Kleesiek: arXiv (2024). https://arxiv.org/abs/2404.04067 (accessed June 2024).
71 J. Li, Y. Zhang, and Y. Zhang: Exploring the Prospects of LLMs in Hospital Management: A Perspective on Medical Data Inquiry (2024). https://doi.org/10.21203/RS.3.RS-3990012/V1
72 K. Singhal, S. Azizi, T. Tu, A. Mahdavi, J. Wei, R. L. Chung, A. Scales, G. Tanwani, V. M. Cole-Lewis, T. Pfohl, N. S. Seneviratne, S. S. Gamble, S. Lynn, K. Ray, S. Shmulenson, J. M. Liu, P. S. Ghorbani, S. G. Rohou, J. A. Dickerson, M. A. Ebner, A. Kumar, A. S. Chou, A. K. Karthikesalingam, G. M. Bream, G. M. Corrado, E. D. Brainard, and M. C. Matias: Nature **620** (2023) 172. https://doi.org/10.1038/s41586-023-06291-2
73 S. A. Imran, M. N. H. Khan, S. Biswas, and B. Islam: arXiv (2024). https://arxiv.org/abs/2406.14498 (accessed May 2024).
74 M. Chiras: Exploration of Different Large Language Models for Retrieval-Augmented Generation in Analyzing Wearable Running Data for Sports Physiotherapy (Bachelor's thesis, University of Twente, 2024). https://essay.utwente.nl/101017
75 J. F. Gaitan-Guerrero, J. L. Lopez, M. Espinilla, and C. Martinez-Cruz: IEEE Internet Things J. **11** (2024) 31080. https://doi.org/10.1109/JIOT.2024.3419260
76 L. Cardenas, K. Parajes, M. Zhu, and S. Zhai: Proc. 2024 IEEE 14th Annual Computing Communication Conf. (IEEE, 2024) 375–379. https://doi.org/10.1109/CCWC60891.2024.10427622
77 C. U. Om Kumar, N. Gowtham, M. Zakariah, and A. Almazyad: IEEE Access **12** (2024) 1. https://doi.org/10.1109/ACCESS.2024.3425953
78 H. Yu, P. Guo, and A. Sano: Proc. 3rd Machine Learning for Health Symposium (PMLR, 2023) 650–663.
79 D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman (the PRISMA Group): Ann. Intern. Med. **151** (2009) 264. https://doi.org/10.7326/0003-4819-151-4-200908180-00135
80 Zotero: https://www.zotero.org (accessed August 2024).
81 P. Khera, N. Kumar, and R. Das: IEEE Sens. J. **23** (2023) 23804. https://doi.org/10.1109/JSEN.2023.3305282
82 S. Du, X. Ma, J. Wang, Y. Mi, J. Zhang, C. Du, X. Li, H. Tan, C. Liang, T. Yang, W. Shi, G. Zhang, and Y. Tian: BMC Geriatr. **23** (2023) 1. https://doi.org/10.1186/S12877-023-04281-7
83 A. S. Buchman, T. Wang, S. Oveisgharan, A. R. Zammit, L. Yu, P. Li, K. Hu, J. M. Hausdorff, A. S. P. Lim, and D. A. Bennett: Sensors **23** (2023) 4152. https://doi.org/10.3390/S23084152
84 I. Rawtaer, T. H. Pink, K. B. A. Jabbar, X. Liu, W. S. Liang, and T. Ng: Int. Psychogeriatr. **32** (2020) 141. https://doi.org/10.1017/S1041610220002781
85 J. Cherian, V. Rajanna, D. Goldberg, and T. Hammond: Proc. 11th Int. Conf. Pervasive Computing Technologies for Healthcare (ACM, 2017) 48–57. https://doi.org/10.1145/3154862.3154866

86 J. Liu, J. Sohn, and S. Kim: J. Healthc. Eng. **2017** (2017) 8934816. https://doi.org/10.1155/2017/8934816
87 A. Ghaffari, O. Rahbek, R. E. K. Lauritsen, A. Kappel, S. Kold, and J. Rasmussen: Sensors **22** (2022) 5289. https://doi.org/10.3390/S22145289
88 T. Pereira, G. Pires, D. Jorge, and D. Santos: Blood Press. Monit. **25** (2020) 227. https://doi.org/10.1097/MBP.0000000000000443
89 H. Chen, L. Yang, Y. Zhao, T. Zhao, S. Li, and S. Zhao: Proc. 2022 IEEE Int. Conf. Consumer Electronic (IEEE, 2022) 93–94. https://doi.org/10.1109/ICCE-Taiwan55306.2022.9869010
90 H. Simi, B. Guggenberger, P. Holler, M. Christova, R. Pilz, and W. Staubmann: Eur. J. Public Health **32** (2022) 33. https://doi.org/10.1093/eurpub/ckac130.055
91 P. Joosen, D. Piette, J. Buekers, J. Taelman, D. Berckmans, and P. De Boever: J. Telemed. Telecare **25** (2019) 611. https://doi.org/10.1177/1357633X18790170
92 F. M. Garcia-Moreno, M. Bermudez-Edo, E. Rodríguez-García, J. M. Pérez-Mármol, J. L. Garrido, and M. J. Rodríguez-Fórtiz: Int. J. Med. Inform. **157** (2022) 104625. https://doi.org/10.1016/j.ijmedinf.2021.104625
93 K. L. Hammett, C. J. Paquin, R. J. Pasquel, and D. E. Dow: Proc. 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conf. (IEEE, 2022) 76–83. https://doi.org/10.1109/UEMCON54665.2022.9965660
94 S. Majumder, M. A. Rahman, M. S. Islam, and D. Ghosh: Proc. 4th Int. Conf. Electrical Engineering Information & Communication Technology (IEEE, 2018) 86–91. https://doi.org/10.1109/CEEICT.2018.8628077
95 K. Pei, S. Wang, Y. He, and L. Qin: Proc. IEEE 5th Information Technology, Networking, Electronic, and Automation Control Conf. (IEEE, 2021) 413–418. https://doi.org/10.1109/ITNEC52019.2021.9587101
96 O. Y. Tham, M. A. Markom, A. H. A. Bakar, E. S. M. M. Tan, and A. M. Markom: Proc. 1st Int. Conf. Information Technology, Advance Mechanical and Electrical Engineering. (IEEE, 2020) 128–133. https://doi.org/10.1109/ICITAMEE50454.2020.9398455
97 T. F. Bernadus, L. B. Subekti, and Y. Bandung: Proc. Int. Conf. ICT Smart Society (IEEE, 2019). https://doi.org/10.1109/ICISS48059.2019.8969845
98 R. K. Nath and H. Thapliyal: Proc. Computer Society Annual Symposium on VLSI (IEEE, 2021) 390–395. https://doi.org/10.1109/ISVLSI51109.2021.00077
99 Y. Gu, P. Wang, H. Duan, H. Chen, and Y. Ren: Proc. IEEE 8th Joint Int. Information Technology and Artificial Intelligence Conf. (IEEE, 2019) 1259–1262. https://doi.org/10.1109/ITAIC.2019.8785440
100 S. Bhowmick, T. Ferdous, R. Momtaz, and M. G. R. Alam: Proc. 4th Int. Conf. Information and Communication Technology (IEEE, 2021) 98–103. https://doi.org/10.1109/ICOIACT53268.2021.9564005
101 I. Sadek and B. Abdulrazak: Health Syst. **12** (2023) 264. https://doi.org/10.1080/20476965.2022.2072777
102 J. Nie, H. Shao, Y. Fan, Q. Shao, H. You, M. Preindl, and X. Jiang: arXiv (2022). https://arxiv.org/abs/2403.10779 (accessed May 2024).
103 M. Z. Uddin, W. Khaksar, and J. Torresen: Sensors **18** (2018) 2027. https://doi.org/10.3390/s18072027
104 D. Marín, J. Llano-Viles, Z. Haddi, A. Perera-Lluna, and J. Fonollosa: Sens. Actuators, B **393** (2023) 134036. https://doi.org/10.1016/j.snb.2023.134036
105 S. Sharma and A. Ghose: Proc. IEEE Int. Conf. Pervasive Computing and Communications Workshops (IEEE, 2018) 179–184. https://doi.org/10.1109/PERCOMW.2018.8480148
106 Q. Zhang, M. Varnfield, L. Higgins, V. Smallbon, J. Bomke, J. O'Dwyer, J. M. Byrnes, M. Sum, J. Hewitt, W. Lu, and M. Karunanithi: JMIR Res. Protoc. **11** (2022) e31970. https://doi.org/10.2196/31970
107 P. Elena-Anca, B. Ovidiu, P. Cristian, and M. Ianculescu: Proc. 23rd Int. Conf. Control Systems and Computer Science (IEEE, 2021) 495–500. https://doi.org/10.1109/CSCS52396.2021.00087
108 I. Susnea, E. Pecheanu, C. Sandu, and A. Cocu: Appl. Sci. **12** (2021) 235. https://doi.org/10.3390/app12010235
109 H. C. Chou: J. Intell. Fuzzy Syst. **36** (2019) 1323. https://doi.org/10.3233/JIFS-169903
110 H. X. Tan and H. P. Tan: Proc. IEEE World Forum on Internet of Things Conf. (IEEE, 2018) 207–212. https://doi.org/10.1109/WF-IoT.2018.8355204
111 A. Ghorayeb, R. Comber, and R. Gooberman-Hill: Int. J. Hum. Comput. Stud. **147** (2021) 102571. https://doi.org/10.1016/j.ijhcs.2020.102571
112 W. Guo, X. Liu, C. Lu, and L. Jing: Electronics. **13** (2024) 1066. https://doi.org/10.3390/electronics13061066
113 M. K. Batra, T. Chaspari, and R. C. Ahn: Proc. 44th Annual Int. Conf. the IEEE Engineering in Medicine & Biology Society (IEEE, 2022) 2839–2843. https://doi.org/10.1109/EMBC48229.2022.9871436
114 T. H. Tan, M. Gochoo, F. R. Jean, S. C. Huang, and S. Y. Kuo: IEEE Access **5** (2017) 10734. https://doi.org/10.1109/ACCESS.2017.2711495
115 A. Borhani, M. Pätzold, and K. Yang: Sensors **19** (2019) 3557. https://doi.org/10.3390/s19163557

116 M. Bennasar, B. A. Price, A. Stuart, D. Gooch, C. McCormick, V. Mehta, and B. Nuseibeh: Proc. Innovation in Medicine and Healthcare Systems, and Multimedia Conf. (Springer, 2019) 590–599. https://doi.org/10.1016/j.procs.2019.09.21
117 X. Ouyang and M. Srivastava: Proc. IEEE 3rd Workshop on Machine Learning on Edge Sensor Systems (IEEE, 2024) 9–14. https://doi.org/10.1109/SenSys-ML62579.2024.00007
118 X. Chen, J. Cumin, F. Ramparany, and D. Vaufreydaz: arXiv (2024). https://arxiv.org/abs/2407.09529 (accessed May 2024).
119 M. Buzzelli, A. Albé, and G. Ciocca: Appl. Sci. **10** (2020) 374. https://doi.org/10.3390/app10010374
120 Y. Htet, T. T. Zin, P. Tin, H. Tamura, K. Kondo, and E. Chosa: Int. J. Environ. Res. Public Health **19** (2022) 12055. https://doi.org/10.3390/ijerph191912055
121 G. Albani, C. Ferraris, R. Nerino, A. Chimienti, G. Pettiti, F. Parisi, G. Ferrari, N. Cau, V. Cimolin, C. Azzaro, L. Priano, and A. Mauro: Sensors **19** (2019) 4764. https://doi.org/10.3390/s19214764
122 J. Kondragunta and G. Hirtz: Proc. 42nd Annual Int. Conf. the IEEE Engineering in Medicine and Biology Society (IEEE, 2020) 5798–5801. https://doi.org/10.1109/EMBC44109.2020.9175766
123 A. Agarwal, B. Wallace, R. Goubran, F. Knoefel, and N. Thomas: Proc. IEEE Sensors Applications Symposium Conf. (IEEE, 2022). https://doi.org/10.1109/SAS54819.2022.9881252
124 S. Payandeh and J. Park: Int. J. Telemed. Appl. **2020** (2020) 8867926. https://doi.org/10.1155/2020/8867926
125 B. X. B. Yu and K. C. C. Chan: Proc. IEEE Int. Conf. Big Knowledge (IEEE, 2017) 284–289. https://doi.org/10.1109/ICBK.2017.18
126 D. Fuentes, L. Correia, N. Costa, A. Reis, J. Ribeiro, C. Rabadão, J. Barroso, and A. Pereira: Sensors **21** (2021) 6051. https://doi.org/10.3390/s21186051
127 Y. Sun and J. Ortiz: arXiv (2024). https://arxiv.org/abs/2407.02606 (accessed June 2024).
128 S. Sreeram, T.-H. Wang, A. Maalouf, G. Rosman, S. Karaman, and D. Rus: arXiv (2024). https://arxiv.org/abs/2405.05956 (accessed June 2024).
129 A. Sanchez-Comas, K. Synnes, D. Molina-Estren, A. Troncoso-Palacio, and Z. Comas-González: Sensors **21** (2021) 4210. https://doi.org/10.3390/s21124210
130 M. Hamim, S. Paul, S. I. Hoque, M. N. Rahman, and I. Al Baqee: Proc. Int. Conf. Robot, Electrical and Signal Processing Techniques (IEEE, 2019) 533–538. https://doi.org/10.1109/ICREST.2019.8644514
131 P. K. Chattopadhyay, A. J. Bond, and M. H. Lader: J. Psychiatr. Res. **12** (1975) 265–270. https://doi.org/10.1016/0022-3956(75)90005-9
132 R. R. Singh, S. Conjeti, and R. Banerjee: Biomed. Signal Process. Control **8** (2013) 740–754. https://doi.org/10.1016/J.BSPC.2013.06.014
133 P. P. Lakshmi, H. Nair, N. Pravin, and A. Sreedher: Proc. 2nd IEEE Int. Conf. Advance Technologies in Intelligent Control, Environment, Computing & Communication Engineering (IEEE, 2022) 1–5. https://doi.org/10.1109/ICATIECE56365.2022.10047781
134 A. Bhide, A. Ganguly, T. Parupudi, M. Ramasamy, S. Muthukumar, and S. Prasad: ACS Omega **6** (2021) 6031. https://doi.org/10.1021/acsomega.0c06209
135 C. D. Flynn, D. Chang, A. Mahmud, H. Yousefi, J. Das, and K. T. Riordan: Nat. Rev. Bioeng. **1** (2023) 560. https://doi.org/10.1038/s44222-023-00067-z
136 F. He, K. Li, X. Lv, Q. Zeng, Y. Zhu, X. Li, and Y. Deng: Microchim. Acta **189** (2022) 3. https://doi.org/10.1007/s00604-022-05469-1
137 M. N. Munshi: Diabetes Technol. Ther. **25** (2023) S56. https://doi.org/10.1089/dia.2023.0111
138 S. Chen, M. Guevara, S. Moningi, F. Hoebers, H. Elhalawani, B. H. Kann, F. E. Chipidza, J. Leeman, H. J. W. L. Aerts, T. Miller, G. K. Savova, J. Gallifant, L. A. Celi, R. H. Mak, M. Lustberg, M. Afshar, and D. S. Bitterman: Lancet Digit. Health **6** (2024) e379. https://doi.org/10.1016/S2589-7500(24)00060-8
139 C. Martinez-Cruz, J. F. G. Guerrero, J. L. L. Ruiz, A. J. Rueda, and M. Espinilla: Proc. of the 15th Int. Conf. Ubiquitous Computing & Ambient Intelligence (Springer, 2023) 842. https://doi.org/10.1007/978-3-031-48642-5_4
140 A. K. Dey: Int. J. Diabetes Technol. **2** (2023) 66. https://doi.org/10.4103/ijdt.ijdt_31_23
141 S. C. Chavan and A. Chavan: Proc. Int. Conf. Intelligent Computing and Control Systems (IEEE, 2017) 1135–1140. https://doi.org/10.1109/ICCONS.2017.8250644
142 H. Y. Teh, A. W. Kempa-Liehr, and K. I. K. Wang: J. Big Data **7** (2020) 1. https://doi.org/10.1186/s40537-020-0285-1
143 H. Zhang, J. Liu, and A. C. Pang: Comput. Netw. **143** (2018) 166. https://doi.org/10.1016/j.comnet.2018.07.009
144 E. W. Dereszynski and T. G. Dietterich: ACM Trans. Sens. Netw. **8** (2011) 1. https://doi.org/10.1145/1993042.1993045

145 H. Liu, J. Chen, F. Huang, and H. Li: Proc. 14th Int. Symp. Pervasive Systems, Algorithms and Networks (2017) 430–435. https://doi.org/10.1109/ISPAN-FCST-ISCC.2017.29
146 Y. Li and L. E. Parker: Inf. Fusion **15** (2014) 64. https://doi.org/10.1016/j.inffus.2012.08.007
147 G. D'Aniello, M. Gaeta, and T. P. Hong: IEEE Trans. Emerg. Top. Comput. Intell. **2** (2018) 65. https://doi.org/10.1109/TETCI.2017.2782800
148 J. Tang, G. Zhang, Y. Wang, H. Wang, and F. Liu: Transp. Res. Part C Emerging Technol. **51** (2015) 29. https://doi.org/10.1016/j.trc.2014.11.003
149 P. Xu, W. Ruan, Q. Z. Sheng, T. Gu, and L. Yao: Proc. the 14th EAI Int. Conf. Mobile and Ubiquitous Systems: Computing, Networking and Services (2017) 442–451. https://doi.org/10.1145/3144457.3144474
150 J. Rabatel, S. Bringay, and P. Poncelet: Expert Syst. Appl. **38** (2011) 7003. https://doi.org/10.1016/j.eswa.2010.12.014
151 K. H. Eom, S. J. Lee, Y. S. Kyung, C. W. Lee, M. C. Kim, and K. K. Jung: Sensors **11** (2011) 10266. https://doi.org/10.3390/s111110266
152 R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, and B. Qureshi: Sensors **20** (2020) 6076. https://doi.org/10.3390/s20216076
153 A. de Cheveigné and J. Z. Simon: J. Neurosci. Methods **168** (2008) 195. https://doi.org/10.1016/j.jneumeth.2007.09.012
154 R. Dunia, S. J. Qin, T. F. Edgar, and T. J. McAvoy: Comput. Chem. Eng. **20** (1996) S713. https://doi.org/10.1016/0098-1354(96)00128-7
155 S. Cai, M. Shao, M. Du, G. Bao, and B. Fan: IEEE Sens. J. **23** (2023) 2663. https://doi.org/10.1016/j.jbiomech.2022.111358
156 M. S. Tasjid and A. A. Marouf: Int. J. Interact. Mob. Technol. **15** (2021) 167. https://doi.org/10.3991/ijim.v15i24.25891
157 P. Albuquerque, J. P. Machado, T. T. Verlekar, P. L. Correia, and L. D. Soares: Diagnostics **11** (2021) 1824. https://doi.org/10.3390/diagnostics11101824
158 N. Yala, B. Fergani, and A. Fleury: Proc. Int. Symp. Innovations in Intelligent Systems and Applications (IEEE, 2015). https://doi.org/10.1109/INISTA.2015.7276759.
159 T. Okita, K. Ukita, K. Matsuishi, M. Kagiyama, K. Hirata, and A. Miyazaki: Proc. Int. Conf. Pervasive and Ubiquitous Computing (2023) 499–504. https://doi.org/10.1145/3594739.3610745
160 Y. Tong, J. Bai, and X. Chen: J. Phys. Conf. Series. **1624** (2020) 032046. https://doi.org/10.1088/1742-6596/1624/3/032046
161 A. De Paola, S. Gaglio, G. Lo Re, and M. Ortolani: Proc. Artificial Intelligence Around Man and Beyond (Springer, 2011) 6934. https://doi.org/10.1007/978-3-642-23954-0_33
162 A. De Paola, P. Ferraro, S. Gaglio, G. Lo Re, and S. K. Das: IEEE Trans. Mob. Comput. **16** (2017) 1502. https://doi.org/10.1109/TMC.2016.2599158
163 C. Scholl, M. Spiegler, K. Ludwig, B. M. Eskofier, A. Tobola, and D. Zanca: Sensors **23** (2023) 83798. https://doi.org/10.3390/s23083798
164 H. Xu, L. Han, Q. Yang, M. Li, and M. Srivastava: Proc. of the 25th Int. Workshop on Mobile Computing Systems and Applications (2024) 1–7. https://doi.org/10.1145/3638550.3641130.
165 F. Cruciani, I. Cleland, C. Nugent, P. McCullagh, K. Synnes, and J. Hallberg: Sensors **18** (2018) 2203. https://doi.org/10.3390/S18072203.
166 A. Diete, T. Sztyler, and H. Stuckenschmidt: Proc. IEEE Int. Conf. Pervasive Computing and Communications Workshops (IEEE, 2017) 111–116. https://doi.org/10.1109/PERCOMW.2017.7917542
167 N. Pius Owoh, M. Mahinderjit Singh, and Z. F. Zaaba: Sensors **18** (2018) 2134. https://doi.org/10.3390/s18072134
168 C. F. Martindale, F. Hoenig, C. Strohrmann, and B. M. Eskofier: Sensors **17** (2017) 2328. https://doi.org/10.3390/s17102328
169 N. Gruver, M. Finzi, S. Qiu, and A. G. Wilson: Adv. Neural Inf. Process. Syst. **36** (2023) 19622.
170 Y. Kim, X. Xu, D. McDuff, C. Breazeal, and H. W. Park: arXiv (2024). https://arxiv.org/abs/2401.06866 (accessed May 2024).
171 N. Rajkumar, R. Li, and D. Bahdanau: arXiv (2022). https://arxiv.org/abs/2204.00498 (accessed May 2024).
172 C. Jeong: arXiv (2024). https://arxiv.org/abs/2401.02981 (accessed May 2024).
173 T. Han, L. C. Adams, J.-M. Papaioannou, P. Grundmann, T. Oberhauser, A. Löser, D. Truhn, and K. K. Bressem: arXiv (2023). https://arxiv.org/abs/2304.08247 (accessed May 2024).
174 A. Toma, P. R. Lawler, J. Ba, R. G. Krishnan, B. B. Rubin, and B. Wang: arXiv (2023). https://arxiv.org/abs/2305.12031 (access June 2024).
175 W. Alshikh and K. Kamble: Palmyra-Med (2023) 1. https://doi.org/10.13140/RG.2.2.30939.75046.

176 C. Wu, W. Lin, X. Zhang, Y. Zhang, W. Xie, and Y. Wang: J. Am. Med. Inform. Assoc. **31** (2024) 1833. https://doi.org/10.1093/jamia/ocae045
177 SuperAnnotate: https://www.superannotate.com/blog/llm-fine-tuning (accessed April 29, 2024).
178 K. Lv, Y. Yang, T. Liu, Q. Gao, Q. Guo, and X. Qiu: arXiv (2023). https://arxiv.org/abs/2306.09782 (accessed April 2024).
179 C. Christophe, P. K. Kanithi, P. Munjal, T. Raha, N. Hayat, R. Rajan, A. Al-Mahrooqi, and A. Gupta: arXiv (2024). https://arxiv.org/abs/2404.14779 (accessed April 2024).
180 J. Chen, H. Lin, X. Han, and L. Sun: Proc. AAAI Conf. Artificial Intelligence **38** (2024) 17754. https://doi.org/10.1609/aaai.v38i16.29728
181 T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei: Proc. Advances in neural information processing systems. (2020) 1877–1901.
182 V. Thambawita, S. A. Hicks, H. Borgli, H. K. Stensland, D. Jha, M. K. Svensen, S. A. Pettersen, P. Schmidt, M. Riegler, T. de Lange, and P. Halvorsen: Proc. the 11th ACM Multimedia Systems Conf. (AMC, 2020) 231– 236. https://doi.org/10.1145/3339825.3394926
183 Langchain: https://python.langchain.com/docs/community (accessed May 2024).
184 S. Valliappan, B. P. R. Mohan, and S. R. Kumar: Proc. Int. Conf. IoT Applications (IEEE, 2017) 1–7. https://doi.org/10.1109/ICIOTA.2017.8073612
185 A. S. Kumar, D. Gobinath, N. Nithiyanandam, P. Vijayakarthik, S. Dhanasekaran, and G. V. Rajkumar: Proc. Int. Conf. Inventive Computation Technology (IEEE, 2022) 1016. https://doi.org/10.1109/ICICT54344.2022.9850896
186 M. Toyama and H. R. Fuller: Innovation in Aging **1** (2019) S618. https://doi.org/10.1093/geroni/igz038.2305
187 X. Liu, D. McDuff, G. Kovacs, I. Galatzer-Levy, J. Sunshine, J. Zhan, M.-Z. Poh, S. Liao, P. Di Achille, and S. Patel: arXiv (2023). https://arxiv.org/abs/2305.15525 (accessed May 2024).
188 M. Balestra, O. Shaer, J. Okerlund, L. Westendorf, M. Ball, and O. Nov: J. Med. Internet Res. **18** (2016) e197. https://doi.org/10.2196/jmir.5662
189 J. Katusiime and N. Pinkwart: Health Inform. J. **25** (2019) 935. https://doi.org/10.1177/1460458217733121
190 M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, and Q. Wen: arXiv (2023). https://arxiv.org/abs/2310.01728 (accessed April 2024).
191 N. Wretblad and F. Gordh Riseby: Bridging Language & Data: Optimizing Text-to-SQL Generation in Large Language Models. https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-200605 (accessed July 2024).
192 A. Berenguer, A. Morejón, D. Tomás, and J. N. Mazón: Appl. Sci. **14** (2024) 2506. https://doi.org/10.3390/app14062506
193 S. Suh, V. F. Rey, and P. Lukowicz: Knowl. Based Syst. **260** (2023) 110143. https://doi.org/10.1016/j.knosys.2022.110143
194 H. Kaneko and S. Inoue: Proc. ACM Int. Joint Conf. Pervasive and Ubiquitous Computing & the 2023 ACM Int. Symposium on Wearable Computing (ACM, 2023) 475. https://doi.org/10.1145/3594739.3610741
195 M. Webber and R. F. Rojas: IEEE Sens. J. **21** (2021) 16979. https://doi.org/10.1109/JSEN.2021.3079883
196 P. Dongre: Proc. HCI Conf. Human Factors in Computing Systems (ACM, 2024) 1–5. https://doi.org/10.1145/3613905.3651132
197 A. Basit, K. Hussain, M. A. Hanif, and M. Shafique: arXiv (2024). https://arxiv.org/abs/2403.00830 (accessed May 2024).

About the Authors

Ziaullah Momand earned his B.Sc CS. degree from Kabul University in Afghanistan and his master's degree from King Mongkut's University of Technology Thonburi (KMUTT) in Thailand. He is currently pursuing his Ph.D. degree at KMUTT, where his research focuses on developing digital twin technology for elderly health monitoring.

Pornchai Mongkolnam received his Ph.D. degree in computer science from Arizona State University, with a focus on computer-aided geometric design. He is currently an associate professor with the School of Information Technology, King Mongkut's University of Technology Thonburi. His research interests include data science and machine learning as well as user-centered applied information technology.

Jonathan H. Chan is an associate professor of computer science at the School of Information Technology, King Mongkut's University of Technology Thonburi (KMUTT), Thailand. He holds B.A.Sc., M.A.Sc., and Ph.D. degrees from the University of Toronto and returned there as a visiting professor on several occasions. He is also a visiting scientist at the Centre for Applied Genomics at Sick Kids Hospital in Toronto and the director of the Innovative Cognitive Computing (IC2) Research Center at KMUTT. In addition, he is a founding member and a past chairperson of the IEEE-CIS Thailand Chapter, the immediate past president of the Asia Pacific Neural Network Society (APNNS), and a senior member of IEEE, ACM, INNS, and APNNS. His research interests include intelligent systems, cognitive computing, biomedical informatics, data science, and machine learning in general.

Nipon Charoenkitkarn received his B.Eng. degree in computer engineering from King Mongkut's Institute of Technology Ladkrabang, Thailand, in 1987, his M.Sc. degree in engineering management from California State University, Northridge, USA, in 1990, and his Ph.D. degree in information systems from the University of Toronto, Canada, in 1996. He is currently an associate professor with the School of Information Technology, King Mongkut's University of Technology Thonburi. His research interests include management information systems, decision support systems, and information retrieval.