The Role of Large Language Models in Addressing IoT Challenges: A Systematic Literature Review

Gabriele De Vito[1], Fabio Palomba[1], Filomena Ferrucci[1]

[a]*Department of Computer Science, University of Salerno, Via Giovanni Paolo II 132, 84084 Fisciano, Salerno, Italy*

Abstract

The Internet of Things (IoT) has revolutionized various sectors by enabling devices to communicate and interact seamlessly. However, developing IoT applications has data management, security, and interoperability challenges. Large Language Models (LLMs) have shown promise in addressing these challenges due to their advanced language processing capabilities. This Systematic Literature Review assesses the role of LLMs in addressing IoT challenges, exploring the strategies, hardware, and software configurations used, and identifying directions for future research. We extensively searched databases like Scopus, IEEE Xplore, and ACM Digital Library, initially screening 1,419 studies and identifying an additional 1,167 through snowballing, ultimately focusing on 55 relevant papers. The findings reveal LLMs' potential to address key IoT challenges such as security and scalability. However, they also highlight significant obstacles, including high computational demands and the complexities of training and tuning these models. Future research should aim to develop methods to reduce the computational requirements of LLMs, improve training datasets, simplify implementation processes, and explore the ethical and privacy implications of using LLMs in IoT applications.

Keywords: Internet-of-Things, Large Language Models, Systematic Literature Reviews

1. Introduction

Internet of Things (IoT) is a rapidly evolving technology that involves the interconnection of physical devices and objects to the Internet, enabling them to communicate and interact with each other [1]. This technology is transforming various sectors, including healthcare, transportation, manufacturing, and smart homes [1, 2, 3, 4], by embedding sensors, actuators, and connectivity capabilities into everyday objects. These "smart" devices can collect and exchange data, monitor their environment, and perform automated actions based on the data they receive [5, 6]. However, the use of IoT also presents several challenges, including ensuring the security and privacy of IoT devices and the data they generate [7, 8], managing and analyzing the vast amounts of data [9], and ensuring interoperability among different IoT devices and platforms [10, 11].

Large Language Models (LLMs) represent a type of artificial intelligence model that has revolutionized the field of Natural Language Processing (NLP) [12]. Models, such as GPT-3, GPT-4, and BERT [13, 14, 15] are characterized by their extensive parameter count [16, 17, 18] and are trained using vast datasets, enabling them to learn and generate language with remarkable proficiency [16, 17, 18]. LLMs have demonstrated exceptional performance in various language-related tasks, such as translation and summarization, and have been used across various disciplines [19, 20]. However, adopting LLMs can pose challenges due to issues related to data privacy and security, resource intensity, context size constraints, and training data required [14].

Given that, we conducted a systematic literature review with a dual purpose. First, we aim to provide a comprehensive overview of the current state of research at the intersection of IoT and LLMs, identifying the critical challenges in the IoT domain that have been addressed using LLMs. Second, we aim to identify the strategies used in LLM-based solutions to address IoT challenges, such as fine-tuning and data preprocessing, and explore the hardware and software configurations used in implementing these solutions. For this purpose, four different full-text and bibliographic databases were used: Scopus, IEEE Xplore, ScienceDirect, ACM Digital Library, and Springer. We identified 55 articles for our review, after a systematic analysis and selection, according to the well-established guidelines by Kitchenham et al. [21]. By synthesizing the findings from these studies, we hope to provide valuable insights for researchers and practitioners in the field and to stimulate further research in this promising area. Moreover, our review identifies gaps in the current literature, indicating where further research is needed.

Structure of the paper. Section 2 describes the background concerning IoT and LLMs and the related works. Section 3 presents the methodology used to conduct the SLR. Section 4 reports the analysis of the results. Section 5 discusses the results and their implications, and provides the future research lines. Section 6 reports the threats to the validity and the mitigation strategies applied. Finally, Section 7 concludes the paper.

Email addresses: gadevito@unisa.it (Gabriele De Vito), fpalomba@unisa.it (Fabio Palomba), fferrucci@unisa.it (Filomena Ferrucci)

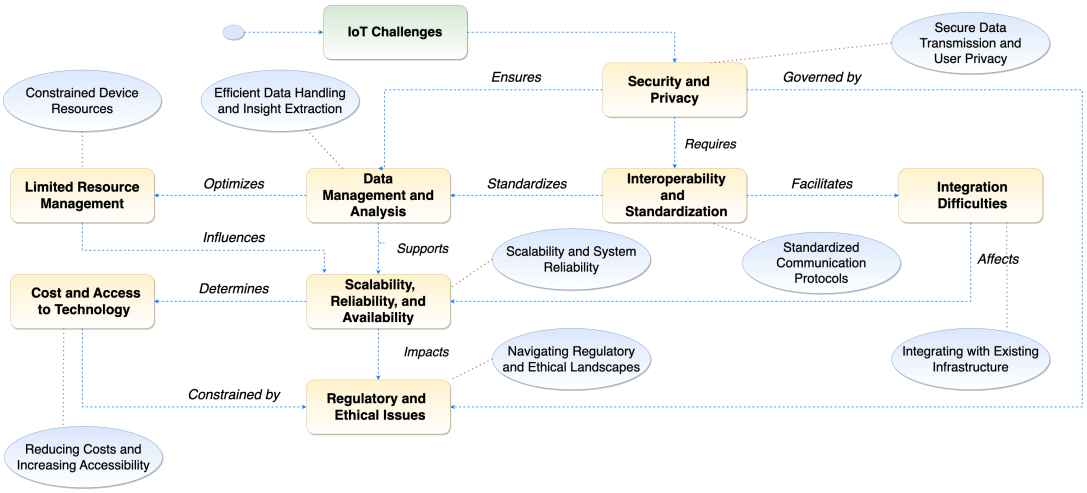


Figure 1: IoT challenges and open issues. Yellow boxes represent challenges, while Blue boxes depict open issues

2. Background and Related Work

This section overviews the two main themes of our work, namely (i) IoT and its challenges in different domains; and (ii) LLMs, their capabilities, and adoption problems.

2.1. Internet of Things

IoT connects physical devices to the Internet, enabling communication and interaction [1]. These devices can collect and exchange data, monitor environments, and perform automated actions, supporting applications like remote monitoring and intelligent automation [5, 6]. IoT's exponential growth has led to billions of connected devices, enhancing efficiency and decision-making [22]. It has revolutionized various domains, including healthcare, transportation, manufacturing, and smart homes [1, 2, 3, 4]. Lohiya et al. [23] categorize the domains of IoT applications into Healthcare, Smart Grid, Transportation, Smart Home and Building, Smart Cities, Agriculture, Industry, Military, and Others (i.e., Social, Education, and Logistics). However, the use of IoT also presents several challenges and issues that need to be addressed.

A few literature reviews and surveys report IoT challenges and open issues (see Table 1).

Yang et al. [24] discuss IoT-based healthcare challenges like high power consumption, limited resources, and security issues, suggesting machine-learning (ML) algorithms and cloud integration as potential solutions. Security and privacy issues are also examined by Conoscenti et al. [25], Asghari et al. [26], and Giordano et al. [27]. Conoscenti et al. propose Blockchain for privacy, while Asghari et al. advocate for interoperable systems and energy-efficient solutions. Giordano et al. suggest AI-enhanced methods for security. Syed et al. [28] overview IoT in Smart Cities, recommending secure, interoperable systems. Goudarzi et al. [29] analyze IoT in Smart Grids, suggesting ML and Blockchain for resilience. Sinha et al. [30] explore IoT in Smart Agriculture, recommending cost reduction and improved security technologies. Stoyanova et al. [31] discuss IoT challenges in digital forensics, emphasizing standardization. Malekshahi Rad et al. [32] review the Social IoT ecosystem, highlighting the need for efficient service discovery. Ullo et al. [33] review smart environment monitoring, recommending robust ML methods. Song et al. [34] highlight IoT opportunities in logistics, suggesting Blockchain and AI integration. Oladimeji et al. [35] discuss IoT in smart transportation, recommending ML for predictive models. Almusaylim et al. [36] review smart home automation, calling for integration into smart city designs. Khan et al. [37] discuss Industrial IoT, highlighting data privacy and cloud strategies. Nizetic et al. [38] address IoT challenges in energy and smart cities, proposing energy efficiency solutions. Babun et al. [7] survey IoT evolution, identifying scalability and security challenges. Compare et al. [5] discuss predictive maintenance in Industry 4.0, calling for research on cost models. Sicari et al. [8] review 5G-enabled IoT networks, highlighting AI's role. Karale et al. [39] examine IoT challenges, proposing future research on ethics and regulatory aspects. From the analyzed work, several patterns emerge regarding challenges and open issues in the field of IoT (see Figure 1).

Table 1: IoT Challenges emerging from the related work.

IoT Challenges	Studies
Security and Privacy	[28, 29, 24, 30, 31, 35, 36, 37, 32, 33, 34, 38, 7, 5, 8, 39, 27]
Interoperability and Standardization	[28, 35, 36, 37, 32, 38, 7, 8, 39]
Integration Difficulties	[36, 37, 32, 39]
Data Management and analysis	[35, 37, 33, 34, 38, 7, 5, 8]
Limited Resource Management	[24, 30, 36, 37, 32, 34, 38, 39]
Scalability, Reliability and Availability	[5, 7, 8, 39]
Cost and Access to Technology	[30, 37, 38]
Regulatory and Ethical Issues	[39]

2.2. Large Language Models

LLMs [12] are language models that, leveraging the Transformer architecture with self-attention mechanisms introduced by Vaswani et al. [40], have revolutionized the NLP field. Notable examples include paid models like GPT-3 [13] and GPT-

4 [14], as well as open-source models like BERT [15], FLAN-T5 [41], LLama [42], BLOOM [43], and GLM [44]. These models are trained on extensive text datasets and often have hundreds of billions of parameters [16, 17, 18]. The initial "pre-training" phase is computationally intensive, but it is required to allow LLMs to perform NLP tasks like translation and summarization [16, 17, 18] with high capability. LLMs can then be specialized through a fine-tuning process, using smaller datasets and tailoring them to execute specific NLP tasks (i.e., question-answering) or the same tasks but in different domains. Several emergent abilities have been discovered in the context of LLMs. LLMs' most common critical abilities are "In-context learning", "Instruction following", and "Step-by-step reasoning". "In-context learning," introduced by GPT-3, allows models to perform tasks based on examples without additional training. "Instruction following" enables task execution from instructions alone, and "Step-by-step reasoning" helps solve complex problems through chain-of-thought prompting.

3. Research Method

This section illustrates the review protocol we employed for the SLR. We adopted the well-established approach proposed by Kitchenham et al. [21]. Moreover, we followed the *"General Standard"* and *"Systematic Reviews"* guidelines provided by the ACM/SIGSOFT Empirical Standards[1] when organizing and reporting the results. The protocol comprises three main phases: Planning the Review, Conducting the Review, and Reporting the Review. The subsequent subsections offer a comprehensive explanation of each step in our SLR.

3.1. Review Planning

In the Review Planning phase, we established the objectives of our research and outlined the methodology for gathering and assessing scholarly literature from the search databases. The Review Planning phase comprises three sub-phases: 1) the Initiation Phase, 2) the Search Phase, and 3) the Eligibility Criteria Phase.

3.1.1. Initiation Phase

The main objective of our work is to analyze and synthesize existing literature to understand the role of LLMs in providing solutions to the prevalent challenges in the IoT domain and how LLMs have been applied to deliver solutions. This consideration gave rise to the first research question:

> **RQ_1—What specific IoT challenges have been addressed using LLMs, and how effective are these solutions in improving IoT systems?**

The IoT landscape presents numerous challenges, such as data security, management, integration, and scalability (see Section 2 for details). Innovative solutions are needed to address these issues. LLMs have shown significant potential in various domains (see Section 2 for details) and could offer new solutions to IoT challenges. For example, LLMs can analyze extensive, unstructured textual data from IoT devices to improve access control and threat detection. In healthcare, LLMs can parse and classify extensive data to monitor health, predict risks, and provide timely intervention alerts, enhancing patient outcomes and reducing costs. So, it is worth identifying the specific challenges within IoT that LLMs have targeted and evaluating the effectiveness of these solutions in enhancing IoT systems. After identifying the solutions, the second question investigates the methodologies of their implementation:

> **RQ_2—What methodologies and optimization techniques have been employed to implement LLM-based solutions for IoT applications, and what are the associated technical requirements and constraints?**

Understanding the implementation of LLM-based solutions is crucial. This research question explores the methodologies and optimization techniques for deploying LLM-based solutions in IoT applications. It focuses on the technical and methodological aspects, including the necessary computational resources, software frameworks, optimization strategies, and the constraints and new challenges introduced by LLMs.

3.1.2. Search Phase

The Search Phase comprises two steps: 1) Data Sources Selection, and 2) Search String Definition.

The Data Sources Selection step aims to identify the most reliable databases to extract the literature for starting our process. This step is crucial for the literature review's success [21]. To collect the studies to review, we selected the following data sources: Scopus, IEEE Xplore, ScienceDirect, ACM Digital Library, Springer. In the Search String Design step, we have included search terms by employing alternative terms and synonyms of related terms using the Boolean operator OR and combining the main terms via the Boolean operator AND.

The final search string is shown in the box below.

Search string

("LLM?" OR "Large Language Model?" OR chatgpt OR gpt OR bert) AND (?iot OR "internet of things")*

Notably, we did not incorporate critical terms from our research questions, such as "challenges," as our objective was to gather as much studies as possible, even if this phase required significant effort. We utilized the references of the identified papers and downloaded additional studies. Consequently, we initiated a formal search using the defined keywords, followed by a manual search in the references of the initial pool within our field of study.

3.1.3. Eligibility Criteria Definition

In this phase, we define criteria to filter and identify relevant literature based on our research questions. The eligibility criteria are divided into three sets for each paper:

[1] ACM/SIGSOFT Empirical Standards: `https://github.com/acmsigsoft/EmpiricalStandards`

- Exclusion Criteria: If any are met, the article is eliminated.
- Inclusion Criteria: If all are met, the article is included.
- Quality Criteria: Used to grade articles; those below a certain threshold are eliminated.

Tables 2 and 3 list the exclusion and inclusion criteria.

Table 2: Inclusion criteria

Code	Name	Description
IC1	Focus on IoT Challenges and LLM's Solutions	The paper should specifically focus on identifying the challenges associated with IoT, and explicitly discuss how LLMs can be identified as viable solutions. The criterion ensures relevance to our primary research objective of studying LLM applications in IoT.
IC2	Practical Application	The paper provides cases or examples of the practical application of LLMs to solve IoT issues. The criterion validates that the research is not purely theoretical.
IC3	Implementation Details	The paper must provide explicit strategies, methods, or steps on the implementation or integration of LLM solutions to the IoT ecosystem. We aim to gather actionable insights from the studies.
IC4	Evidence of Effectiveness	The paper provides empirical evidence (qualitative or quantitative) to support the effectiveness of the proposed LLM solution. Evidence-backed research is essential for proving the feasibility of the studied solutions.

Table 3: Exclusion criteria

Code	Name	Description
EC1	Off-topic Papers	Papers not primarily focused on using LLMs in the IoT ecosystem. For example, papers that just mention these terms in passing or as not central to their study should be excluded.
EC2	Document Type	Conference proceedings summary, book chapters, books, and other forms of documents leaving only peer-reviewed journal articles and conference papers must beexpluded.
EC3	Access Type	Articles that are not freely accessible or do not have a full-text available online.
EC4	Duplicated articles	The same articles that are in more than one digital library must be considered once.
EC5	Articles written before 2017	Articles that has been written before 2017 must be excluded.

Table 4: Quality criteria

Cod	Name	Question	Rationale
QC1	Adequate Methodology	Is the methodology comprehensively described, including the specific LLMs used, their application to IoT challenges, and how data was collected and analyzed?	Ensures the study is detailed enough for reproducibility and clearly explains the application of LLMs to IoT challenges
QC2	Clear Objectives and Research Questions	Are the research objectives and research questions clearly stated that align with investigating the role of large language models in solving IoT challenges?	Ensures the study directly addresses our research questions and objectives on LLMs in IoT.
QC3	Relevant Results	Are the findings of the paper directly relevant to the research topic? The results should either display the advantages of using large language models in the IoT field or show areas that need to be further improved.	Assessing the practical impact and relevance of the findings ensures valuable insights into the efficacy and improvement areas of LLMs in IoT.
QC4	Limitations and threats to validity	Are the limitations of the study analyzed explicitly?	Analyzing study limitations enhances reliability, contextual applicability, and identifies areas needing caution or further research.

We defined criteria based on guidelines [21] and previous literature reviews in IoT and LLM fields [25, 45, 26, 46]. Each criterion was binary (True/False). These filters excluded preliminary research (e.g., workshops, posters) and duplicates. EC1 ensured relevance to our topic. Inclusion criteria were aligned with our objectives and research questions, similar to other SLRs [25, 26, 46]. We used EC5—Articles written before 2017—to account for significant NLP advancements post-2017 due to transformers and attention mechanisms by Vaswani et al. [40] (see Section 2). Other exclusion criteria were standard in SLRs and applied via database filters. For quality, we evaluated each study's reliability and relevance, as suggested by Higgins et al. [47], using questions listed in Table 4.

The questions can be answered as "Yes," "Partially," or "No." Each answer corresponded to a numerical value, i.e., "1.5", "1.0", and "0.5". The sum of these values reflects the quality score of the article. We excluded the score from the review if it was lower than 3.5.

3.2. Review Conducting

The Review Conducting phase pertains to the compilation of the literature set, which is designed based on the planning from the preceding phase, to address the research questions. It comprises three sub-phases: 1) Study Selection, 2) Data Extraction, and 3) Data Synthesis and Analysis.

3.2.1. Study Selection Phase

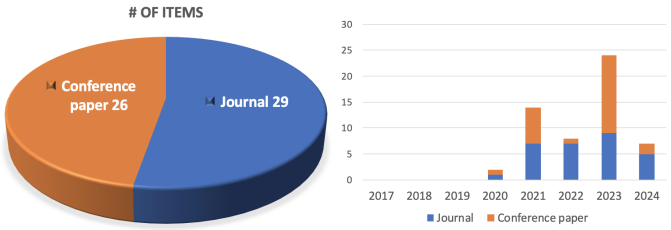
The Selection Phase aimed to gather relevant literature using a planned strategy [21]. Initially, we searched the identified databases, retrieving 1,419 articles. Applying Exclusion and Inclusion criteria reduced the articles to 55. Quality criteria further refined it to 50. We reviewed references from identified studies to address potential missing studies, adding 1,167 more. After applying the same criteria, we included five additional studies, resulting in a final set of 55 articles for data extraction. The resulting dataset includes 29 journal articles and 26 conference papers (see Figure 2a). Most papers were published between 2021 and 2024 (see Figure 2b), reflecting recent growing interest in the subject. The studies span 48 distinct venues, suggesting that a primary venue dedicated to the amalgamation of IoT and LLMs is yet to emerge, possibly due to the novelty of this subject within the scientific community.

3.2.2. Data Extraction Phase

We aligned our objectives and research questions to create data extraction forms during data extraction. These forms, developed collaboratively, were implemented by the first author. We used a pre-defined form to extract relevant data from each source, including general information like authors' names, publication years, and keywords, as per literature review guidelines [21]. Table 5 shows the extracted information. We conducted a preliminary extraction on 15 randomly selected sources to ensure validity and reliability. All extracted data were compiled into a spreadsheet for analysis, available in our online appendix [48]. We ensured traceability between forms and research questions, as recommended by SLR guidelines [21]. All papers address RQ_1, while [SLR 6, 28, 37, 47, 52] do not address RQ_2.

3.2.3. Data Synthesis and Analysis Phase

During the data synthesis phase, we combined the collected data to make it comprehensible and applicable to the intended



(a) Number of papers for source type. (b) Number of papers per year.

Figure 2: Statistics on the final dataset composed of 55 papers.

Table 5: The information extracted from articles.

Code	Information	Description
M1	Authors	Authors of the article
M2	Title	Title of the article
M3	Year	Year of publication of the article
M4	Source	Source of the article
M5	Source type	Source type of the article (i.e., Journal)
M6	Publisher	Publisher of the article
M7	Citations	Number of citations of the article
M8	Abstract	Abstract of the article
M9	Keywords	Keywords used by the authors to describe the topics of the article

audience. The first two authors analyzed and segregated the data into various form groups designed to extract specific information for each research question. This collaborative approach ensured that no crucial information was overlooked or misinterpreted [21]. We used two qualitative analysis methods to ensure accurate data synthesis: narrative synthesis and thematic analysis. Narrative synthesis involves describing and interpreting primary evidence, while thematic analysis summarizes studies based on recurring themes. Initially, we used narrative synthesis for preliminary analysis, followed by thematic analysis to classify the data. The first two authors jointly executed these steps to identify potential discussion points for a subset of items. Upon completion, we derived new insights under ten major categories, five for each research question. All findings are documented in the online appendix [48].

3.3. Result Reporting

In the final phase, we compiled and presented findings to ensure clear insight and comprehension. We classified results by research question for coherence, with each finding substantiated by references from our SLR. Visual aids, such as tables, charts, and diagrams, were employed to summarize results and highlight significant points. In Section 5, we elaborated on the results, discussing their implications and highlighting future research opportunities to make our review valuable for researchers and practitioners in the IoT and LLM fields and foster academic progress.

4. Analysis of the Results

This section presents the findings from our review. For readability, we provide only few examples. For detailed information and data, the reader can refer to the online appendix [48]. Specifically, the *Data_Extraction* file includes information from Table 6 for each article, and the *Additional_result_details* file offers extra results and tables for each category and subcategory.

Table 6: Data extracted from the analyzed studies.

Information	Description
IoT field	IoT field or domain of the article
IoT Challenges	IoT challenges addressed by the LLM-solution
Methodology	Research methodology used
Assessment	Evaluation approach and achieved results
LLM used	LLM used for the proposed solution
SW configuration	Software Configurations used to develop the LLM solution
HW Configuration	Hardware Configurations (Type, CPU, RAM, GPU, etc.)
Pre-training and fine-tuning	Pre-training and Fine-tuning approaches for LLMs
Other ML techniques	Machine learning techniques integrated with LLMs
Optimization techniques	Optimization techniques used to enhance LLMs performance
Extraction techniques	Extraction techniques used in the LLM-based solution to process and analyze the IoT data
Evaluation techniques	Evaluation metrics for assessing LLM-based solutions
Continuous monitor	Strategies used to continuously monitor and adjust the performance of LLM-based solutions to improve their effectiveness
Data Collection and preprocessing	Data Collection and Pre-processing approaches used for the LLM-based solution
LLM issues	Limitations and issues introduced by LLMs
Future direction	Recommendations or future direction suggested

*4.1. **RQ$_1$**—What specific IoT challenges have been addressed using LLMs, and how effective are these solutions in improving IoT systems?*

The first research question aims to identify the solutions proposed to address existing challenges in IoT by using LLMs. It explores how LLMs have been applied in various IoT domains and their effectiveness. We have organized the solutions into primary categories, each with sub-categories for readability (see Section 3 for category breakdown details). Figure 3 illustrates the distribution of the studies among the categories.

4.1.1. Security and Privacy

The "Security and Privacy" category, representing 36.4% of the dataset, involves studies using LLMs to improve IoT security and privacy. These papers focus on applying LLMs to detect and mitigate cyber threats like intrusion and malware detection, and identifying privacy-violating rules in IoT platforms.

Cyber Threat Detection and Access Control. In the Security context, 27.3% of studies have exploited LLMs to enhance cyber threat detection and access control in IoT environments. A

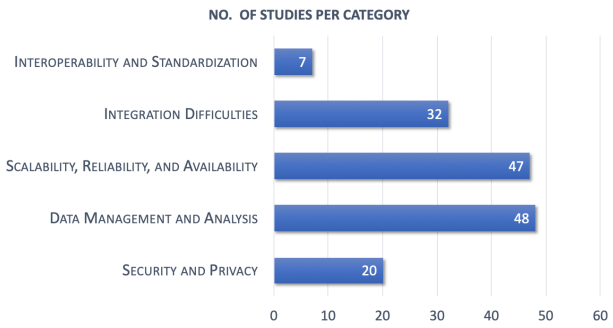


Figure 3: RQ1: Categories distribution of the studies.

practical application is the dynamic access control model for IoT-based smart grids in [SLR 2]. Using BERT for feature extraction from log access records, it achieved 87.73% accuracy and 92.27% recall on a provincial power grid system, demonstrating its effectiveness in monitoring user access, dynamically adjusting rights, and managing policies in real-time. Another application is reported in [SLR 8], where BERT was employed to detect Advanced Persistent Threats (APTs) in Industrial IoT (IIoT). By fine-tuning the pre-trained BERT model for identifying sequences indicative of APT attacks, the study achieved over 99% accuracy across various sequence lengths, underscoring the model's precision and adaptability.

Data Privacy. Despite the limited number of studies (only 12.7% of the research), LLMs have shown promising results in improving data privacy in IoT applications, particularly in the healthcare, industrial, and smart home sectors. The SecurityBERT model [SLR 38] exemplifies LLMs' role in this context by efficiently and accurately identifying real-time network-based attacks on IoT devices. It uses a Privacy-Preserving Fixed-Length Encoding technique and a Byte-level Byte-Pair Encoder tokenizer to handle network traffic data. With 98.2% accuracy in detecting 14 attack types, SecurityBERT outperforms traditional methods. Its inference time is under 0.15 seconds on an average CPU, and its size is 16.7MB, making it ideal for real-time traffic analysis on resource-limited IoT devices.

Security in Communication. Only 10.9% of the analyzed studies used LLMs in communication security. Nevertheless, LLM applications spanning diverse IoT sectors, including IIoT, transportation and smart home systems, have tackled a plethora of security concerns. For instance, [55, SLR] introduces "CANBERT," a BERT-based intrusion detection system for securing the Controller Area Network (CAN) in modern vehicles. CANBERT uses a self-supervised "masked language model" training approach to predict masked CAN IDs, effectively identifying unauthorized message injections. Experimental results show superior performance, surpassing models like Isolation Forest, with an F1-score of 0.81 to 0.99. Its high accuracy and real-time detection (0.8 to 3 ms inference times) underscore its potential to enhance automotive IoT security.

Security and Privacy: Summing Up

LLMs enhance IoT security and privacy by addressing challenges like cyber threat detection, data privacy, and communication security and enhancing IoT systems' accuracy, recall, robustness, precision, and real-time response.

4.1.2. Data Management and Analysis

LLMs have been extensively used to improve the management, processing, and analysis of IoT-generated data, representing 83.6% of the dataset. These studies address large-scale, heterogeneous, and real-time data stream challenges, demonstrating LLMs' effectiveness in this context.

Efficient Data Processing and Analysis. Much of the analyzed literature (87.3% of the dataset) focuses on using LLMs to improve real-time data processing from IoT devices while reducing computational overhead. Practical applications of LLMs in this context are [SLR 17, 54]. In [SLR 17], fine-tuned LLMs like BERT and DistilRoBERTa performed sentiment analysis on log data for anomaly detection in drone forensic timelines. By identifying negative sentiments in log messages, the system accurately distinguished between normal and anomalous events in real time. DistilRoBERTa achieved the best performance with an accuracy of 92.527% and F1-score of 90.556%. In [SLR 54], the LIMU-BERT model was used to extract features from unlabeled IMU sensor data for human activity recognition (HAR) and device placement classification (DPC). Designed for mobile deployment, it achieved up to 98.4% accuracy in HAR and DPC tasks.

Semantic Data Handling. LLMs (used in 61.8% of cases) excel in semantic service clustering, similarity models, and rule-based analysis, enhancing data interoperability and understanding. The ability of LLMs in semantic data handling is presented in [SLR 22]. The study employs a BERT-based model to assess the semantic similarity between different data points, such as sensor readings and device-generated messages, leading to improved accuracy in data categorization with an F1-score of 0.89. Another example is detailed in [SLR 27], which introduces a model optimized to be deployed in resource-constrained edge environments. This model, based on BERT and RoBERTa, leverages LLMs to handle semantic data dynamically to optimize resource allocation and improve real-time decision-making. The study shows a 30% latency reduction and a 25% throughput improvement in a simulated 5G environment over conventional edge computing models.

Data Classification and Categorization. LLMs have also played a crucial role in enhancing the classification process's accuracy and efficiency (in 58.2% of the analyzed studies), addressing the IoT data classification issue. [SLR32 32] exemplifies the application of LLMs in detecting security and privacy violation rules within trigger-action IoT platforms, such as IFTTT. The researchers used BERT-based models to classify applets by potential risks, achieving 88-93% precision and recall on a dataset of 76,741 rules. Another application is

illustrated by [SLR36 36], where LLMs were used to detect Distributed Denial of Service attacks in IoT device logs. The study used OpenAI's GPT-3.5 and GPT-4 models for few-shot learning and fine-tuning on security datasets, achieving 95% and 96% accuracy and demonstrating LLMs' effectiveness in identifying network security threats.

Data Integration and Fusion. Data integration and fusion pose significant challenges in IoT domains like healthcare and industry due to raw sensor data's non-interoperability nature. Despite LLMs' innate capability to convert data into structured formats, ensuring consistent analysis and actionable insights, only a modest percentage of the analyzed works (14.5% of the dataset) have investigated their application in data fusion and integration. An elucidative study in this context is [SLR 40], which addresses the challenge of converting raw sensor data from non-interoperable formats into structured formats like JSON or XML. This study utilizes GPT-4 to enhance the reusability of sensor data in IIoT, making it more accessible and valuable for third-party applications. Specifically, GPT-4 demonstrated a precision of 93.51% and a recall of 85.33% in transforming HTML sensor data into structured formats.

Data Management And Analysis: Summing Up

LLMs efficiently manage and analyze large-scale IoT data, boosting semantic analysis, classification, and anomaly detection. They enhance processing speed, semantic understanding, and data integration, aiding decision-making and operational efficiency in IoT systems.

4.1.3. Scalability, Reliability, and Availability

Most of the research, representing 85.5% of the dataset, investigates how LLMs improve IoT performance, handle growing loads, and ensure continuous functionality. It covers key areas such as network optimization and service enhancement, as described below.

Enhanced Scalability. LLMs have been leveraged to support the seamless expansion and efficient performance of IoT networks across domains like healthcare, military, smart homes, industry, and transportation. In the military IoT domain, LLMs address scalability challenges by optimizing computational demands and robust performance even in adversarial conditions in mission-critical scenarios [SLR 46]. By leveraging GPT-3 for natural language understanding and context interpretation, as well as specialized foundation models for specific tasks, complex tasks are decomposed into manageable subtasks. This approach achieved 90.82% accuracy in out-of-context object detection. In smart environments, [SLR 26] introduced PipeBERT, based on BERT, to optimize ARM CPU clusters in edge devices by splitting BERT models and mapping them onto ARM architecture. PipeBERT showed a 48.6% increase in average throughput and a 61% reduction in the energy-delay product compared to homogeneous inference, showcasing its scalability and efficiency for various IoT applications on resource-constrained devices.

Improved Reliability. A key application of LLMs, comprising 56.4% of the dataset, is enhancing systems' reliability in environmental monitoring and smart cities by improving data accuracy, anomaly detection, and decision-making consistency. For instance, [SLR33 33] investigates how LLMs can interact with the physical world via IoT sensors and actuators. The study demonstrates ChatGPT's ability to reliably interpret sensor data and handle tasks like activity sensing and heartbeat detection. ChatGPT-4 achieved up to 100% accuracy in motion detection and a mean absolute error of 1.92 beats per minute in heartbeat detection. Another study, [SLR53 53], examines the use of LLMs for automated building operations monitoring, focusing on the semantic mapping of operational data to create digital twins. The system standardizes heterogeneous data points by employing BERT-based models, which enhances reliability and reduces errors. The study reports an impressive F1 score of over 95% in data classification, showcasing the model's robustness with diverse datasets.

Increased Availability. Only a few studies (7.3% of the dataset) focused on healthcare, industry, and edge computing explore how IoT services can be more widely accessible. [SLR11 11] highlights the role of LLMs in addressing the IoT availability challenge through a lightweight BERT-based service embedding for dynamic service recommendations in edge computing. The system uses content-based filtering and semantic clustering to ensure that services are recommended efficiently and remain available even as the service environment evolves. The model reduced time complexity by 19% to 56% and achieved 80-100% precision in service recommendations, demonstrating its efficacy.

Scalability, Reliability, And Availability: Summing Up

LLMs have improved IoT scalability, reliability, and availability by optimizing computational demands, data processing accuracy, and service recommendations, significantly enhancing performance across IoT applications.

4.1.4. Integration Difficulties

The "Integration Difficulties" category (58.2% of the dataset) investigates how LLMs address the complexities of connecting diverse IoT systems and technologies using different protocols and standards. The studies show that LLMs enhance integration via advanced natural language processing, improving communication and heterogeneous data interpretation.

Ease of System Integration and Configuration. Integrating IoT devices and automating system configurations is complex. However, LLMs can simplify these tasks, as reflected in 56.4% of the dataset. One notable application of LLMs in this context is shown in [SLR9 9], where a BERT-based model efficiently disaggregates energy in smart grids by classifying household appliances using energy data, eliminating the need for extra sensors. The model attained a mean accuracy of 89% for fridge classification and a mean precision of 71%, significantly

simplifying the setup and maintenance of energy monitoring systems. LLMs have also been effective in complex device interaction configurations without extensive labeled data. Indeed, [SLR20 20] introduces DeviceGPT is an LLM pre-trained on large datasets that automatically learns interactions among IoT devices without requiring extensive labeled data, simplifying the configuration process. Tested on a real-world dataset, it achieved 82.45% device identification accuracy and geolocation accuracies of 32.44% for distances under 10km.

Semantic Integration. LLMs have improved semantic understanding and data integration in IoT systems (23.6% of the dataset). Enhancements include more intuitive trigger-action programming in smart homes, as show in [SLR 41]. The study introduces ChatIoT, a system that employs ChatGPT for the zero-code generation of Trigger-Action Programs in IoT environments. This system allows users to create IoT rules through natural language interactions, significantly simplifying the integration of IoT devices and services. ChatIoT utilizes a Prompts Manager to optimize input processing and rule specificity and a Cross-modal Model Zoo to handle multimodal sensor data. Evaluated with a dataset and integrated with Home Assistant, it achieved a rule generation accuracy of 94.1% to 98.5%, highlighting the potential of LLMs to enhance semantic integration and simplify IoT device programming.

Technological Compatibility. In IoT fields like Industry, Smart Homes, and Smart Grids, LLMs have enhanced technological compatibility by facilitating seamless communication and data exchange across devices with different protocols and formats (12.7% of the dataset). An example is the zero-sample face retrieval method that integrates GPT-3 with visual base models [SLR 39]. This approach eliminates the need for extensive data collection and model training, which is traditionally required in IoT text retrieval methods. The method converts discrete facial features into natural language descriptions using cue words as prompts and employs the CLIP model to align vector representations of text and images. It attained a top-1 accuracy of 72% and a top-3 accuracy of 93%, reducing data and computational costs and enhancing the compatibility of face recognition systems across various IoT devices.

Integration Difficulties: Summing Up

LLMs have effectively tackled IoT integration challenges, simplifying system configuration, enhancing semantic integration, and improving technological compatibility, leading to better device interaction, rule generation, and data exchange across protocols and formats.

4.1.5. Interoperability and Standardization

Interoperability and standardization remain significant challenges in IoT systems due to the diversity of devices and communication protocols, which can lead to complexities in data exchange and system functionality without standardization. While only a few studies (12.7% of the entire dataset) have investigated the application of LLMs in this area, the results have been encouraging in enhancing seamless interaction and communication standards. Studies show that applying LLMs reduces integration complexities, facilitating smoother interoperability and generating interoperable interfaces.

Protocol and Data Format Harmonization. Although only 5.5% of the dataset explores this area, initial findings suggest that LLMs can significantly enhance interoperability and standardization within IoT ecosystems. These models harmonize communication protocols and data formats as shown in [SLR 1]. The study introduces a Neural RFC Knowledge Graph using LLMs to detect contradictions in IoT protocol documents. It automatically parses these documents, builds knowledge graphs, and detects contradictions using BERT and GPT-2-xl models. This approach ensures clear and consistent protocol specifications, reducing security risks and enhancing interoperability across IoT devices. The model showed high accuracy in entity recognition (up to 99%) and effective contradiction detection in IoT messaging protocols.

Service Discovery and Composition. Research shows that LLMs can effectively tackle service discovery and composition challenges in IoT applications, enhancing system efficiency and robustness by addressing interoperability and standardization issues. Despite limited studies (7.3% of the dataset), LLMs demonstrate the potential to reduce complexities in managing IoT service interactions across devices and platforms, as shown in [SLR 13]. The study explores semantic service clustering using a lightweight BERT-based model to capture semantic information from service invocation sequences and generate service embeddings with neural language models. It employs convolutional attention within a transformer architecture for efficiency and uses K-means clustering to form semantic clusters. Evaluation metrics like purity and entropy showed the model's effectiveness, achieving purity values between 50% and 77% and optimal performance at K=400 clusters. The model also reduced time complexity by 19% to 56% and cut training time from 10 hours to about 6 hours.

Interoperability and Standardization: Summing Up

LLMs boost IoT interoperability by harmonizing protocols and data formats, improving service discovery, achieving high accuracy in entity recognition, and cutting time complexity and training time.

*4.2. **RQ$_2$**—What methodologies and optimization techniques have been employed to implement LLM-based solutions for IoT applications, and what are the associated technical requirements and constraints?*

In addressing the second research question, we explore methodologies and optimization techniques used to implement LLM-based solutions for IoT applications, along with the associated technical requirements and constraints and new challenges introduced by LLMs. We have organized these aspects in the following categories.

4.2.1. *Pre-training and Fine-tuning Large Language Models*

Pre-training and fine-tuning LLMs are standard techniques (76.4%) for boosting performance. Pre-training uses large text corpora like Wikipedia for general language understanding, followed by fine-tuning on smaller, task-specific datasets, such as feature extraction or text generation. Some studies (9.1%) [SLR 1, 15, 18, 24, 54] employed iterative training for gradual improvements. For instance, [SLR 18] iteratively trained BERT to detect APT attack sequences in IIoT better. Pre-training or fine-tuning LLMs on specific IoT datasets is challenging due to the diverse, noisy nature of IoT data and resource-intensive demands, which complicates preprocessing and adaptability and leads to generalization issues in new or varied IoT environments. As a result, only 22% of studies have used pre-training, and only 20% have used both techniques. Fine-tuning is more common, as reported by 75% of the reviewed studies. In particular, the combined use of both techniques has been primarily employed to address the IoT challenges of scalability, reliability, availability, security, and privacy (see Figure 4). On the other hand, prompt engineering and in-context learning (guiding model responses using input design) are less resource-intensive. These methods, combined with techniques like Retrieval Augmented Generation (RAG) [49], incorporate external information into the model's context, allowing it to adapt to new tasks and elaborate new knowledge. Some studies [SLR 33, 36, 37, 39, 40, 41, 42, 44, 45, 46, 52] used prompt engineering for desired responses without LLM weight modification, and a few [SLR 41, 46] utilized RAG for external knowledge integration.

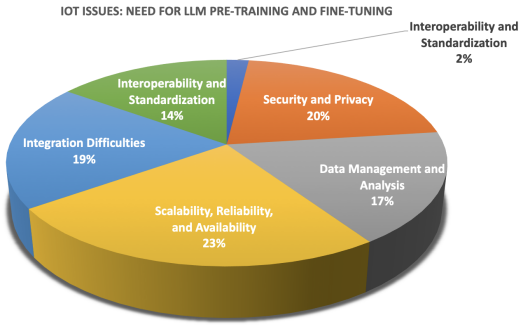


Figure 4: RQ2: IoT challenges addressed by pre-training and fine-tuning.

Pre-Training And Fine-Tuning: Summing Up

Pre-training and fine-tuning LLMs on domain-specific datasets enhance IoT performance despite data diversity and resource constraints. These techniques also pose challenges like high computational demands and storage needs, affecting adaptability. Generalization issues are also significant, as models may struggle with new, unseen data or different domains common in IoT environments.

4.2.2. *Software and Hardware Configurations*

The category focuses on setting up software tools, libraries, and hardware, including programming languages, development environments, and hardware like servers and processors for implementing LLMs in IoT applications. For Software Configuration, BERT and its variants are the most commonly used LLMs (67.3%). Python is the leading programming language (70.9%), while popular frameworks include TensorFlow (18.2%), Hugging Face Transformers (18.2%), scikit-learn (16.4%), and Pytorch (15.63%). Key Python libraries are OpenAI API (18.2%), Keras (9.1%), and spaCy (9.1%).

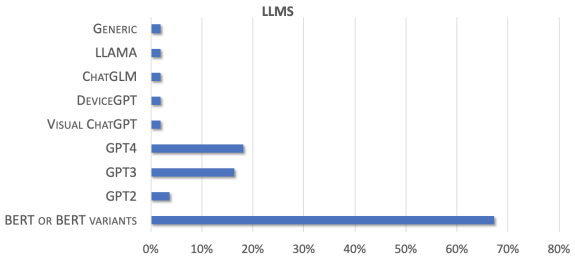


Figure 5: RQ2: LLMs used to develop LLM-based solutions

BERT's appeal is due to its open-source nature, robust tool support, and ability to handle complex NLP tasks. Python is favored for its simplicity, versatility, and solid library ecosystem. Combining TensorFlow, Hugging Face Transformers, scikit-learn, and Pytorch with Python underlines their compatibility and preference for ML model development and training. Finally, spaCy and Keras are critical for NLP and deep-learning tasks, while the growing adoption of the OpenAI API indicates a trend toward leveraging GPT-3 and GPT-4 models.

Various hardware configurations, including high-end GPUs, cloud environments, and embedded systems like Raspberry Pi, were used, reflecting different computational needs for LLMs in IoT applications. Operating environments mentioned include Windows 10 (5.5%), Ubuntu (7.3%), cloud environments (9.1%), and embedded systems (16.2%). Most studies used GPUs (40%) and multi-core CPUs (23.6%). Some studies specified RAM requirements (18.2%) ranging from 8 GB to 1 TB. It is worth noting that high computing resources, such as GPU and multicore CPU, are used in 71% of the studies addressing interoperability issues in IoT and in 38%- 45% of the studies addressing the other IoT challenges.

Software and Hardware Configurations: Summing Up

In IoT research, BERT and its variants are the most commonly used large language models. Python is the preferred programming language, paired with frameworks like TensorFlow, Hugging Face Transformers, scikit-learn, and PyTorch. Frequently used Python libraries include OpenAI API, Keras, and spaCy. The primary operating systems are Windows 10 and Ubuntu, with some use of cloud environments and embedded systems. Hardware setups often feature high-end GPUs and multi-core CPUs, with 8GB to 1TB RAM to accommodate various computational needs. However, LLMs also introduce new challenges, such as increased computational demands and the necessity for scalable, efficient resource management.

4.2.3. Use of Specific Techniques

LLM-based solutions often require integration with other techniques to address IoT issues, mainly to enhance LLM performance or process and analyze IoT data. Figure 6 shows how combining LLM with other techniques contributed to addressing IoT issues.

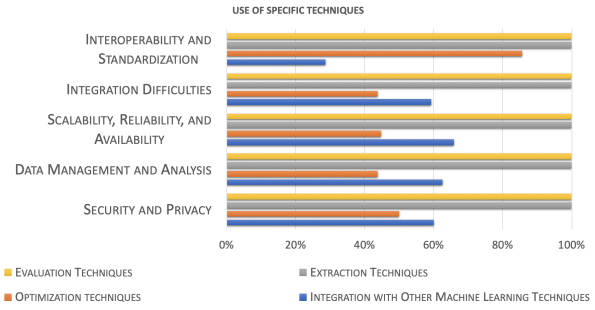


Figure 6: RQ2: IoT Issues - LLM integration with other techniques

Integration with Other Machine Learning Techniques. The integration of LLMs with other ML techniques has been widely observed in various studies (63.6% of the entire dataset), both in Deep Learning (DL) architectures (34.38%) and traditional ML techniques (40%). Specifically, 66% of studies addressing Scalability, Reliability, and Availability, 62% addressing Data Management and Analysis, 60% addressing Security and Privacy, 59.4% addressing Integration Difficulties, and 28.6% addressing Interoperability and Standardization have integrated LLM solutions with other machine-learning techniques. Integrating DL techniques enhances feature extraction, classification accuracy, and complex data handling. For instance, [SLR 3] combined BERT with CNN and BiGRU, improving classification accuracy on short-text datasets and a 5G-enabled IoT social dataset. Similarly, [SLR 10] used BERT with CNN and LSTM for malware detection in IoT devices, achieving higher detection rates and lower false positives. In [SLR 16], LLMs combined with DL-based computer vision techniques enhanced remote sensing and image processing in IoT, enabling tasks like object detection, image segmentation, and NLP. Traditional ML techniques aid information retrieval, category identification, and pattern recognition. In the context of LLMs, clustering techniques, such as K-means, are used to group similar data points based on semantic similarity. For example, [SLR 13] used K-means for service clustering, while [SLR 47] utilized Random Forest for feature selection from datasets providing URL features without the URL strings before passing the URLs to the LLM for further analysis. However, integrating LLMs with other ML techniques introduces challenges like increased computational overhead, complex data management, heightened security concerns and system complexity.

Optimization Techniques. Several studies (43.6% of the dataset) employed optimization algorithms to enhance LLM performance in specific IoT tasks. These mathematical methods fine-tune LLM parameters, guiding the iterative learning process by minimizing a loss function. For example, [SLR 9] used the AdaX algorithm to optimize a BERT model for energy disaggregation in smart grids. In contrast, [SLR 1] applied hyperparameter optimization to BERT and DistilBERT for detecting contradictions in CoAP and MQTT specification documents. Optimization algorithms are vital for LLM performance in IoT, particularly in areas like Security, Privacy, Scalability, Reliability, Availability, Data Management, and Integration, being used in nearly 50% of studies addressing these challenges. They are even more prevalent (over 85%) in studies addressing Interoperability and Standardization. Despite their importance, these algorithms are computationally intensive and complex, requiring specialized knowledge for effective implementation. Additionally, the non-deterministic nature of LLMs introduces further challenges in optimization. Non-determinism can lead to variability in the optimization results, making it difficult to consistently achieve the best performance.

Extraction Techniques. Extraction techniques are vital in processing and analyzing data in IoT solutions, particularly those based on LLM. Various methods facilitate interpretability, operational efficiency, and decision-making within IoT ecosystems.

"Relation Extraction" (10.9% of the dataset) identifies and classifies semantic relationships between entities in text, facilitating the understanding of connections, such as inferring relationships from sensor data to determine steps counted and activity levels (e.g., [SLR 33]). "Rule Extraction" (14.5%) derives explicit rules from complex datasets or models, automating processes and improving decision-making, as seen in formulating "if-then" logic for IoT TAPs from natural language inputs (e.g., [SLR 41]). "Named Entity Recognition (NER)" (20%) classifies named entities like persons, organizations, and monetary values within the text, enhancing data handling by mapping user intents to network policies for practical network configurations (e.g., [SLR 19], [SLR 51]). "Feature Extraction for Sequential Data Analysis" (50.9%) isolates relevant features from time-series data (such as continuous sensor readings) to facilitate effective modeling and prediction by learning algorithms, aiding in anticipating future states, such as extracting place names for crisis response in Social IoT ([SLR 6]). "Feature Extraction for Textual Data Analysis" (36.4%) transforms raw textual data into actionable insights, aiding monitoring and management in applications like Smart Cities, where textual descriptions generate decision-making logic (e.g., [SLR 44]). "Cross-modal Feature Extraction and Alignment for Data Retrieval" (7.3%) correlates features from different data sources, enhancing tasks like image captioning and visual question answering by aligning visual information with natural language (e.g., [SLR 46]). Each technique contributes to IoT systems' seamless integration and functionality, but requires advanced methods to handle context-rich data for accurate semantic relationship identification, rule derivation, and feature extraction. Figure 7 shows in which percentage the extraction techniques have been used to address the different types of IoT challenges.

Evaluation Techniques. Assessing the performance and effectiveness of LLM-based solutions in IoT applications relies

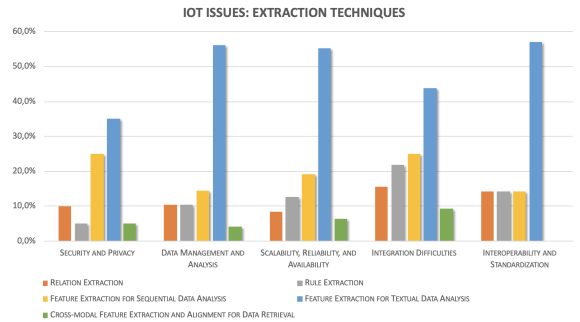


Figure 7: RQ2: IoT Issues - Extraction Techniques used in LLM-based solutions in addressing IoT challenges.

on a variety of metrics. Metrics like accuracy, efficiency, and reliability offer insights into various performance aspects. The primary metrics used include Accuracy, Precision, Recall, F1-score, BLEU, ROUGE, Mean Average Precision (MAP), Mean Reciprocal Rank (MRR), Mean Absolute Error (MAE), and Area Under the ROC Curve (AUC).

Accuracy, Precision, Recall, and F1-score are commonly used, particularly for addressing Security and Privacy issues in IoT, appearing in 55% to 60% of studies and 70% to 75% of security-focused research. BLEU and ROUGE, although used in only 2% to 5% of the studies, are crucial for evaluating the linguistic quality of machine-generated text. MAP and MRR are important for precise data retrieval and are prominent in IoT Integration, Interoperability, and Standardization studies. MAE, relevant in applications like smart grid energy forecasting, is prevalent in research on Scalability, Reliability, and Availability. AUC, which has been used extensively in 85.7% of Interoperability and Standardization studies, measures classification performance effectively across different thresholds. These metrics help determine the robustness, accuracy, and utility of LLM-based solutions. However, given the intrinsic nature of LLM, assessing them for IoT applications should include metrics like token generation time, latency, throughput, compute efficiency, memory bandwidth, power consumption, and other hardware capabilities.

Use of Specific Techniques: Summing Up

Integrating LLMs with other ML techniques enhances feature extraction, classification accuracy, and complex data handling but introduces challenges like increased computational overhead, system complexity, and non-determinism. Optimization and extraction techniques are crucial for improving LLM performance in IoT tasks. In contrast, evaluation techniques must consider traditional and new metrics specific to LLMs, such as latency and compute efficiency.

4.2.4. Continuous Monitoring and Adjustment

Continuous monitoring and adjustment of LLM-based solutions enhance effectiveness in dynamic IoT contexts by adapting to new data patterns and user behaviors. It is also beneficial because LLMs struggle with complex queries and long data sequences, posing challenges such as high computational demands, real-time adaptation difficulties, data privacy concerns, and scalability issues. Figure 8 shows the contribution of the approaches to address IoT challenges.

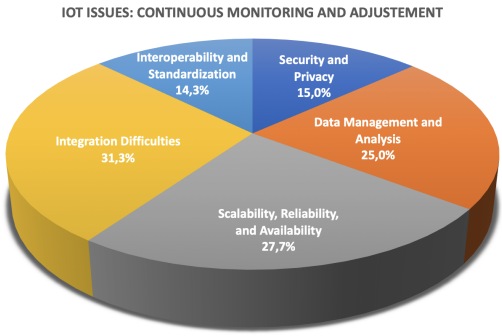


Figure 8: RQ2: Iot Issues - Continuous Monitoring and Adjustment usage.

Real-time Monitoring and Adjustment. This strategy involves continuously monitoring LLM-based IoT solutions and adjusting real-time operations based on current conditions or incoming data, constituting 16.4% of the dataset. For example, [SLR 2] implemented real-time monitoring to adjust access policies, while [SLR 5] used the Weights and Biases tool to track model performance during training. Additionally, [SLR 24] monitored a semantic similarity model's performance, allowing for adjustments to improve speed and disk space usage. Despite its complexity and high computational demands, this strategy promises to enhance real-time IoT performance.

Adaptive Learning and Feedback Incorporation. Refinement of LLM-based solutions based on past performance and user feedback is crucial for adjusting to dynamic IoT environments. Employed in 14.5% of the dataset, this approach enhances model accuracy and responsiveness. For example, [SLR 3] refined sentence embeddings with human annotator feedback, while [SLR 33] improved activity recognition and health monitoring. However, challenges remain in real-time adaptation, data privacy, and scalability.

Continuous Monitoring & Adjustment: Summing Up

Continuous monitoring and adjustment are vital in dynamic IoT environments to enhance performance by adapting to new data and user behaviors. These techniques can also address LLMs' lack of human-like understanding and difficulty with complex queries and long data sequences. However, real-time adaptation and data privacy concerns limit their broader use.

4.2.5. Data Collection and Preprocessing

"Data Collection and Preprocessing" entails gathering and preparing data for analysis using LLMs, crucial steps that directly influence model performance and accuracy in data-driven

studies. Figures 9 and 10 illustrate the approaches distribution in addressing IoT challenges.

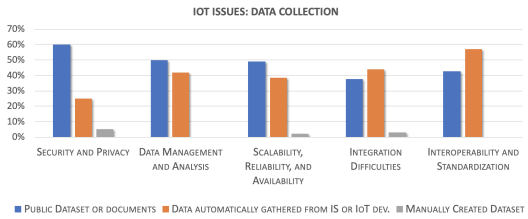


Figure 9: RQ2: IoT issues - Data Collection approaches.

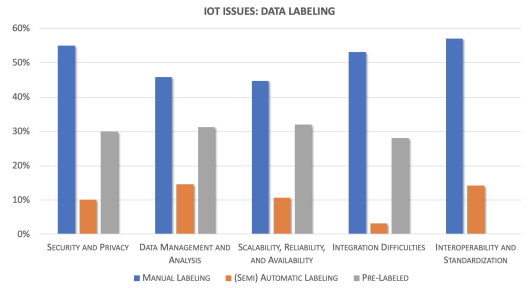


Figure 10: RQ2: IoT issues - Data Labeling approaches.

Data Collection. This step concerns gathering relevant data from various sources for processing and feeding into LLMs, with the approach depending on the specific IoT field or challenge. Most studies (87.3%) reported their data collection methods, with 40% emphasizing the need for qualitative data, highlighting the importance of careful planning. Public datasets or documents were used in 50.9% of the studies; 36.4% automatically acquired data from Information Systems or sensor platforms, and only 1.8% manually created datasets. This diversity in methods underscores the dynamic nature of the field.

Data Preprocessing. This step converts raw data into a format efficiently processed by LLMs, enhancing performance. Most studies (83.6%) highlight its importance, involving cleaning (25.5%) to remove irrelevant data, reduce noise, and improve reliability. It includes tokenization, stop word removal, and lemmatization for dimensionality reduction and standardization. Although balancing data (e.g., SMOTE) to prevent bias is important [SLR 47], it is often overlooked. While most studies (81.8%) focus on transforming or standardizing data, data augmentation is reported in only 1.8% of studies [SLR 7].

Data Labeling. Labeling quality significantly impacts model performance. About 47.3% of studies use domain experts' manual labeling, ensuring high accuracy, which is crucial for context-sensitive scenarios. However, this method is time-consuming and less scalable. Around 12.7% of studies use automated or semi-automated labeling to speed up the process, though it may compromise accuracy. Pre-labeled datasets, used in 30.9% of studies, expedite preparation but often yield lower performance than manually labeled data.

Data Collection and Preprocessing: Summing Up

Effective LLM-based IoT solutions require high-quality data, proper cleaning and transformation, and accurate labeling for supervised learning. Automated techniques can speed up labeling, but LLMs still need extensive, diverse datasets and pose privacy, ethical, and bias challenges.

5. Discussion, Implications, and Future work

The SLR highlighted several observations that warrant further exploration, serving as a starting point for future research.

5.1. Addressing Research Questions

LLMs contribute to developing more effective IoT applications and services. They enhance various aspects of IoT, including security, data management, scalability, reliability, and interoperability. They improve access control, threat detection, and security violation identification in IoT security and privacy. LLMs also strengthen IoT data processing and analysis, leading to improved processing speeds, semantic data handling, and classification. They enhance IoT systems' scalability, reliability, and availability and simplify IoT ecosystem integration. Moreover, the multimodality of LLMs enriches their applicability in IoT through comprehensive data analysis and interpretation. [SLR 16, 39, 46, 50].

Therefore, given the results provided in Section 4, we can report the summary of the answer to RQ_1:

Answer to RQ_1: Summing up

LLMs effectively enhance IoT applications by improving security through advanced threat detection, managing data by parsing and classifying unstructured information, and facilitating scalability and interoperability.

Another important aspect that emerges from our analysis is that LLM-based solutions must be implemented through a meticulous workflow to address IoT challenges effectively. This workflow involves data collection and preprocessing, pre-training, and fine-tuning on domain-specific datasets, tailored software and hardware configurations, integration of advanced machine-learning techniques and optimization algorithms, continuous monitoring and adaptive adjustments, and thorough evaluation of results. Implementing LLMs in IoT aims to develop robust solutions that are efficient, secure, adaptable, and user-friendly. However, implementing LLMs in IoT introduces several challenges. First, LLMs are computationally intensive and require substantial resources for training and inference [50, 51]. Secondly, acquiring large amounts of labeled data for effective training and achieving dynamic adaptation of LLMs without extensive retraining remains challenging. Additionally, LLMs lack human-like understanding and have limitations in handling complex queries and longer data sequences. The opaque internal workings of LLMs create a "black box" effect, and their non-deterministic nature complicates consistency and

Figure 11: Relationships between key findings (green), challenges (orange), and problems addressed (blue) in reviewed studies (grey).

reliability in IoT applications. Furthermore, the cost and dependence on external providers for state-of-the-art LLMs introduce latency, privacy concerns, and the risk of service disruptions. Bias and generalization issues also pose significant challenges, as LLMs can propagate societal biases present in the training data and may struggle to generalize effectively in varied IoT contexts. Finally, ethical and privacy concerns are significant, as LLMs can inadvertently expose sensitive information.

Thus, we can provide the summary of the answer toRQ_2:

Answer to RQ_2: Summing up

LLM-based solutions for IoT challenges are implemented through pre-training on large datasets and fine-tuning for specific applications. These implementations often require hardware like GPUs or TPUs and software frameworks like TensorFlow or PyTorch. Real-time monitoring, feedback loops, and thorough data handling processes are crucial, but they require significant expertise. Nonetheless, LLMs introduce challenges such as high computational costs, data labeling difficulties, lack of human-like understanding, and their "black box" nature. Additionally, reliance on external providers raises latency and privacy concerns, while biases and ethical issues complicate deployment.

Figure 11 illustrates the relationships between key findings and challenges emerged from the analyzed papers.

5.2. Future Directions

The potential of LLMs in addressing IoT challenges is vast and largely untapped. Future research could focus on several promising directions, including exploring new applications of LLMs in IoT, adopting privacy-preserving techniques, reducing the LLMs' computational requirements, and investigating the use of LLMs in emerging IoT fields. We report some research directions, prioritized based on their potential impact.

5.2.1. *Exploring New Applications*

LLMs have shown promise in various IoT domains, but their potential must be explored. In smart agriculture, LLMs could analyze data from weather stations, soil sensors, and satellite images to improve crop yield and reduce resource wastage, e.g., they could interpret weather forecasts and historical crop data to predict future yields, supporting planning and resource management. In smart transportation, LLMs could analyze traffic data to provide real-time updates and route recommendations. By interpreting data from traffic cameras, GPS, and social media, LLMs could predict traffic conditions and suggest optimal routes, reducing congestion and enhancing system efficiency. In smart cities, LLMs could analyze data from social media, sensors, and IoT devices to improve city planning and management. Future research could also explore using LLMs in real-time healthcare applications to monitor health, predict risks, interpret clinical reports, and provide timely intervention alerts.

5.2.2. *Investigating the Use of LLMs in Emerging IoT Fields*

As the IoT landscape evolves, new fields and applications constantly emerge, presenting unique challenges and opportunities. With their text understanding and generation capabilities, LLMs could address these challenges and capitalize on these opportunities. In the Internet of Bio-Nano Things, involving nanotechnology and biotechnology, LLMs could analyze and interpret data from bio-nano devices, enhancing their

communication and interaction. For instance, LLMs could interpret signals from bio-nano sensors, improving their environmental detection and response. In Quantum IoT, which integrates quantum technologies, LLMs could analyze quantum data, enhancing quantum communication and computation. For example, LLMs could interpret quantum computation results, making them more effective for IoT applications. In Space IoT, involving space exploration and research, LLMs could analyze space data, improving space communication and research. For instance, LLMs could interpret signals from space probes, enhancing their environmental detection and response. Future research could explore the use of LLMs in these and other emerging IoT fields, leveraging their unique capabilities to address these areas' specific challenges and opportunities.

*5.2.3. **Privacy-preserving Techniques***

Privacy-preserving techniques are required for implementing LLMs in IoT applications. Differential privacy adds noise to data to protect individual privacy while allowing for pattern analysis [52]. Federated learning enhances privacy by training a centralized model using distributed data from multiple IoT devices, thus keeping raw data local [53]. Despite these advantages, each technique faces specific issues. Federated learning's challenges include handling non-IID data, communication overhead, and robustness against malicious clients [54]. Security risks involve poisoning attacks, backdoor attacks, and membership inference attacks, which can compromise model integrity and privacy. Additionally, attacks based on Generative Adversarial Networks (GANs) generate synthetic data that mimics real data, potentially leading to data leakage or manipulation. Differential privacy also has vulnerabilities. While it protects data by adding noise, it is still susceptible to reverse-engineering attacks, allowing adversaries to expose sensitive information [54]. Several defense mechanisms are needed to mitigate these risks. Anomaly detection, secure aggregation protocols, and adversarial training form the first line of defense. Secure Multiparty Computation and Homomorphic Encryption allow computations on encrypted data without exposing it [55], ensuring only aggregated updates are accessible to the central server. Blockchain technology can provide a tamper-proof log of transactions, and Trusted Execution Environments can offer secure areas for sensitive computations, enhancing data integrity and confidentiality in federated learning systems [55].

Future research should improve the efficiency and scalability of these privacy-preserving techniques for LLM-based IoT applications. Additionally, enhancing the interpretability and transparency of LLMs within federated environments is crucial to comply with ethical guidelines and standards.

*5.2.4. **Reducing Computational Requirements***

The computational demands of LLMs present a challenge for IoT applications. Future research should reduce these requirements, making LLMs more suitable for IoT. One approach is compression techniques, which reduce model size and complexity without significant performance loss. Knowledge distillation, for example, trains a smaller model (student) to mimic a larger, pre-trained model (teacher), learning its generalization ability. Pruning removes unnecessary parameters or layers, and quantization reduces parameter precision, minimizing size and computational needs with minimal performance impact. Recent advancements in 1-bit LLMs, like BitNet b1.58, show promise in reducing memory and computational demands while maintaining performance [56]. BitNet b1.58, using ternary weights -1, 0, 1, matches full-precision LLMs but with lower latency, memory, and energy use. This approach could lead to the development of hardware optimized for 1-bit LLMs, enhancing efficiency and suitability for IoT applications. Another critical aspect is evaluating LLM solutions based on key performance metrics such as time to generate a token, latency, throughput, and hardware utilization. It is essential to assess how efficiently the model uses computational resources, memory bandwidth, power consumption, and other hardware capabilities in IoT environments to identify the most suitable LLM solution. Edge computing is another promising direction, moving computation closer to IoT devices to reduce latency and bandwidth. Research could focus on strategies for implementing LLMs in edge environments, using compression techniques, developing efficient architectures for streaming data, or exploring distributed learning like federated learning. Developing energy-efficient GPU architectures or other suitable hardware accelerators for IoT applications could also be beneficial.

*5.2.5. **Ethical Implications and Biases***

Using LLMs in IoT applications raises significant ethical and bias concerns. A primary issue is the tendency of LLMs to hallucinate or generate false or misleading information based on their internal patterns and biases [50]. This issue can be particularly problematic in sensitive areas such as healthcare, where misinformation can lead to a loss of trust and potentially harm patients. Additionally, biased data can lead to skewed outcomes, especially in sensitive areas like healthcare, exacerbating issues of fairness and equity [57].

To address the ethical concerns, future research should incorporate domain-specific knowledge (i.e., medical knowledge in healthcare) into LLMs to enhance their accuracy and reliability. Developing improved evaluation metrics, benchmark datasets, and mitigation methods could reduce the risk of hallucination and improve the faithfulness of AI in medical and other critical applications. Concrete frameworks and solutions should be established to address these ethical concerns holistically. Implementing XAI techniques can provide transparency in LLM decision-making processes. Adopting Federated Learning combined with Differential Privacy ensures data privacy and security while maintaining model performance (Section 5.2.3). Ethical guidelines and standards, such as those proposed by the EU Artificial Intelligence Act, can provide a comprehensive framework for the responsible use of LLMs. These guidelines should address accountability, transparency, fairness, and privacy, ensuring that LLMs respect user rights and promote trust.

To address the issues of bias, techniques such as data augmentation, synthetic data generation, and oversampling of underrepresented groups can improve fairness [58]. Bias detection and correction algorithms, along with XAI solutions, can further mitigate biases. Additionally, developing robust bias de-

tection and mitigation frameworks is essential for fairness and equity in LLM applications [59].

5.2.6. *Pre-training, Fine-tuning, and Implementing LLMs*

Future research could focus on developing new pre-training and fine-tuning methodologies tailored to specific IoT domains. Researchers could explore domain-specific corpora for pre-training LLMs, enhancing their performance in specific IoT applications. New fine-tuning methodologies could also consider IoT data's unique characteristics, such as its temporal nature, high dimensionality, and privacy concerns. Transfer learning techniques could leverage pre-trained LLMs for new IoT tasks, reducing the required training data and speeding up the training process. Future research could also develop new strategies and tools to simplify the LLM's implementation process. Researchers could develop automated machine learning (AutoML) tools, automating LLM selection, hyperparameter tuning, and evaluation. Explainable AI techniques (XAI) could be explored to improve the interpretability and transparency of LLMs, aiding IoT practitioners in understanding and fine-tuning their models. Advanced optimization algorithms could be developed to fine-tune LLM parameters more effectively, improving performance and reducing training time.

5.2.7. *Machine Learning Techniques*

Integrating LLMs with other machine-learning techniques shows promise for future research, potentially enhancing LLM performance and addressing complex IoT challenges more effectively. As reported in Section 4, several studies have demonstrated the benefits of such integration. Future research could explore integrating LLMs with reinforcement learning or ensemble learning. Reinforcement learning could train LLMs for optimal decision-making in IoT applications, while ensemble learning could enhance robustness and accuracy by combining multiple LLM predictions.

5.2.8. *Generating or Collecting Large and Diverse Training Datasets*

The successful implementation of LLMs in IoT applications relies on large and diverse training datasets, which are challenging to generate due to IoT data's dynamic nature and volume. Future research could focus on new data collection and generation techniques, such as synthetic data generation through data augmentation and generative adversarial networks (GANs). Data augmentation creates new data by transforming existing data, while GANs generate new data instances using two neural networks. These methods could enhance the size and diversity of training datasets, improving LLM performance in IoT. Additionally, few-shot learning could be explored to enable LLMs to learn from small datasets, which is beneficial in IoT applications where large labeled datasets are complex to obtain. Moreover, introducing 1-bit LLMs like BitNet b1.58 (see 5.2.4 could facilitate using more extensive and diverse datasets, new experimentations, and deployment in resource-constrained IoT environments by reducing the overall training and inference costs.

5.2.9. *Addressing Generalization in LLMs*

Future research should enhance LLMs' reliability in IoT applications by improving model generalization across diverse environments [57]. Generalizing LLMs across IoT domains is challenging due to unique data characteristics. For this reason, it is vital to use different strategies. For instance, strategies like domain adaptation, transfer learning, and fine-tuning pre-trained models on domain-specific data can enhance performance [18]. In addition, few-shot and zero-shot learning allow LLMs to adapt to new tasks with minimal labeled data, while RAG techniques allow LLMs to incorporate external information into the model's context and elaborate new knowledge. Moreover, ensemble methods combine multiple LLMs trained on different domains for better robustness and generalization. Lastly, real-time monitoring and adaptive feedback mechanisms can address generalization issues as they arise. Adaptive learning techniques enable continuous LLM refinement based on new data and user feedback. Engaging end-users in evaluation provides valuable insights into LLM performance across various IoT contexts, guiding improvements.

6. Threats to Validity

In this section, we discuss the potential threats to the validity of our SLR and the strategies we took to mitigate them. We have categorized these threats into four main types, following the guidelines proposed by Wohlin et al. [60]: construct validity, internal validity, external validity, and conclusion validity.

The main threat to construct validity is the definition of the search string. To mitigate this threat, we used many keywords and their synonyms related to IoT and LLMs. We also used wildcard characters to capture variations of the keywords. Furthermore, we manually searched the references of the initially identified studies to find additional relevant studies that the search string might have missed. The internal validity threat pertains the selection and evaluation of the studies. We defined clear and objective inclusion, exclusion, and quality criteria to mitigate this threat. Two authors selected and evaluated the studies independently, and disagreements were resolved through discussion. Regarding external validity, the coverage of the literature is a potential issue. The literature we reviewed might only represent some of the studies on the role of LLMs in solving IoT challenges. To mitigate this threat, we used multiple databases to search for studies. We also manually searched the references of the initially identified studies to find additional relevant studies. Finally, The synthesis and interpretation of study findings pose a threat to conclusion validity. To mitigate this threat, we used a systematic and transparent approach to synthesize and interpret the study findings. We also detailed our methodology and made our data extraction forms and dataset available for scrutiny.

7. Conclusion

The SLR has examined the role of LLMs in addressing various challenges within the IoT domain. Our findings highlight

the benefits of LLMs in enhancing data management, interoperability, security, privacy, and a wide range of IoT applications. For instance, LLMs have proven effective in detecting security threats in IIoT, identifying malware in IoT devices, and parsing and classifying healthcare data to improve patient care. However, the review also identified several challenges in implementing LLMs in IoT applications, such as high computational requirements and the need for extensive and diverse training datasets. These challenges must be addressed to realize LLMs' potential in IoT fully. To overcome these hurdles, we propose several future research directions. Future studies could focus on developing techniques to reduce the computational demands of LLMs, making them more suitable for IoT applications. Additionally, new data collection and generation methods could be explored to meet the need for diverse training datasets. Further research could also investigate new applications of LLMs in IoT, adopt privacy-preserving techniques, and explore their use in emerging IoT fields. We hope that our review will inspire innovative efforts for more effective IoT applications and services by exploiting LLMs.

References

[1] J. Gubbi, et al., Internet of things (iot): A vision, architectural elements, and future directions, Future Gener Comput Syst 29 (7) (2013) 1645–1660.

[2] S. Muralidharan, et al., Mdp-iot: Mdp based interest forwarding for heterogeneous traffic in iot-ndn environment, Future Gener Comput Syst 79 (2018) 892–908.

[3] T. hoon Kim, et al., Smart city and iot, Future Gener Comput Syst 76 (2017) 159–162.

[4] J. M. Talavera, L. E. Tobón, J. A. Gómez, et al., Review of iot applications in agro-industrial and environmental fields, Computers and Electronics in Agriculture 142 (2017) 283–297.

[5] M. Compare, P. Baraldi, E. Zio, Challenges to iot-enabled predictive maintenance for industry 4.0, IEEE Internet of Things J 7 (5) (2020) 4585–4597.

[6] M. Attaran, The impact of 5g on the evolution of intelligent automation and industry digitization, Journal of Ambient Intelligence and Humanized Computing (2023) 5977–5993.

[7] L. Babun, et al., A survey on iot platforms: Communication, security, and privacy perspectives, Computer Networks 192 (2021) 108040.

[8] S. Sicari, et al., 5g in the internet of things era: An overview on security and privacy challenges, Computer Networks 179 (2020) 107345.

[9] M. A. Ahad, G. Tripathi, S. Zafar, F. Doja, IoT Data Management—Security Aspects of Information Linkage in IoT Systems, Springer International Publishing, Cham, 2020, pp. 439–464.

[10] H. Rahman, M. I. Hussain, A comprehensive survey on semantic interoperability for internet of things: State-of-the-art and research challenges., Transactions on Emerging Telecommunications Technologies (2020).

[11] E. Lee, Y.-D. Seo, S.-R. Oh, Y.-G. Kim, A survey on standards for interoperability and security in the internet of things, IEEE Communications Surveys & Tutorials 23 (2) (2021) 1020–1047.

[12] M. Shanahan, Talking about large language models (2023). arXiv:2212.03551.

[13] T. Brown, et al., Language models are few-shot learners, in: Advances in Neural Information Processing Systems, Vol. 33, 2020, pp. 1877–1901.

[14] OpenAI, Gpt-4 technical report., Tech. rep., OpenAI (2023).

[15] J. Devlin, et al., Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proc. naacL-HLT, 2019, pp. 4171–4186.

[16] M. Chen, J. Tworek, H. Jun, et al., Evaluating large language models trained on code (2021). arXiv:2107.03374.

[17] E. Kasneci, K. Sessler, S. Küchemann, et al., Chatgpt for good? on opportunities and challenges of large language models for education, Learning and Individual Differences 103 (2023) 102274.

[18] W. X. Zhao, K. Zhou, J. Li, T. Tang, et al., A survey of large language models (2023). arXiv:2303.18223.

[19] C. Wu, et al., Visual chatgpt: Talking, drawing and editing with visual foundation models (2023). arXiv:2303.04671.

[20] M. Maaz, et al., Video-chatgpt: Towards detailed video understanding via large vision and language models (2023). arXiv:2306.05424.

[21] B. Kitchenham, S. Charters, Guidelines for performing systematic literature reviews in software engineering, Keele Univ. , U.K., Rep. no. EBSE-2007-01 (2007).

[22] H. Allioui, et al., Exploring the full potentials of iot for better financial growth and stability: A comprehensive survey, Sensors 23 (19) (2023) 8015.

[23] R. Lohiya, A. Thakkar, Application domains, evaluation data sets, and research challenges of iot: A systematic review, IEEE Internet of Things Journal 8 (11) (2021) 8774–8798.

[24] Y. Yang, et al., A review of iot-enabled mobile healthcare: technologies, challenges, and future trends, IEEE Internet of Things Journal 9 (12) (2022) 9478–9502.

[25] M. Conoscenti, et al., Blockchain for the internet of things: A systematic literature review., in: IEEE/ACS 13th AICCSA, 2016, pp. 1–6.

[26] P. Asghari, A. M. Rahmani, H. H. S. Javadi, Internet of things applications: A systematic review., Computer Networks 128 (2010) 241–261.

[27] G. Giordano, F. Palomba, F. Ferrucci, On the use of artificial intelligence to deal with privacy in iot systems: A systematic literature review, Journal of Systems and Software 193 (2022) 111475.

[28] A. S. Syed, D. Sierra-Sosa, A. Kumar, A. Elmaghraby, Iot in smart cities: A survey of technologies, practices and challenges, Smart Cities 4 (2) (2021) 429–475.

[29] A. Goudarzi, et al., A survey on iot-enabled smart grids: Emerging, applications, challenges, and outlook, Energies 15 (19) (2022) 6984.

[30] B. B. Sinha, R. Dhanalakshmi, Recent advancements and challenges of internet of things in smart agriculture: A survey, Future Generation Computer Systems 126 (2022) 169–184.

[31] M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, et al., A survey on the internet of things (iot) forensics: Challenges, approaches, and open issues, IEEE Communications Surveys & Tutorials 22 (2) (2020) 1191–1221.

[32] M. Malekshahi Rad, et al., Social internet of things: vision, challenges, and trends, Human-centric Computing and Information Sciences (2020).

[33] S. L. Ullo, G. R. Sinha, Advances in smart environment monitoring systems using iot and sensors, Sensors 20 (11) (2020) 3113.

[34] Y. Song, et al., Applications of the internet of things (iot) in smart logistics: A comprehensive survey, IEEE IoT-J 8 (6) (2021) 4250–4274.

[35] D. Oladimeji, et al., Smart transportation: an overview of technologies and applications, Sensors 23 (8) (2023) 3880.

[36] Z. A. Almusaylim, N. Zaman, A review on smart home present state and challenges: linked to context-awareness internet of things (iot), Wireless Networks 25 (2019) 3193–3204.

[37] W. Khan, et al., Industrial internet of things: Recent advances, enabling technologies and open challenges, Computers & Electrical Engineering 81 (2020) 106522.

[38] S. Nižetić, P. Šolić, et al., Internet of things (iot): Opportunities, issues and challenges towards a smart and sustainable future, Journal of Cleaner Production 274 (2020) 122877.

[39] A. Karale, The challenges of iot addressing security, ethics, privacy, and laws, Internet of Things 15 (2021) 100420.

[40] A. Vaswani, N. Shazeer, N. Parmar, et al., Attention is all you need., in: Advances in neural information processing systems, Vol. 30, 2017.

[41] H. W. Chung, L. Hou, S. Longpre, B. Zoph, et al., Scaling instruction-finetuned language models (2022). arXiv:2210.11416.

[42] H. Touvron, T. Lavril, G. Izacard, et al., Llama: Open and efficient foundation language models (2023). arXiv:2302.13971.

[43] T. L. Scao, A. Fan, C. Akiki, et al., Bloom: A 176b-parameter open-access multilingual language model (2023). arXiv:2211.05100.

[44] A. Zeng, X. Liu, Z. Du, Z. Wang, et al., Glm-130b: An open bilingual pre-trained model (2022). arXiv:2210.02414.

[45] E. Cavalcante, J. Pereira, M. P. Alves, et al., On the interplay of internet of things and cloud computing: A systematic mapping study, Computer Communications (2016) 17–33.

[46] S. S. Sohail, F. Farhat, et al., Decoding chatgpt: A taxonomy of existing research, current challenges, and possible future directions., Journal of King Saud University-Computer and Information Sciences 101675 (2023).

[47] J. Higgins, S. Green, Handbook for systematic reviews of interventions. version 5.1.0, The Cochrane Collaboration, 2011.

[48] G. De Vito, F. Palomba, F. Ferrucci, Online appendix (2024). URL https://github.com/gadevito/SLR_IoT_LLM

[49] P. Lewis, E. Perez, A. Piktus, et al., Retrieval-augmented generation for knowledge-intensive nlp tasks, Advances in Neural Information Processing Systems 33 (2020) 9459–9474.
[50] L. Fan, L. Li, et al., A bibliometric review of large language models research from 2017 to 2023 (2023). arXiv:2304.02020.
[51] B. Min, H. Ross, E. Sulem, et al., Recent advances in natural language processing via large pre-trained language models: A survey, ACM Comput. Surv. (jun 2023).
[52] Y. Zhao, J. Chen, A survey on differential privacy for unstructured data content, ACM Computing Surveys (CSUR) 54 (10s) (2022) 1–28.
[53] L. Li, Y. Fan, M. Tse, K.-Y. Lin, A review of applications in federated learning, Computers & Industrial Engineering 149 (2020) 106854.
[54] J. Hasan, Security and privacy issues of federated learning, arXiv preprint arXiv:2307.12181 (2023).
[55] N. Truong, et al., Privacy preservation in federated learning: An insightful survey from the gdpr perspective, Computers & Security 110 (2021) 102402.
[56] S. Ma, et al., The era of 1-bit llms: All large language models are in 1.58 bits, arXiv preprint arXiv:2402.17764 (2024).
[57] L. Lin, et al., Investigating bias in llm-based bias detection: Disparities between llms and human perception, arXiv preprint arXiv:2403.14896 (2024).
[58] K.-C. Yeh, J.-A. Chi, D.-C. Lian, S.-K. Hsieh, Evaluating interfaced llm bias, in: Proceedings of the 35th Conference on Computational Linguistics and Speech Processing (ROCLING 2023), 2023, pp. 292–299.
[59] S. G. Ayyamperumal, L. Ge, Current state of llm risks and ai guardrails, arXiv preprint arXiv:2406.12934 (2024).
[60] C. Wohlin, P. Runeson, et al., Experimentation in software engineering., Springer Science & Business Media, 2012.

SLR Literature

[SLR1] X. Feng, Y. Zhang, M. H. Meng, Y. Li, C. E. Joe, Z. Wang, G. Bai, Detecting contradictions from iot protocol specification documents based on neural generated knowledge graph, ISA Transactions (2023). doi:https://doi.org/10.1016/j.isatra.2023.04.025.
URL https://www.sciencedirect.com/science/article/pii/S0019057823001945
[SLR2] R. Qiu, X. Xue, M. Chen, J. Zheng, S. Jing, Y. Li, A fine-grained dynamic access control method for power iot based on kformer., Infocommunications Journal 14 (4) (2022) 79–85. doi:https://doi.org/10.36244/ICJ.2022.4.11.
[SLR3] X. Luo, Z. Yu, Z. Zhao, W. Zhao, J.-H. Wang, Effective short text classification via the fusion of hybrid features for iot social data, Digital Communications and Networks 8 (6) (2022) 942–954. doi:https://doi.org/10.1016/j.dcan.2022.09.015.
URL https://www.sciencedirect.com/science/article/pii/S2352864822001900
[SLR4] M. M., Predictive analytics based on digital twins, generative ai, and chatgpt, in: Proceedings of the 27th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2023, International Institute of Informatics and Cybernetics, 2023, pp. 168–174. doi:https://doi.org/10.54808/WMSCI2023.01.168.
[SLR5] R. Silva Barbon, A. T. Akabane, Towards transfer learning techniques—bert, distilbert, bertimbau, and distilbertimbau for automatic text classification from different languages: A case study, Sensors 22 (21) (2022) 8184. doi:10.3390/s22218184.
URL http://dx.doi.org/10.3390/s22218184
[SLR6] H. Zhu, P. Tiwari, A. Ghoneim, M. S. Hossain, A collaborative ai-enabled pretrained language model for aiot domain question answering, IEEE Transactions on Industrial Informatics 18 (5) (2022) 3387–3396. doi:10.1109/TII.2021.3097183.
[SLR7] H. Zhang, L. Zhu, L. Zhang, T. Dai, X. Feng, L. Zhang, K. Zhang, Y. Yan, Smart objects recommendation based on pre-training with attention and the thing–thing relationship in social internet of things, Future Generation Computer Systems 129 (2022) 347–357. doi:https://doi.org/10.1016/j.future.2021.11.006.
URL https://www.sciencedirect.com/science/article/pii/S0167739X21004350
[SLR8] K. Yu, L. Tan, S. Mumtaz, S. Al-Rubaye, A. Al-Dulaimi, A. K. Bashir, F. A. Khan, Securing critical infrastructures: Deep-learning-based threat detection in iiot, IEEE Communications Magazine 59 (10) (2021) 76–82. doi:10.1109/MCOM.101.2001126.
[SLR9] I. H. Çavdar, V. Feryad, Efficient design of energy disaggregation model with bert-nilm trained by adax optimization method for smart grid, Energies 14 (15) (2021). doi:10.3390/en14154649.
URL https://www.mdpi.com/1996-1073/14/15/4649
[SLR10] S. A. Hamad, D. H. Tran, Q. Z. Sheng, W. E. Zhang, Bertdeep-ware: A cross-architecture malware detection solution for iot systems, in: 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2021, pp. 927–934. doi:10.1109/TrustCom53373.2021.00130.
[SLR11] K. Zeng, I. Paik, Dynamic service recommendation using lightweight bert-based service embedding in edge computing, in: 2021 IEEE 14th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC), 2021, pp. 182–189. doi:10.1109/MCSoC51149.2021.00035.
[SLR12] A. Aljubairy, A. Alhazmi, W. E. Zhang, Q. Z. Sheng, D. H. Tran, Towards a deep learning-driven service discovery framework for the social internet of things: A context-aware approach, in: W. Zhang, L. Zou, Z. Maamar, L. Chen (Eds.), Web Information Systems Engineering – WISE 2021, Springer International Publishing, Cham, 2021, pp. 480–488.
[SLR13] K. Zeng, I. Paik, Semantic service clustering with lightweight bert-based service embedding using invocation sequences, IEEE Access 9 (2021) 54298–54309. doi:10.1109/ACCESS.2021.3069509.
[SLR14] A. Sriram, Y. Li, A. Hadaegh, Mining social media to understand user opinions on iot security and privacy, in: 2021 IEEE International Conference on Smart Computing (SMARTCOMP), 2021, pp. 252–257. doi:10.1109/SMARTCOMP52413.2021.00056.
[SLR15] S. Xu, W. Zhang, F. Zhang, Multi-granular bert: An interpretable model applicable to internet-of-thing devices, in: 2020 IEEE International Conference on Energy Internet (ICEI), 2020, pp. 134–139. doi:10.1109/ICEI49372.2020.00032.
[SLR16] L. P. Osco, E. L. d. Lemos, W. N. Gonçalves, A. P. M. Ramos, J. Marcato Junior, The potential of visual chatgpt for remote sensing, Remote Sensing 15 (13) (2023). doi:10.3390/rs15133232.
URL https://www.mdpi.com/2072-4292/15/13/3232
[SLR17] S. Silalahi, T. Ahmad, H. Studiawan, Transformer-based sentiment analysis for anomaly detection on drone forensic timeline, in: 2023 11th International Symposium on Digital Forensics and Security (ISDFS), 2023, pp. 1–6. doi:10.1109/ISDFS58141.2023.10131749.
[SLR18] F. Ventirozos, M. Jacobo-Romero, S. Clinch, R. Batista-Navarro, Interactive clustering of cooking recipe instructions: Towards the automatic detection of events involving kitchen devices, in: 2021 IEEE 15th International Conference on Semantic Computing (ICSC), 2021, pp. 341–346. doi:10.1109/ICSC50631.2021.00064.
[SLR19] C. F. Iglesias, R. Guo, P. Nucci, C. Miceli, M. Bolic, Automated extraction of iot critical objects from iot storylines, requirements and user stories via nlp, in: 2023 10th IEEE Swiss Conference on Data Science (SDS), 2023, pp. 104–107. doi:10.1109/SDS57534.2023.00022.
[SLR20] Y. Ren, J. Wang, H. Li, H. Zhu, L. Sun, Devicegpt: A generative pre-training transformer on the heterogenous graph for internet of things, in: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 1929–1933. doi:10.1145/3539618.3591972.
URL https://doi.org/10.1145/3539618.3591972
[SLR21] B. Breve, G. Cimino, G. Desolda, V. Deufemia, A. Elefante, On the user perception of security risks of tap rules: A user study, in: L. D. Spano, A. Schmidt, C. Santoro, S. Stumpf (Eds.), End-User Development, Springer Nature Switzerland, Cham, 2023, pp. 162–179.
[SLR22] J. Ye, L. Zhang, P. Lan, H. He, D. Yang, Z. Wu, Improved intelligent semantics based chinese sentence similarity computing for natural language processing in iot, in: B. Li, C. Li, M. Yang, Z. Yan, J. Zheng (Eds.), IoT as a Service, Springer International Publishing, Cham, 2021, pp. 234–246.
[SLR23] P. Ni, Y. Li, G. Li, V. Chang, Natural language understanding approaches based on joint task of intent detection and slot filling for

iot voice interaction, Neural Computing and Applications 32 (20) (2020). doi:10.1007/s00521-020-04805-x.
URL https://doi.org/10.1007/s00521-020-04805-x

[SLR24] A. Kotha, K. Manohar, V. U, Iaasi: a device based interoperability as a service for iomt devices, Journal of Ambient Intelligence and Humanized Computing (2023). doi:10.1007/s12652-023-04669-8.
URL https://doi.org/10.1007/s12652-023-04669-8

[SLR25] A. L. Alfeo, M. G. C. A. Cimino, G. Vaglini, Technological troubleshooting based on sentence embedding with deep transformers, Journal of Intelligent Manufacturing 32 (2021) 1699–1710. doi:10.1007/s10845-021-01797-w.
URL https://doi.org/10.1007/s12652-023-04669-8

[SLR26] H. Y. Chang, S. H. Mozafari, C. Chen, J. J. Clark, B. H. Meyer, W. J. Gross, Pipebert: High-throughput bert inference for arm big.little multi-core processors, Journal of Signal Processing Systems (2022). doi:10.1007/s11265-022-01814-y.

[SLR27] R. Teixeira, M. Antunes, D. Gomes, R. L. Aguiar, Comparison of semantic similarity models on constrained scenarios, Information Systems Frontiers (2022). doi:10.1007/s10796-022-10350-w.

[SLR28] S. M. Aldossary, Smart vehicles networks: Bert self-attention mechanisms for cyber-physical system security, International Journal of System Assurance Engineering and Management (2023). doi:10.1007/s13198-023-02065-1.

[SLR29] F. Alloatti, A. Bosca, L. Di Caro, F. Pieraccini, Diabetes and conversational agents: the aida project case study, Discover Artificial Intelligence 1 (1) (2021) 4.

[SLR30] Y. Yuan, X. Cai, A human–machine interaction scheme based on background knowledge in 6g-enabled iot environment, IEEE Internet of Things Journal 8 (20) (2021) 15292–15302. doi:10.1109/JIOT.2021.3050880.

[SLR31] K. Zeng, I. Paik, Semantic service clustering with lightweight bert-based service embedding using invocation sequences, IEEE Access 9 (2021) 54298–54309. doi:10.1109/ACCESS.2021.3069509.

[SLR32] B. Breve, G. Cimino, V. Deufemia, Identifying security and privacy violation rules in trigger-action iot platforms with nlp models, IEEE Internet of Things Journal 10 (6) (2023) 5607–5622. doi:10.1109/JIOT.2022.3222615.

[SLR33] H. Xu, L. Han, Q. Yang, M. Li, M. Srivastava, Penetrative ai: Making llms comprehend the physical world, in: Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications, 2024, pp. 1–7.

[SLR34] F. Gao, Z. Xiao, S. Chen, R. Yu, X. Li, Medgcn: An iot-edge thrombus graph convolutional network for accurate prediction and prescription diagnosis of vascular occlusive diseases from unstructured clinical reports, Computer Communications 214 (2024) 123–135.

[SLR35] N. Zhong, Y. Wang, R. Xiong, Y. Zheng, Y. Li, M. Ouyang, D. Shen, X. Zhu, Casit: Collective intelligent agent system for internet of things, IEEE Internet of Things Journal (2024).

[SLR36] M. Guastalla, Y. Li, A. Hekmati, B. Krishnamachari, Application of large language models to ddos attack detection, in: International Conference on Security and Privacy in Cyber-Physical Systems and Smart Vehicles, Springer, 2023, pp. 83–99.

[SLR37] Y. Wang, Z. Wang, W. Wang, Q. Chen, K. Huang, A. Nguyen, S. De, Zero-shot medical information retrieval via knowledge graph embedding, in: International Workshop on Internet of Things of Big Data for Healthcare, Springer, 2023, pp. 29–40.

[SLR38] M. A. Ferrag, M. Ndhlovu, N. Tihanyi, L. C. Cordeiro, M. Debbah, T. Lestable, N. S. Thandi, Revolutionizing cyber threat detection with large language models: A privacy-preserving bert-based lightweight model for iot/iiot devices, IEEE Access (2024).

[SLR39] J. Lu, M. Chen, Zero-sample face retrieval combining large language model and visual base model for iot, Internet Technology Letters (2024) e506.

[SLR40] A. Berenguer, A. Morejón, D. Tomás, J.-N. Mazón, Using large language models to enhance the reusability of sensor data, Sensors 24 (2) (2024) 347.

[SLR41] F. Li, J. Huang, Y. Gao, W. Dong, Chatiot: Zero-code generation of trigger-action based iot programs with chatgpt, in: Proceedings of the 7th Asia-Pacific Workshop on Networking, 2023, pp. 219–220.

[SLR42] P. M. Sooriya Patabandige, S. A. O. Waskito, K. Li, K. J. Leow, S. Chakrabarty, A. Varshney, Poster: Rethinking embedded sensor data processing and analysis with large language models, in: Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services, 2023, pp. 561–562.

[SLR43] F. Jaafar, D. Ameyed, L. Titare, M. Nematullah, Iot phishing detection using hybrid nlp and machine learning models enhanced with contextual embedding, in: 2023 IEEE 23rd International Conference on Software Quality, Reliability, and Security Companion (QRS-C), IEEE, 2023, pp. 340–349.

[SLR44] N. Nascimento, P. Alencar, D. Cowan, Gpt-in-the-loop: Supporting adaptation in multiagent systems, in: 2023 IEEE International Conference on Big Data (BigData), IEEE, 2023, pp. 4674–4683.

[SLR45] N. Petrović, S. Koničanin, S. Suljović, Chatgpt in iot systems: Arduino case studies, in: 2023 IEEE 33rd International Conference on Microelectronics (MIEL), IEEE, 2023, pp. 1–4.

[SLR46] S. Jha, A. Roy, A. Cobb, A. Berenbeim, N. D. Bastian, Challenges and opportunities in neuro-symbolic composition of foundation models, in: MILCOM 2023-2023 IEEE Military Communications Conference (MILCOM), IEEE, 2023, pp. 156–161.

[SLR47] M.-Y. Su, K.-L. Su, Bert-based approaches to identifying malicious urls, Sensors 23 (20) (2023) 8499.

[SLR48] M. A. Ferrag, M. Debbah, M. Al-Hawawreh, Generative ai for cyber threat-hunting in 6g-enabled iot networks, in: 2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW), IEEE, 2023, pp. 16–25.

[SLR49] J. Gao, Y. Zhang, Y. Chen, T. Zhang, B. Tang, X. Wang, Unsupervised human activity recognition via large language models and iterative evolution, in: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2024, pp. 91–95.

[SLR50] D. A. Hafeth, G. Lal, M. Al-Khafajiy, T. Baker, S. Kollias, Cloud-iot application for scene understanding in assisted living: Unleashing the potential of image captioning and large language model (chatgpt), in: 2023 16th International Conference on Developments in eSystems Engineering (DeSE), IEEE, 2023, pp. 150–155.

[SLR51] Y. Njah, A. Leivadeas, J. Violos, M. Falkner, Toward intent-based network automation for smart environments: A healthcare 4.0 use case, IEEE Access 11 (2023) 136565–136576.

[SLR52] B. Zhao, W. Jin, J. Del Ser, G. Yang, Chatagri: Exploring potentials of chatgpt on cross-linguistic agricultural text classification, Neurocomputing 557 (2023) 126708.

[SLR53] M. Both, B. Kämper, A. Cartus, J. Beermann, T. Fessler, J. Müller, C. Diedrich, Automated monitoring applications for existing buildings through natural language processing based semantic mapping of operational data and creation of digital twins, Energy and Buildings 300 (2023) 113635.

[SLR54] H. Xu, P. Zhou, R. Tan, M. Li, G. Shen, Limu-bert: Unleashing the potential of unlabeled data for imu sensing applications, in: Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, 2021, pp. 220–233.

[SLR55] N. Alkhatib, M. Mushtaq, H. Ghauch, J.-L. Danger, Can-bert do it? controller area network intrusion detection system based on bert language model, in: 2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA), IEEE, 2022, pp. 1–8.