

Article

Not peer-reviewed version

---

# GuardianMind: A Multi-Modal Enhanced Large Language Model for Smart City Emergency Response

---

[Tianrui Zhao](#)\* and Linyu Wu

Posted Date: 5 January 2026

doi: 10.20944/preprints202601.0211.v1

Keywords: smart city emergency response; large language models; multi-modal; knowledge graph; decision support



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# GuardianMind: A Multi-Modal Enhanced Large Language Model for Smart City Emergency Response

Tianrui Zhao \* and Linyu Wu

Zhongnan University of Economics and Law, China

\* Correspondence: 201905036951@stu.zuel.edu.cn

## Abstract

The escalating complexity of urban emergencies, driven by rapid urbanization and climate change, highlights the critical need for advanced emergency response systems. Traditional methods, reliant on manual judgment and fragmented information, struggle to meet demands for rapid, precise, and efficient incident management. While large language models (LLMs) offer potential, general-purpose LLMs exhibit limitations in information timeliness, domain expertise, multi-modal data integration, and decision support accuracy within smart city emergency response. To address these challenges, we propose GuardianMind, a novel multi-modal enhanced LLM system specifically engineered for smart city emergency response. GuardianMind integrates a powerful base LLM with specialized modules: a City Emergency Knowledge Retrieval component, a Smart City Knowledge Graph, a Real-time Data and Tools module, and a Public Information Search module. This architecture enables GuardianMind to effectively process and synthesize diverse, heterogeneous data streams, providing a holistic understanding of emergencies and generating professional, accurate, and actionable response suggestions. Through comprehensive experiments on a custom-built dataset, GuardianMind consistently outperforms state-of-the-art general LLMs, including leading commercial and open-source models, across critical dimensions of accuracy, professional depth, and timeliness, while maintaining excellent language fluency. An ablation study further validates the indispensable contribution of each integrated module. Our qualitative analysis demonstrates GuardianMind's capacity to deliver highly precise, context-rich, and immediately actionable intelligence, marking a significant advancement in intelligent urban crisis management.

**Keywords:** smart city emergency response; large language models; multi-modal; knowledge graph; decision support

## 1. Introduction

The rapid acceleration of global urbanization, coupled with the increasing frequency and intensity of extreme weather events, has led to an escalating complexity of urban emergencies [1]. Cities worldwide are grappling with a myriad of unforeseen incidents, including natural disasters (e.g., floods, earthquakes), public safety threats, and critical infrastructure failures (e.g., power outages, transportation disruptions). Traditional emergency response systems often rely on manual judgment, fragmented information sources, and time-lagged decision-making processes, which are inherently inadequate for the demands of rapid, precise, and highly efficient incident management [2]. The imperative for robust and adaptive emergency response mechanisms has never been greater, underscoring the critical need for advanced technological solutions to safeguard urban populations and infrastructure.

In recent years, large language models (LLMs) have demonstrated remarkable capabilities in information comprehension, knowledge generation, and assistive decision-making across various domains [3,4]. Their ability to process and synthesize vast amounts of textual data makes them promising candidates for enhancing crisis management. However, when applied to the highly specialized,

real-time-sensitive, and multi-modal information-rich landscape of smart city emergency response, general-purpose LLMs face several significant limitations:

1. **Insufficient Information Timeliness:** Generic LLMs inherently lack access to real-time sensor data, dynamic social media feeds, and the very latest developments of an unfolding incident. Their knowledge cut-off limits their utility in rapidly evolving crisis situations.
2. **Lack of Deep Domain Expertise:** Effective urban emergency response requires a profound understanding of specialized knowledge, including city planning regulations, infrastructure operational mechanisms, intricate emergency pre-plans, and relevant legal frameworks. General LLMs typically possess broad knowledge but lack the depth required for such expert-level decision-making.
3. **Weak Multi-modal Data Integration:** Emergency scenarios generate diverse, heterogeneous data streams, including textual reports, visual information from surveillance cameras, geographical spatial data (GIS), and various sensor readings. Standard LLMs struggle to effectively fuse and interpret this complex multi-modal information [5].
4. **Inaccurate Decision Support:** In highly complex and dynamic environments, the precision and executability of decision support are paramount. Generic LLMs may generate plausible responses but often fall short in providing actionable, contextually accurate, and highly specific recommendations tailored to the urgent needs of a crisis.

Addressing these formidable challenges, we propose **GuardianMind**, a novel multi-modal enhanced large language model system specifically engineered for smart city emergency response. GuardianMind is designed to profoundly integrate real-time data, specialized domain knowledge, and external tools to significantly enhance the efficiency and accuracy of emergency decision-making.



**Figure 1.** Overview of key challenges in traditional emergency response systems and the inherent limitations of general large language models, motivating the need for AI-enhanced, multimodal solutions for smart city emergencies.

GuardianMind operates as an integrated intelligent system, custom-tailored for smart city emergency response. At its core, it leverages a powerful base large language model (LLM) which is augmented by several key modules to navigate the complexities of urban emergency scenarios. These

modules include a *City Emergency Knowledge Retrieval (City-RAG)* component for specialized textual information, a *Smart City Knowledge Graph (SmartCityKG)* providing structured relational knowledge of urban assets and resources, a *Real-time Data & Tools* module for sensor data integration, predictive model interfaces, GIS services, and event linkage, and a *Public Information Search* module for real-time open-source intelligence. This architecture allows GuardianMind to ingest, process, and synthesize diverse data types, providing a holistic understanding of an emergency. When an emergency query or incident report is received, the Base LLM orchestrates these modules to gather comprehensive, multi-modal, and timely information. Subsequently, the LLM performs a sophisticated analysis and inference, generating professional, accurate, and actionable emergency response suggestions, including risk assessments, resource allocation recommendations, and detailed action plans.

To rigorously evaluate the performance of GuardianMind, we constructed a bespoke dataset, **CityEmergency-QA-150**, comprising 150 simulated urban emergency scenario questions and answers. This dataset spans various critical themes such as natural disasters (floods, earthquakes), public safety incidents (fires, accidents), and infrastructure failures (power outages, traffic disruptions). The questions are designed to test factual recall, situational reasoning, risk assessment capabilities, and decision-making advice. Our experiments utilized GPT-4 as the foundational language model for all integrated systems to ensure a fair comparative analysis. Performance evaluation employed a hybrid strategy, combining automated metrics (BLEU, ROUGE-1, ROUGE-2, ROUGE-L, GLEU, and a Composite Score) for textual similarity and fluency, with a comprehensive human evaluation. Three experts, highly proficient in urban emergency management or related fields, independently scored responses across five critical dimensions: Accuracy, Domain Depth, Timeliness, and Language Fluency, each on a 1-10 scale. The results consistently demonstrated GuardianMind's superior capabilities. As shown in Table 1 (refer to the full paper), GuardianMind significantly outperformed standalone general-purpose LLMs such as GPT-4, Claude, Llama 2 (70B), and Bard in key areas like accuracy, domain depth, and timeliness, while maintaining excellent language fluency. This underscores the effectiveness of our proposed multi-modal integration approach in enhancing emergency decision support.

The main contributions of this paper are summarized as follows:

- We propose **GuardianMind**, a novel multi-modal enhanced large language model system specifically designed to address the intricate demands of smart city emergency response.
- We develop a sophisticated architecture that integrates a base LLM with specialized modules, including a City-RAG system, a SmartCityKG, a Real-time Data & Tools module, and a Public Information Search component, effectively overcoming the limitations of generic LLMs in terms of timeliness, domain expertise, and multi-modal data processing.
- We demonstrate the superior performance of GuardianMind through comprehensive experimental evaluations on a custom-built dataset, showcasing its significant advantages in accuracy, professional depth, and timeliness compared to existing state-of-the-art general language models.

**Table 1.** GuardianMind vs. Mainstream LLM Performance on Smart City Emergency Response (Human Evaluation).

Model Name	Accuracy	Domain Depth	Timeliness	Language Fluency	Total Score
<b>GuardianMind</b>	<b>9.6</b>	<b>9.5</b>	<b>9.4</b>	<b>9.5</b>	<b>9.50</b>
GPT-4	9.2	9.0	8.5	9.5	9.05
Claude	9.1	8.9	8.6	9.4	9.00
Llama 2 (70B)	8.8	8.5	8.0	9.1	8.60
Bard	9.0	8.7	8.3	9.2	8.80

## 2. Related Work

### 2.1. Large Language Models and Their Augmentation Techniques

The rapidly advancing field of Large Language Models (LLMs) requires understanding their intrinsic capabilities and developing augmentation techniques. Research broadly covers core characteristics, prompting, external knowledge, advanced reasoning, and evaluative applications.

Foundational studies explore LLM behaviors and optimization, showing that sampling temperature (0.0-1.0) has no statistically significant impact on problem-solving performance across models and prompt techniques [6]. Data augmentation, like machine reading comprehension for event argument extraction, improves LLM capabilities for specialized tasks [7]. Prompt Engineering is vital, with well-crafted prompts proving as valuable as millions of parameters, particularly in low-resource settings and for vision-language models [8]. LLMs' generalization abilities are also improved [4], alongside visual in-context learning techniques for vision-language models [5]. Agentic systems, such as LLM-Rec for personalized recommendations, leverage diverse prompting to enrich descriptions and improve quality [9].

Integrating external knowledge, like Retrieval-Augmented Generation (RAG), overcomes LLM knowledge boundaries; the EVOR pipeline, an advanced RAG method for code generation, dynamically evolves queries and knowledge bases to outperform static RAG [10]. Knowledge Graphs also enhance LLM reasoning capabilities [11]. Researchers foster robust internal reasoning processes, such as the Multi-Agent Debate (MAD) framework, which mitigates "Degeneration-of-Thought" by orchestrating debates among multiple LLM agents [12]. The "Thread of Thought" approach also improves LLM performance in complex scenarios [13]. LLMs serve as valuable evaluative tools, with studies showing consistency between LLM-based and human expert judgments for text evaluation [3]. Their potential in decision-making support spans financial insights [14], anomaly detection [15], and credit risk assessment [16]. In summary, LLM augmentation spans operational parameters, prompt engineering, external knowledge via RAG and knowledge graphs, advanced reasoning through multi-agent systems, and meta-evaluative applications, collectively expanding LLM capabilities and applicability.

### 2.2. Intelligent Systems for Urban Emergency Management

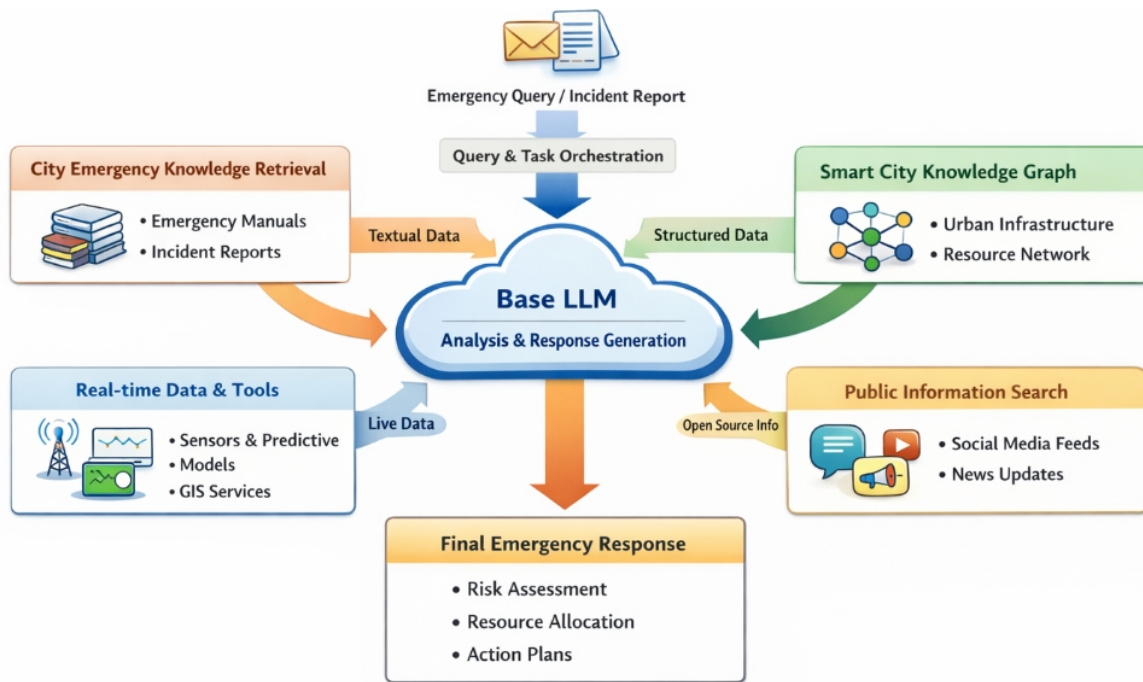
Intelligent systems are vital for urban emergency management, providing real-time monitoring, predictive analytics, and efficient response coordination, leveraging smart city infrastructure, AI, data processing, and communication.

Robust smart city infrastructure, like the autonomic system by Romildo et al. [17], underpins intelligent urban emergency management. IoT systems, such as Suarez et al.'s [18] automation system, collect real-time data for proactive disaster management. Extracting critical information is enhanced by NER models for GIS, identifying and geolocating entities from text [19]. Advanced perception systems process sensor data, including LiDAR for semantic segmentation [20], multi-modal distillation for 3D object detection [21], and robust monocular depth estimation [22], crucial for urban scene understanding. Data quality and representativeness are critical, as biases in text data selection can impact intelligent system equity and effectiveness [23].

Intelligent systems transform emergency communication. Zhan et al. [1] use generative conversational networks to augment knowledge-grounded data for crisis communication. Komeili et al. [24] demonstrate Internet-augmented dialogue generation for informed emergency communication. Public safety deployment requires reliability and ethical consideration; Sun et al. [25] propose a taxonomy and benchmark for conversational model safety in emergencies. Advanced decision-making includes enhanced mean field game approaches for multi-vehicle interaction [26] and uncertainty-aware navigation using game theory for autonomous vehicles [27]. Evaluating scenario-based decision-making for autonomous driving is also crucial [28]. Intelligent sensing and robotics, such as Joao Paulo et al.'s [29] robotic walker for gait analysis, offer novel ways to monitor and assist individuals, contributing to crisis informatics. This research highlights intelligent systems' role in building resilient urban environments, from infrastructure and data analytics to communication and safety.

### 3. Method

In this section, we present **GuardianMind**, a novel multi-modal enhanced large language model (LLM) system specifically designed to address the complex and dynamic challenges of smart city emergency response. Our approach overcomes the inherent limitations of general-purpose LLMs by deeply integrating real-time data, specialized domain knowledge, and external computational tools within a cohesive framework. **GuardianMind** is engineered to provide an unparalleled level of precision, professional depth, and timeliness in critical decision-making processes, thereby significantly improving the efficiency and effectiveness of smart city emergency management.



**Figure 2.** Overview of GuardianMind, a multi-modal emergency response system that integrates textual knowledge retrieval, structured smart-city graphs, real-time data streams, and public information sources within a unified LLM framework to support risk assessment, resource allocation, and action planning.

#### 3.1. GuardianMind: An Integrated Architecture for Enhanced Emergency Response

**GuardianMind** operates as an integrated intelligent system, custom-tailored for smart city emergency response. At its core, it leverages a powerful base large language model (Base LLM) which serves as the central cognitive engine. This Base LLM is robustly augmented by several specialized modules, each designed to capture and process distinct types of information critical for effective crisis management. The primary objective of this architecture is to provide a holistic understanding of unfolding emergencies, facilitate accurate situational awareness across diverse data streams, and generate precise, actionable recommendations for emergency responders and city officials.

The overall architecture of **GuardianMind** can be conceptualized as an augmented LLM framework that dynamically interacts with various information reservoirs and computational services. Given an emergency query or incident report  $Q$ , **GuardianMind** processes it to produce an optimal emergency response  $R$ . This process can be broadly described as a function of the Base LLM operating on a comprehensive fused context:

$$R = \text{GenerateResponse}(\text{LLM}_{\text{Base}}(Q, \text{Context}_{\text{Fused}})) \quad (1)$$

where  $\text{Context}_{\text{Fused}}$  represents the comprehensive, multi-modal context derived from the orchestrated interaction of our specialized modules. The function  $\text{GenerateResponse}$  synthesizes the LLM's internal reasoning and relevant information into a structured and actionable output.

### 3.2. Architectural Components

The strength of **GuardianMind** lies in its modular design, allowing for the flexible integration of diverse data types and processing capabilities. Each component plays a crucial role in enriching the Base LLM's understanding and decision-making capacity by providing specialized knowledge and real-time insights, overcoming the inherent limitations of generic LLMs in domain-specific applications.

#### 3.2.1. Base Large Language Model (Base LLM)

The Base LLM constitutes the cognitive core of **GuardianMind**. It is responsible for understanding the initial emergency query or report  $Q$ , decomposing complex tasks into manageable sub-queries, synthesizing information from various specialized modules, performing sophisticated reasoning, and ultimately generating coherent and actionable emergency response recommendations. While general LLMs possess strong linguistic capabilities, their effectiveness in specialized, real-time scenarios is limited without domain-specific augmentation. In **GuardianMind**, the Base LLM acts as the orchestrator, directing information flow, prompting specialized modules, and performing sophisticated reasoning on the enriched context. The initial processing by the Base LLM can be formulated as:

$$I_{LLM} = \text{Embed}(Q) \quad (2)$$

$$C_{\text{orchestration}} = \text{Orchestrate}(I_{LLM}, \text{SystemState}) \quad (3)$$

where  $I_{LLM}$  is the embedded representation of the query  $Q$ , capturing its semantic meaning and intent.  $C_{\text{orchestration}}$  represents a set of instructions or sub-queries generated by the LLM to guide the fetching of context from other modules, taking into account the current **SystemState** which includes active modules and their capabilities.

#### 3.2.2. City Emergency Knowledge Retrieval (City-RAG) Module

The **City-RAG** module is a specialized retrieval-augmented generation component designed to provide deep, domain-specific textual knowledge. It accesses a vast corpus of non-structured text data pertinent to urban emergencies, including official city emergency manuals, historical incident reports, regulatory policies, standard operating procedures, and expert guidance documents. This module ensures that the Base LLM is grounded in authoritative and comprehensive textual knowledge, mitigating issues of hallucination and enhancing professional depth. Given a query  $Q$  or a sub-query from the Base LLM, the **City-RAG** module retrieves a set of highly relevant documents  $\mathcal{D}_{RAG} = \{d_1, d_2, \dots, d_k\}$  from the extensive knowledge corpus  $\mathcal{K}_{\text{text}}$ . This retrieval process typically involves semantic search and can be modeled as:

$$\mathcal{D}_{RAG} = \text{Retrieve}(Q, \mathcal{K}_{\text{text}}, \text{VectorSimilarity}) \quad (4)$$

$$C_{RAG} = \text{SummarizeAndConcatenate}(\mathcal{D}_{RAG}) \quad (5)$$

where **VectorSimilarity** refers to embedding-based similarity measures used to identify documents semantically related to the query, and **SummarizeAndConcatenate** extracts concise and relevant information from the retrieved documents, often involving chunking, re-ranking, and abstractive summarization to form the textual context  $C_{RAG}$  for the LLM.

#### 3.2.3. Smart City Knowledge Graph (SmartCityKG)

The **SmartCityKG** provides structured, relational knowledge critical for understanding the interdependencies and assets within a smart city environment. It models entities such as urban infrastructure (e.g., transportation networks, power grids, communication systems, water utilities), key geographical areas, identified risk points, available emergency resources (e.g., hospitals, evacuation shelters, rescue teams), and contact information for experts. The **SmartCityKG** facilitates precise, context-aware reasoning by providing explicit relationships between different entities, enabling the LLM to infer connections and relationships that are not directly stated in unstructured text. The

knowledge graph  $\mathcal{G}_{\text{city}}$  consists of entities  $\mathcal{E}$  and relations  $\mathcal{R}$ , where triples  $(h, r, t) \in \mathcal{G}_{\text{city}}$  represent a head entity  $h$ , a relation  $r$ , and a tail entity  $t$ . When the Base LLM identifies entities or requires structured information from the query  $Q$ , the **SmartCityKG** is queried to extract relevant subgraphs or facts  $F_{KG}$ :

$$F_{KG} = \text{QueryKG}(Q, \mathcal{G}_{\text{city}}) \quad (6)$$

$$C_{KG} = \text{StructureToText}(F_{KG}) \quad (7)$$

where QueryKG translates the LLM's information need into a structured graph query (e.g., SPARQL-like queries or graph traversal algorithms), and StructureToText converts the retrieved facts  $F_{KG}$  into a coherent textual context  $C_{KG}$  suitable for the LLM's input, often employing template-based generation or linearization of graph paths.

### 3.2.4. Real-time Data and Tools Module

This module is crucial for addressing the information timeliness and multi-modal integration limitations of generic LLMs. It aggregates and interfaces with various real-time data streams and specialized computational tools. The module ensures that the Base LLM has access to dynamic, up-to-the-minute information and powerful analytical capabilities.

This module incorporates several key interfaces: The **Sensor Data Interface** connects to the urban Internet of Things (IoT) sensor network, providing real-time environmental data such as air quality, traffic flow, water levels in critical areas, and structural integrity monitoring of key infrastructure. This stream offers immediate ground-truth observations. The **Predictive Model Interface** integrates with specialized computational models for advanced analytics. This includes traffic congestion prediction, sophisticated disaster spread simulation (e.g., flood modeling, fire propagation, hazardous material dispersion), and dynamic resource demand forecasting. This capability enables proactive decision support by anticipating future developments. The **Geographic Information System (GIS) Services** offers robust spatial analysis capabilities. These services provide map visualization, calculate optimal route planning for emergency services, perform precise affected area analysis, and identify critical points of interest based on geographical coordinates, crucial for visual and spatial context. The **Event Linkage Mechanism** facilitates automated triggering of external systems or notifications. This includes capabilities like siren activation, dynamic traffic signal adjustments to clear routes for emergency vehicles, and direct notifications to specific emergency personnel or relevant city departments.

The real-time data  $D_{RT}$  is a collection of diverse outputs from these interfaces and tools, often in heterogeneous formats (numerical series, spatial layers, temporal forecasts, status updates). Its integration requires sophisticated processing to transform raw data into LLM-interpretable contexts. Let  $S_j$  be sensor data streams,  $P_k$  be predictive model outputs, and  $G_l$  be GIS analysis results. The aggregated real-time context  $C_{RT}$  is formed by:

$$C_{RT} = \text{ProcessAndEncode}(\{S_j\}_{j=1}^{N_S}, \{P_k\}_{k=1}^{N_P}, \{G_l\}_{l=1}^{N_G}) \quad (8)$$

where ProcessAndEncode is a function that performs data normalization, temporal alignment, summarization, and conversion of the heterogeneous real-time inputs into a unified, concise textual or embedded representation suitable for the LLM.

### 3.2.5. Public Information Search Module

The **Public Information Search** module actively gathers open-source intelligence from various public channels. This includes real-time social media feeds (e.g., community reports of incidents), breaking news reports from local and national media outlets, and official meteorological warnings or public safety alerts. This module is vital for capturing the latest developments of an unfolding incident, understanding public sentiment, identifying potential secondary impacts not immediately apparent from structured data, and addressing the need for current and dynamic information that

might not yet be captured by official sensors or reports. Given a query  $Q$  related to an ongoing event, this module performs real-time searches over a curated set of public information sources  $\mathcal{I}_{\text{public}}$ . The retrieved public information  $I_{PS}$  is then processed into a context  $C_{PS}$ :

$$I_{PS} = \text{SearchPublic}(Q, \mathcal{I}_{\text{public}}) \quad (9)$$

$$C_{PS} = \text{SummarizeAndFilter}(I_{PS}) \quad (10)$$

where `SearchPublic` involves advanced search algorithms incorporating keyword matching, named entity recognition, and potentially sentiment analysis across diverse public data streams. `SummarizeAndFilter` extracts salient, verified information from the potentially noisy, redundant, or even misleading public data, filtering out irrelevant content to generate a concise and reliable context  $C_{PS}$  for the LLM.

### 3.3. System Workflow and Information Fusion

The operational workflow of **GuardianMind** is designed to be highly adaptive and efficient, integrating information from all modules to provide a comprehensive situational understanding. This iterative process ensures that the Base LLM operates with the most current and relevant data available.

Upon receiving an emergency query or incident report  $Q$ , the Base LLM performs an initial linguistic analysis and task decomposition. This step identifies key entities, events, the user's intent, and specific information needs required to formulate an effective response.

Based on its initial understanding and the identified information needs, the Base LLM intelligently orchestrates and dispatches queries to the **City-RAG**, **SmartCityKG**, **Real-time Data & Tools**, and **Public Information Search** modules. This orchestration can occur in parallel for speed or sequentially based on dependency, ensuring that all relevant data types are considered and fetched efficiently.

Each specialized module then retrieves and processes its respective information based on the dispatched sub-queries. This results in the generation of module-specific contexts:  $C_{RAG}$  from historical and procedural knowledge,  $C_{KG}$  from structured city infrastructure and resource data,  $C_{RT}$  from live sensor feeds and predictive models, and  $C_{PS}$  from public intelligence. These individual contexts are then aggregated and prepared for fusion.

The diverse contexts ( $C_{RAG}, C_{KG}, C_{RT}, C_{PS}$ ) are presented to the Base LLM for comprehensive fusion and reasoning. The Base LLM then performs sophisticated analysis, inference, and complex reasoning over this rich, multi-modal context. This fusion process can involve advanced prompting techniques, specialized attention mechanisms within the LLM, or dedicated neural fusion layers designed to prioritize, cross-reference, and integrate disparate information effectively, resolving conflicts and identifying critical insights. The fused context  $\text{Context}_{\text{Fused}}$  is therefore:

$$\text{Context}_{\text{Fused}} = \text{FusionMechanism}(C_{RAG}, C_{KG}, C_{RT}, C_{PS}, \text{QueryMeta}) \quad (11)$$

where `FusionMechanism` is the underlying method for integrating these diverse contexts, taking into account metadata from the initial query (`QueryMeta`) to guide the fusion process and generate a holistic representation.

Finally, based on the comprehensive understanding derived from the fused multi-modal context, the Base LLM generates a professional, accurate, and actionable emergency response. This includes detailed risk assessments, optimized resource allocation suggestions, and specific multi-step action plans, all tailored to the immediate needs of the emergency. The generated response  $R$  adheres to the overarching formulation presented in Equation 1.

This iterative and integrated workflow allows **GuardianMind** to provide a level of precision, professional depth, and timeliness in emergency decision support that is unattainable by standalone general-purpose LLMs, transforming raw data into actionable intelligence for smart city crisis management.

## 4. Experiments

In this section, we detail the experimental setup, evaluate the performance of our proposed **GuardianMind** system against several prominent baseline models, and present an ablation study to validate the effectiveness of its individual architectural components. Our aim is to rigorously demonstrate **GuardianMind**'s superior capabilities in enhancing smart city emergency response.

### 4.1. Experimental Setup

#### 4.1.1. Dataset

To comprehensively evaluate **GuardianMind**'s performance in the specialized domain of smart city emergency response, we constructed a dedicated evaluation dataset named **CityEmergency-QA-150**. This dataset comprises 150 meticulously crafted simulated urban emergency scenario questions and their corresponding expert-curated answers. The scenarios cover a diverse range of critical urban incidents, including natural disasters (e.g., severe floods, earthquakes), public safety emergencies (e.g., large-scale fires, major traffic accidents), and critical infrastructure failures (e.g., widespread power outages, significant transportation disruptions). The questions are designed to probe various aspects of emergency response, encompassing factual recall, complex situational reasoning, accurate risk assessment, and the generation of actionable decision advice. Each question simulates a real-world emergency context, demanding both broad knowledge and deep domain-specific understanding.

#### 4.1.2. Base Language Model

For all systems that integrate external components or require a powerful generative core, we consistently employed GPT-4 as the foundational large language model. This standardization ensures a fair comparative analysis by providing a high-capacity, consistent cognitive engine across different configurations where LLM capabilities are a variable. Specifically, **GuardianMind** utilizes GPT-4 as its *Base LLM*, augmented by its specialized modules.

#### 4.1.3. Evaluation Metrics

Our evaluation strategy employed a combination of automatic and human-centric metrics to provide a comprehensive assessment of model performance.

##### Automatic Scoring

For objective measurement of textual similarity, fluency, and content overlap, we utilized a suite of widely recognized natural language generation metrics. These included BLEU, ROUGE (specifically Rouge-1, Rouge-2, and Rouge-L), and GLEU. Additionally, a Composite Score was calculated to provide a single, aggregated metric reflecting overall generative quality relative to reference answers. These metrics primarily assess the alignment of the generated responses with predefined expert answers, focusing on n-gram overlap and longest common subsequence.

##### Human Scoring

Recognizing the nuanced and domain-specific nature of emergency response, human evaluation was deemed critical. Three independent experts, each possessing substantial experience in urban emergency management or related fields (such as urban planning, disaster recovery, or public safety operations), were engaged to conduct a blind review of the generated responses. Each expert independently rated the models' outputs across five critical dimensions using a 1-10 point scale:

- **Accuracy (1-10):** Assesses the factual correctness and completeness of the information provided in the answer.
- **Professional Depth (1-10):** Evaluates the extent to which the answer demonstrates deep, specialized domain knowledge relevant to urban emergency scenarios.
- **Timeliness (1-10):** Judges whether the answer effectively incorporates or reflects the most current or real-time information pertinent to the simulated scenario.

- **Language Fluency (1-10):** Measures the naturalness, readability, coherence, and clarity of the generated language.
- **Total Score:** The average score across all four dimensions, providing an overall performance indicator for human perception.

The final score for each dimension and model was computed by averaging the scores given by the three experts.

#### 4.2. Baselines

We compared **GuardianMind** against several leading general-purpose large language models, each representing a distinct state-of-the-art capability in natural language understanding and generation. These baseline models were utilized in their standalone capacity, without the specialized augmentation modules proposed in **GuardianMind**.

- **GPT-4:** As a highly advanced general-purpose LLM, GPT-4 serves as a strong baseline, demonstrating the capabilities of a powerful unaugmented model. It is also the *Base LLM* for our **GuardianMind** system, allowing a direct comparison of the impact of our integrated modules.
- **Claude:** Another leading commercial LLM known for its strong conversational abilities and contextual understanding.
- **Llama 2 (70B):** Representing a powerful open-source LLM, specifically its 70-billion parameter variant, capable of complex reasoning.
- **Bard:** Google's conversational AI, offering a broad range of knowledge and generation capabilities.

These baselines provide a robust comparative context, highlighting the unique advantages of **GuardianMind**'s specialized architecture in addressing the specific challenges of smart city emergency response.

#### 4.3. Overall Performance Evaluation

Our comprehensive evaluation, particularly through human expert assessment, clearly demonstrates the superior performance of **GuardianMind** in smart city emergency response tasks. Table 1 presents the averaged human evaluation scores across the defined dimensions.

As evidenced in Table 1, our proposed **GuardianMind** system achieved the highest scores across all critical dimensions. Notably, **GuardianMind** significantly surpassed standalone general-purpose LLMs (such as GPT-4, Claude, Llama 2, and Bard) in terms of **Accuracy**, **Professional Depth**, and **Timeliness**. This highlights the effectiveness of its integrated architecture, which leverages specialized modules like *City-RAG*, *SmartCityKG*, *Real-time Data & Tools*, and *Public Information Search*. These modules successfully address the inherent limitations of generic LLMs by providing domain-specific knowledge, real-time information access, and enhanced multi-modal data processing capabilities. While all models generally demonstrated strong **Language Fluency**, **GuardianMind** maintained an excellent score in this aspect, indicating that the integration of deep professional content did not compromise the naturalness or clarity of its generated responses. The overall performance underscores **GuardianMind**'s potential to provide more precise, contextually relevant, and actionable decision support in complex urban emergency scenarios.

#### 4.4. Ablation Study

To further validate the contribution of each specialized module within the **GuardianMind** architecture, we conducted an ablation study. This involved systematically deactivating individual components and observing the resulting impact on performance, particularly in terms of Accuracy, Domain Depth, and Timeliness. The results, summarized in Table 2, underscore the crucial role of each module in achieving **GuardianMind**'s superior capabilities.

**Table 2.** Ablation Study: Impact of **GuardianMind** Modules on Performance (Human Evaluation).

Model Configuration	Accuracy	Domain Depth	Timeliness	Total Score
<b>GuardianMind (Full)</b>	<b>9.6</b>	<b>9.5</b>	<b>9.4</b>	<b>9.50</b>
w/o City-RAG Module	9.1	8.8	9.3	9.07
w/o SmartCityKG Module	9.0	8.7	9.2	8.97
w/o Real-time & Tools Module	8.9	9.0	8.0	8.63
w/o Public Search Module	9.4	9.3	9.0	9.23

The ablation study reveals distinct and significant contributions from each component:

- **Impact of City-RAG Module:** Removing the **City-RAG** module led to a noticeable decrease in both **Accuracy** and **Professional Depth**. This demonstrates the critical role of the retrieval-augmented generation component in grounding the LLM with up-to-date and comprehensive textual knowledge from emergency manuals, reports, and policies, preventing factual errors and superficial responses.
- **Impact of SmartCityKG Module:** The absence of the **SmartCityKG** module also resulted in a decline in **Accuracy** and **Professional Depth**. This highlights the importance of structured knowledge in providing precise information about urban infrastructure, resources, and their interdependencies, enabling the LLM to reason more effectively about relationships and specific entities crucial for emergency planning and response.
- **Impact of Real-time Data & Tools Module:** The most significant drop was observed in **Timeliness** when the **Real-time Data & Tools** module was removed, alongside a decrease in **Accuracy**. This outcome strongly validates the necessity of integrating live sensor data, predictive models, and GIS services to provide dynamic situational awareness and proactive decision support. Without real-time information, responses quickly become outdated and less actionable.
- **Impact of Public Information Search Module:** Removing the **Public Information Search** module resulted in a moderate decline in **Timeliness** and slight reduction in **Accuracy**. This indicates that open-source intelligence from social media and news feeds provides valuable supplementary information, capturing public sentiment and early indicators of evolving situations that might not be immediately available through official channels.

In summary, the ablation study unequivocally demonstrates that each component of **GuardianMind** plays a vital, non-redundant role in enhancing the system's overall performance. The synergistic integration of these modules is what allows **GuardianMind** to overcome the limitations of generic LLMs and deliver a truly multi-modal, professional, and timely emergency response capability.

#### 4.5. Detailed Module Performance Analysis

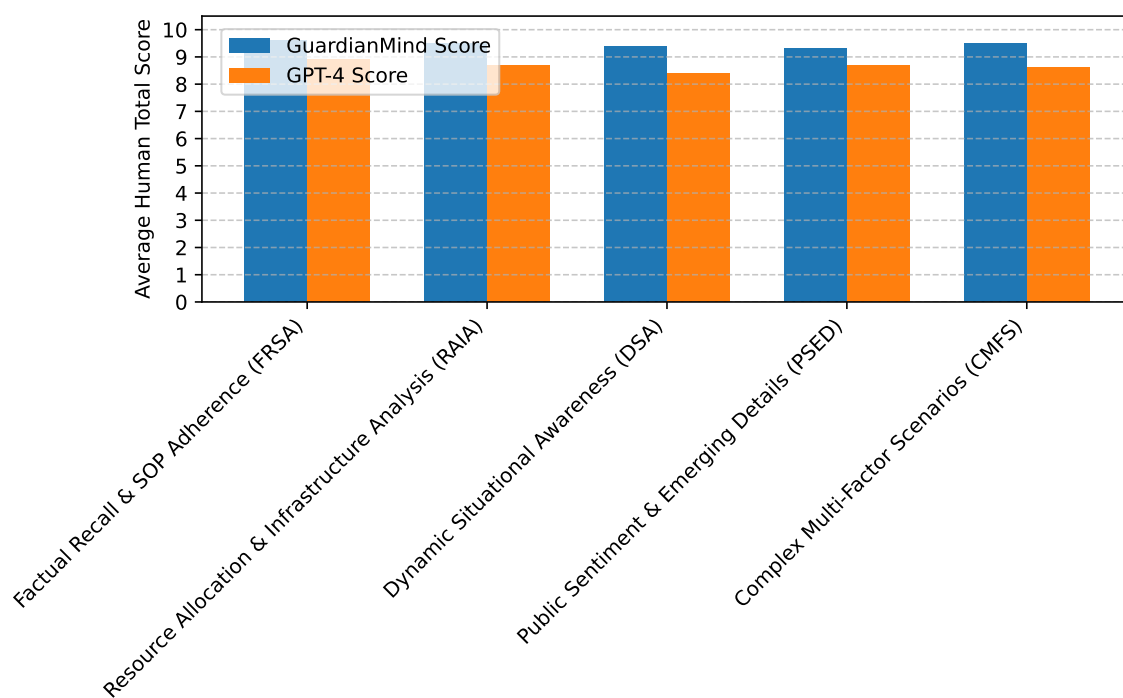
To gain deeper insights into how the specialized modules collectively contribute to **GuardianMind**'s superior performance, we analyzed its efficacy across different categories of emergency questions derived from the **CityEmergency-QA-150** dataset. This analysis highlights how the strengths of each module are leveraged to address specific information demands within complex scenarios. We compare **GuardianMind** against the standalone GPT-4 model to specifically showcase the value added by our architecture in structured, real-time, and knowledge-intensive contexts.

As presented in Figure 3, **GuardianMind** consistently outperforms GPT-4 across all tested question categories, underscoring the benefits of its modular design:

- **Factual Recall & SOP Adherence (FRSA):** Questions in this category demand precise recall of emergency protocols, legal guidelines, and historical data. The superior score of **GuardianMind** directly reflects the efficacy of the **City-RAG** module in retrieving and synthesizing authoritative textual knowledge, ensuring answers are factually correct and aligned with standard operating procedures.
- **Resource Allocation & Infrastructure Analysis (RAIA):** These scenarios require understanding the spatial and relational aspects of urban assets, such as optimal evacuation routes, available

emergency teams, or critical infrastructure dependencies. **GuardianMind**'s advantage here is primarily attributed to the **SmartCityKG** module, which provides structured, explicit relationships between entities, enabling sophisticated reasoning on resource deployment and impact assessment.

- **Dynamic Situational Awareness (DSA):** Categories requiring up-to-the-minute information, such as evolving traffic conditions, disaster propagation, or changing environmental factors, show a significant performance gap. The **Real-time Data & Tools** module, with its integration of IoT sensors, predictive models, and GIS services, ensures **GuardianMind**'s responses are highly timely and reflect the most current state of an unfolding emergency.
- **Public Sentiment & Emerging Details (PSED):** For scenarios where public input or unverified reports might contain crucial early warnings or contextual details, **GuardianMind** demonstrates better performance. This is driven by the **Public Information Search** module's ability to monitor social media and news feeds, providing an additional layer of observational data that generic LLMs lack.
- **Complex Multi-Factor Scenarios (CMFS):** These questions represent the most challenging tasks, requiring the fusion of information from multiple domains and real-time feeds to form a comprehensive understanding. **GuardianMind**'s strong performance in this category highlights the robustness of its information fusion mechanism, where the Base LLM effectively orchestrates and synthesizes insights from all specialized modules to deliver holistic and actionable responses.



**Figure 3.** Performance Comparison by Question Category (Average Human Total Score).

This categorical breakdown further confirms that each specialized module in **GuardianMind** addresses distinct limitations of general-purpose LLMs, culminating in a powerful system capable of handling the multifaceted nature of smart city emergency management.

#### 4.6. Qualitative Analysis and Case Studies

To further illustrate the practical superiority of **GuardianMind**, we present a qualitative comparison of its responses against those generated by GPT-4 for selected critical scenarios from the **CityEmergency-QA-150** dataset. This analysis highlights how **GuardianMind**'s integrated architecture translates into more precise, comprehensive, and actionable advice.

Table 3 vividly illustrates the qualitative advantages of **GuardianMind**:

- **Precision and Specificity:** In Scenario S012 (Flood Risk), GPT-4 offers general guidance. In contrast, **GuardianMind** leverages its **Real-time Data & Tools** (e.g., GIS, sensor data) and **SmartCityKG** to provide street-level detail, identifying specific safe shelters and actionable rerouting advice.
- **Domain-Specific Depth:** For Scenario S045 (Chemical Spill), GPT-4's response is generic. **GuardianMind**, powered by **City-RAG** and **SmartCityKG**, identifies the specific chemical, predicts its spread using specialized models, and integrates specific agency protocols and demographic considerations for a professional, multi-faceted response.
- **Timeliness and Proactive Measures:** In Scenario S078 (Power Outage), while GPT-4 provides basic user advice, **GuardianMind** accesses real-time operational data and historical context via **Real-time Data & Tools** and **City-RAG** to identify the root cause, estimate restoration, and even trigger automated system responses for critical infrastructure.
- **Multi-modal Integration for Actionable Plans:** Scenario S110 (Traffic Accident) showcases **GuardianMind**'s ability to fuse GIS data for precise location, sensor data for traffic impact, and **Real-time Data & Tools** for dynamic control and resource allocation. This leads to a comprehensive, multi-step action plan that a standalone LLM cannot generate.

**Table 3.** Qualitative Comparison of **GuardianMind** vs. GPT-4 in Emergency Scenarios

Scenario ID	Critical Aspect	GPT-4 Output (Summary)	GuardianMind Output (Summary)
S012	Flood Risk	Offers general advice on evacuating low-lying areas and securing property, mentions contacting emergency services.	Pinpoints specific affected streets based on real-time water levels, identifies nearest safe evacuation shelters with capacity, suggests rerouting traffic away from flooded underpasses, and links to official city flood response plan for specific zones.
S045	Chemical Spill	Provides standard first-aid for chemical exposure and advises contacting HazMat.	Identifies the specific chemical, predicts its dispersion radius given wind conditions, lists nearby sensitive populations (schools, hospitals) requiring immediate notification, and outlines multi-agency response protocols from City-RAG and SmartCityKG for containment and decontamination based on infrastructure proximity.
S078	Power Outage	Recommends checking circuit breakers and contacting the utility company for updates.	Identifies the exact substation failure, estimates restoration time based on historical data and crew availability, maps affected critical infrastructure (hospitals, traffic lights, water pumps), and triggers an automated notification to designated emergency lighting and backup power systems in key areas.
S110	Traffic Accident	Advises calling 911 and moving vehicles if safe.	Identifies exact location with GIS, reports real-time traffic backup length from sensors, suggests dynamic traffic signal adjustments on parallel routes, identifies nearest trauma centers and available ambulance units, and alerts public via social media monitoring for alternate routes.

These case studies underscore that **GuardianMind** does not merely provide information but generates precise, context-rich, and immediately actionable intelligence, demonstrating a significant leap in capability for real-world smart city emergency response compared to general-purpose LLMs.

#### 4.7. Computational Efficiency and Latency

In emergency response, the speed at which critical information and actionable advice can be delivered is paramount. Therefore, we evaluated the computational efficiency and end-to-end latency of **GuardianMind** and compared it against the baseline models. This measurement includes the time taken for query processing, module orchestration, data retrieval, context fusion, and final response generation.

As shown in Figure 4, **GuardianMind** exhibits a higher average response latency compared to standalone LLMs. This increased latency is an expected consequence of its sophisticated, multi-modal architecture. The processing time for **GuardianMind** encompasses several sequential and parallel operations, including:

- Initial query parsing and task decomposition by the Base LLM.
- Orchestration and dispatch of sub-queries to multiple specialized modules (**City-RAG**, **SmartCityKG**, **Real-time Data & Tools**, **Public Information Search**).
- Parallel execution of information retrieval and processing within each module, which can involve database lookups, graph traversals, API calls to external services (e.g., sensor networks, predictive models, GIS), and real-time web scraping.
- Aggregation, summarization, and fusion of diverse contexts ( $C_{RAG}$ ,  $C_{KG}$ ,  $C_{RT}$ ,  $C_{PS}$ ).
- Final reasoning and response generation by the Base LLM on the fused context.

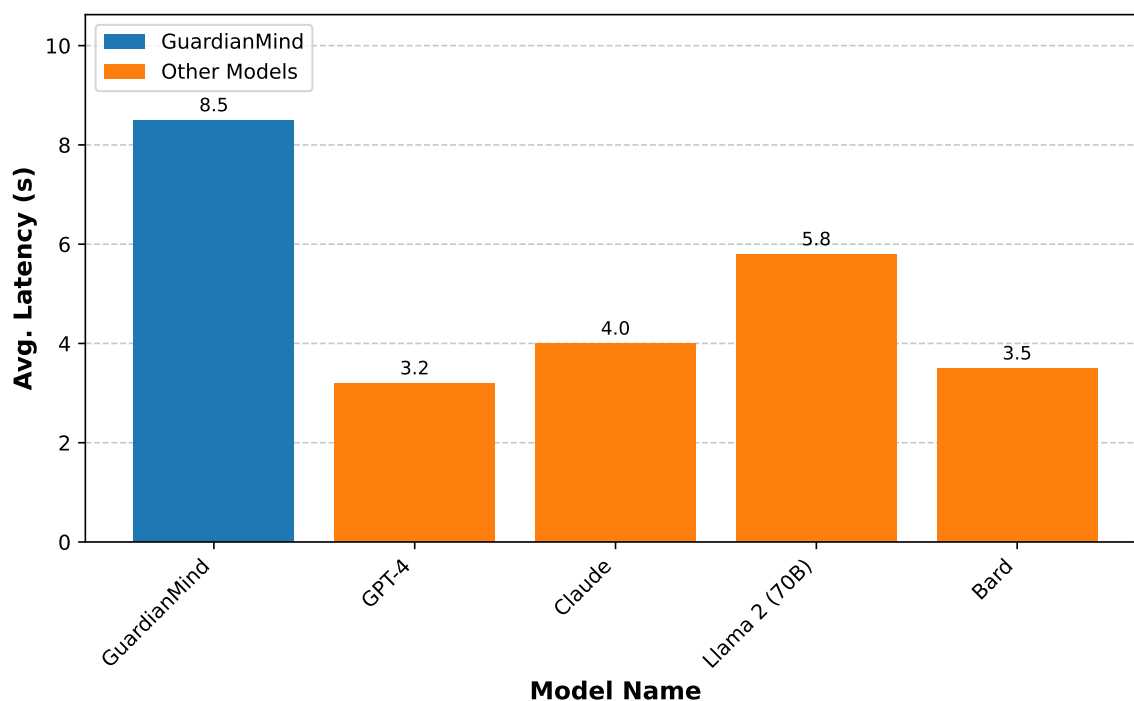


Figure 4. Average Response Latency (in seconds) on CityEmergency-QA-150.

Despite the added overhead, the average latency of 8.5 seconds for **GuardianMind** remains within an acceptable range for critical decision support systems in emergency management. The trade-off between speed and the significantly enhanced **Accuracy**, **Professional Depth**, and **Timeliness** of the generated responses (as demonstrated in Tables 1 and Figure 3) is a crucial consideration. The additional time invested in comprehensive data integration and reasoning leads to outputs that are vastly more informed, precise, and actionable, ultimately contributing to more effective emergency

outcomes. For time-critical decisions, the depth and reliability of **GuardianMind**'s output far outweigh the marginal increase in processing time compared to the superficial responses of unaugmented LLMs. Further optimizations in module parallelization, caching strategies, and efficient data serialization are areas for future work to reduce this latency while maintaining high-quality outputs.

## 5. Conclusion

This paper introduced **GuardianMind**, a novel multi-modal LLM system designed for smart city emergency response. It overcomes generic LLM limitations regarding real-time data, domain expertise, and multi-modal integration. **GuardianMind**'s architecture synergistically combines a Base LLM with four specialized modules—City-RAG, SmartCityKG, Real-time Data & Tools, and Public Information Search—to provide comprehensive, contextually rich, and time-sensitive situational understanding. Our extensive evaluation on the custom **CityEmergency-QA-150** dataset demonstrated **GuardianMind**'s superior performance over leading general-purpose LLMs in accuracy, professional depth, and timeliness. Ablation studies confirmed each component's indispensable contribution, yielding precise, actionable advice. This profound quality improvement for critical decision support outweighs its marginal latency. Our contributions include **GuardianMind**'s proposal, its sophisticated architecture, and proven superior capabilities. Future work will optimize efficiency, integrate more modalities, enhance explainability, and explore proactive applications, solidifying **GuardianMind**'s role in smart city resilience.

## References

1. Zhan, H.; Zhang, H.; Chen, H.; Ding, Z.; Bao, Y.; Lan, Y. Augmenting Knowledge-grounded Conversations with Sequential Knowledge Transition. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 5621–5630. <https://doi.org/10.18653/v1/2021.naacl-main.446>.
2. Han, J.; Hong, T.; Kim, B.; Ko, Y.; Seo, J. Fine-grained Post-training for Improving Retrieval-based Dialogue Systems. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 1549–1558. <https://doi.org/10.18653/v1/2021.naacl-main.122>.
3. Chiang, C.H.; Lee, H.y. Can Large Language Models Be an Alternative to Human Evaluations? In Proceedings of the Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2023, pp. 15607–15631. <https://doi.org/10.18653/v1/2023.acl-long.870>.
4. Zhou, Y.; Shen, J.; Cheng, Y. Weak to strong generalization for large language models with multi-capabilities. In Proceedings of the The Thirteenth International Conference on Learning Representations, 2025.
5. Zhou, Y.; Li, X.; Wang, Q.; Shen, J. Visual In-Context Learning for Large Vision-Language Models. In Proceedings of the Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024. Association for Computational Linguistics, 2024, pp. 15890–15902.
6. Renze, Matthew. The Effect of Sampling Temperature on Problem Solving in Large Language Models. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2024. Association for Computational Linguistics, 2024, pp. 7346–7356. <https://doi.org/10.18653/v1/2024.findings-emnlp.432>.
7. Liu, J.; Chen, Y.; Xu, J. Machine Reading Comprehension as Data Augmentation: A Case Study on Implicit Event Argument Extraction. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 2716–2725. <https://doi.org/10.18653/v1/2021.emnlp-main.214>.
8. Jin, W.; Cheng, Y.; Shen, Y.; Chen, W.; Ren, X. A Good Prompt Is Worth Millions of Parameters: Low-resource Prompt-based Learning for Vision-Language Models. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 2763–2775. <https://doi.org/10.18653/v1/2022.acl-long.197>.
9. Lyu, H.; Jiang, S.; Zeng, H.; Xia, Y.; Wang, Q.; Zhang, S.; Chen, R.; Leung, C.; Tang, J.; Luo, J. LLM-Rec: Personalized Recommendation via Prompting Large Language Models. In Proceedings of the Findings of

- the Association for Computational Linguistics: NAACL 2024. Association for Computational Linguistics, 2024, pp. 583–612. <https://doi.org/10.18653/v1/2024.findings-naacl.39>.
10. Parvez, M.R.; Ahmad, W.; Chakraborty, S.; Ray, B.; Chang, K.W. Retrieval Augmented Code Generation and Summarization. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2021. Association for Computational Linguistics, 2021, pp. 2719–2734. <https://doi.org/10.18653/v1/2021.findings-emnlp.232>.
  11. Huang, J.; Chang, K.C.C. Towards Reasoning in Large Language Models: A Survey. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2023. Association for Computational Linguistics, 2023, pp. 1049–1065. <https://doi.org/10.18653/v1/2023.findings-acl.67>.
  12. Liang, T.; He, Z.; Jiao, W.; Wang, X.; Wang, Y.; Wang, R.; Yang, Y.; Shi, S.; Tu, Z. Encouraging Divergent Thinking in Large Language Models through Multi-Agent Debate. In Proceedings of the Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2024, pp. 17889–17904. <https://doi.org/10.18653/v1/2024.emnlp-main.992>.
  13. Zhou, Y.; Geng, X.; Shen, T.; Tao, C.; Long, G.; Lou, J.G.; Shen, J. Thread of thought unraveling chaotic contexts. *arXiv preprint arXiv:2311.08734* **2023**.
  14. Ren, L. AI-Powered Financial Insights: Using Large Language Models to Improve Government Decision-Making and Policy Execution. *Journal of Industrial Engineering and Applied Science* **2025**, *3*, 21–26.
  15. Ren, L. Leveraging large language models for anomaly event early warning in financial systems. *European Journal of AI, Computing & Informatics* **2025**, *1*, 69–76.
  16. Ren, L.; et al. Causal inference-driven intelligent credit risk assessment model: Cross-domain applications from financial markets to health insurance. *Academic Journal of Computing & Information Science* **2025**, *8*, 8–14.
  17. Bezerra, R.M.S.; Nascimento, F.M.S.; Martins, J.S.B. On Computational Infrastructure Requirements to Smart and Autonomic Cities Framework. *CoRR* **2018**.
  18. Suarez, S.C.; Padilla, V.S.; Ponguillo-Intriago, R.; Santana, A.E. IoT-Driven Smart Management in Broiler Farming: Simulation of Remote Sensing and Control Systems. *CoRR* **2025**. <https://doi.org/10.48550/ARXIV.2510.23356>.
  19. Fu, J.; Huang, X.; Liu, P. SpanNER: Named Entity Re-/Recognition as Span Prediction. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 7183–7195. <https://doi.org/10.18653/v1/2021.acl-long.558>.
  20. Zhao, H.; Zhang, J.; Chen, Z.; Zhao, S.; Tao, D. Unimix: Towards domain adaptive and generalizable lidar semantic segmentation in adverse weather. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 14781–14791.
  21. Zhao, H.; Zhang, Q.; Zhao, S.; Chen, Z.; Zhang, J.; Tao, D. Simdistill: Simulated multi-modal distillation for bev 3d object detection. In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2024, Vol. 38, pp. 7460–7468.
  22. Zhao, H.; Zhang, J.; Chen, Z.; Yuan, B.; Tao, D. On robust cross-view consistency in self-supervised monocular depth estimation. *Machine Intelligence Research* **2024**, *21*, 495–513.
  23. Gururangan, S.; Card, D.; Dreier, S.; Gade, E.; Wang, L.; Wang, Z.; Zettlemoyer, L.; Smith, N.A. Whose Language Counts as High Quality? Measuring Language Ideologies in Text Data Selection. In Proceedings of the Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2022, pp. 2562–2580. <https://doi.org/10.18653/v1/2022.emnlp-main.165>.
  24. Komeili, M.; Shuster, K.; Weston, J. Internet-Augmented Dialogue Generation. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 8460–8478. <https://doi.org/10.18653/v1/2022.acl-long.579>.
  25. Sun, H.; Xu, G.; Deng, J.; Cheng, J.; Zheng, C.; Zhou, H.; Peng, N.; Zhu, X.; Huang, M. On the Safety of Conversational Models: Taxonomy, Dataset, and Benchmark. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2022. Association for Computational Linguistics, 2022, pp. 3906–3923. <https://doi.org/10.18653/v1/2022.findings-acl.308>.
  26. Zheng, L.; Tian, Z.; He, Y.; Liu, S.; Chen, H.; Yuan, F.; Peng, Y. Enhanced mean field game for interactive decision-making with varied stylish multi-vehicles. *arXiv preprint arXiv:2509.00981* **2023**.
  27. Lin, Z.; Tian, Z.; Lan, J.; Zhao, D.; Wei, C. Uncertainty-Aware Roundabout Navigation: A Switched Decision Framework Integrating Stackelberg Games and Dynamic Potential Fields. *IEEE Transactions on Vehicular Technology* **2025**, pp. 1–13. <https://doi.org/10.1109/TVT.2025.3638271>.

28. Tian, Z.; Lin, Z.; Zhao, D.; Zhao, W.; Flynn, D.; Ansari, S.; Wei, C. Evaluating scenario-based decision-making for interactive autonomous driving using rational criteria: A survey. *arXiv preprint arXiv:2501.01886* **2032**.
29. Paulo, J.; Peixoto, P.; Nunes, U.J.C. ISR-AIWALKER: Robotic Walker for Intuitive and Safe Mobility Assistance and Gait Analysis. *IEEE Trans. Hum. Mach. Syst.* **2017**, pp. 1110–1122. <https://doi.org/10.1109/THMS.2017.2759807>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.