Australasian Transport Research Forum 2024 Proceedings
27-29 November, Melbourne, Australia
Publication website: https://australasiantransportresearchforum.org.au/

Incident reporting for autonomous shuttles via LLMs: Eglinton case study

Xiangrui Kong[1], Thomas Bräunl[1]

[1] The University of Western Australia, School of Engineering

Email for correspondence: xiangrui.kong@research.uwa.edu.au

Abstract

The rise of autonomous driving signifies a critical advancement in the evolution of transportation systems, integrating smoothly into daily routines due to its numerous benefits over traditional vehicles, yet it also introduces complex challenges that contribute to accidents and injuries on an annual basis. The existing incident reporting system primarily uses forms and descriptive content, offering a relatively small amount of information for analysis. In contrast, autonomous driving systems produce extensive data from sensors, loggers, and CAN bus, highlighting the need for an advanced reporting and recording approach to accommodate this richer, more detailed information. Large language models (LLMs) have shown exceptional abilities in understanding, reasoning, and summarizing information. We propose a multi-agent LLM-based automatic traffic log generation system. This system generates traffic reports by allowing LLMs to access formatted log data packages from autonomous vehicles while restricting access to sensitive traffic data. Using nearly a year's worth of service data from autonomous shuttle buses in WA, Elginton as a test case, our framework demonstrates the capability to reduce log generation time and capture more granular traffic details.

1 Introduction

According to the California Department of Motor Vehicles, more than 732 incident reports caused by autonomous driving vehicles in California have been reported as of August 12, 2024 (California DMV, 2023). The current incident reporting system relies heavily on forms and descriptive content, providing only a limited amount of information for analysis. In contrast, autonomous driving systems generate vast amounts of data from sensors, loggers, and CAN bus, underscoring the necessity for a more sophisticated reporting and recording system that can effectively handle and utilize this richer, more detailed information.

In recent years, Large Language Models (LLMs) have emerged as a powerful tool in various domains, including traffic control and transportation. LLMs have been applied to a wide range of tasks, such as accident forecasting (de Zarzà et al., 2023), traffic signal control (Lai et al., 2024), and autonomous driving (Cui et al., 2023). These applications demonstrate the potential of LLMs in enhancing safety, efficiency, and decision-making processes within the transportation sector.

However, despite the promising results, the adoption of LLMs in real-world transportation systems has been limited due to several challenges. One major concern is the inherent uncertainty associated with LLMs, which can lead to unexpected or unreliable outputs . Additionally, the lack of extensive real-world applications and case studies has hindered the widespread implementation of LLMs in traffic control and transportation (Greshake et al., 2023). Furthermore, safety issues and privacy concerns have been raised regarding the use of LLMs, particularly in the context of autonomous vehicles and the handling of sensitive data.

The combination of rich, multi-modal traffic data generated by autonomous shuttles and the advanced natural language processing capabilities of LLMs presents a significant opportunity for enhancing safety and efficiency in transportation. By leveraging LLMs to analyze and

interpret the complex data streams from autonomous shuttles, it becomes possible to generate automated safety reports, detect anomalies, and provide real-time recommendations for traffic management.

To the best of our knowledge, this study represents the first attempt to combine rosbag, a commonly used format for storing and analyzing autonomous shuttle data, with LLM-based safety reporter generators. By integrating these two powerful tools, we aim to demonstrate the feasibility and potential benefits of utilizing LLMs in the context of autonomous shuttle traffic management. This research explores the challenges and opportunities associated with this novel approach, paving the way for future advancements in intelligent transportation systems.

2 Related works

A LLM such ChatGPT (OpenAI et al., 2024) and Claude is a type of artificial intelligence model that is trained on vast amounts of text data to understand and generate human-like language. These models use deep learning techniques, such as transformer architectures, to learn patterns and relationships within the training data. These models have revolutionized Natural Language Processing and have been applied in numerous fields, such as chatbots, content creation, and research. Recently LLMs showed their potential in modern transportation such as autonomous driving, traffic control, and urban journey planning.

2.1 LLM in modern transportation

Language models can leverage their broad knowledge not just for language-related tasks, but also to make informed decisions and take actions in dynamic, interactive environments that require goal-oriented reasoning (Mao et al., 2023). LanguageMPC is a system that employs large language models to predict the future dynamics of vehicles. It utilizes a bird's-eye view (BEV) representation to understand complex interactive situations, such as those encountered in roundabouts or other multi-vehicle scenarios. The system takes into account the current state of the vehicles involved, including their positions, velocities, and other relevant parameters, to generate accurate forecasts of their future trajectories and interactions (Sha et al., 2023). The Agent-Driver method introduces a framework that uses LLMs to handle various types of driving-related information. This framework enables LLMs to process and interpret data such as images, point clouds, driving rules, and maps through function calls. By employing a chain-of-thought approach, the LLM can systematically analyze and reason over the diverse input data, facilitating comprehensive decision-making for autonomous driving tasks (Mao et al., 2023). The DriveLLM method integrates rule-based driving methods with LLMs, implementing the LLM for campus driving scenarios, and demonstrates high real-time performance within a stable network, evidenced by the efficient token processing time in GPT-3.5 (Cui et al., 2023).

While existing applications of LLMs primarily focus on agent and prompt design, their potential in urban applications is starting to be explored (Yan et al., 2024). Recent studies have emphasized the impact of LLMs on urban planning (Zhou et al., 2024) and their ability to process urban data effectively. In the domain of human mobility, LLMs have demonstrated their predictive strength (Huang et al., 2022), particularly in forecasting mobility patterns (Su et al., 2024), including during public events (Greshake et al., 2023). This highlights their potential to contribute to urban transportation planning and management. Moreover, LLMs have found applications in various aspects of transportation. They have been used to enhance traffic forecasting, automate accident report generation (Zheng et al., 2024). These diverse use cases underscore the wide applicability and potential of LLMs in urban transportation.

2.2 Cloud-based LLM safety

As the size of both models and data continues to grow, generative LLMs are demonstrating a promising ability to understand and integrate classification tasks into their generative pipelines (Colin, 2020). However, this growth has also brought safety issues related to LLMs into the

spotlight (Tjondronegoro et al., 2022). While Differential Privacy (DP) (Dwork and Roth, 2014) offers a theoretical worst-case privacy guarantee for safeguarded data, current privacy mechanisms significantly reduce the utility of LLMs, rendering many existing approaches impractical.

Recent research has identified three primary safety concerns in the realm of LLMs: prompt injection, data breaches, and model hallucinations. Prompt injection emerges as a significant security risk, where specifically crafted inputs are used to manipulate or exploit the natural language processing capabilities of AI systems. Furthermore, LLMs are vulnerable to unintentional data breaches, in which sensitive information may be leaked through model outputs, often due to the inclusion of confidential datasets during the training phase(Namer et al., 2023). Another critical issue identified in these models is their propensity for hallucination, where they generate erroneous or illogical information with a false sense of confidence, stemming from limitations in their predictive text generation algorithms (Martino, Iannelli and Truong, 2023). These findings highlight the necessity for improved security measures and algorithmic refinements in the development and deployment of LLMs to mitigate these risks.

The alignment of LLMs with human and organizational values presents a crucial area of research in the rapidly evolving field of artificial intelligence, requiring a multi-faceted approach to ensure ethical and effective AI deployment (Wolf et al., 2023). In current research on LLMs, two primary methods influence the alignment of output text. Firstly, the training data of the LLM plays a significant role in shaping the nature of the generated content and its alignment (Ouyang, 2022). Secondly, LLM service providers offer optional API alignment services designed to filter out content that drastically deviates from predefined norms or standards (Smith, 2023). Moreover, LLM customers often tailor alignment requirements to their specific needs, typically employing simpler methods such as Retrieval-Augmented Generation (RAG) (Lewis, 2020) or customized prompting techniques.

3 Methods

In this section we provide details of our method. As depicted in Figure 2, It consists of six main modules that work together to process user requests, retrieve relevant data, apply necessary safeguards, and compose the final traffic report. The user, typically a Traffic Manager, initiates the request for a traffic report by providing the necessary details such as the location (Eglinton, WA) and date (April 14, 2023). The system processes the request through the Cloud-LLM and Request Decomposition modules, applies the necessary guardrails and sensitive rails, retrieves the required data through the Functional Rails, and composes the final traffic report.

3.1 Request decomposition

This reporter utilizes a cloud based LLM to process and understand the user request for a traffic report. It interprets the natural language input and extracts key information such as the location, date, and any specific details requested. Once the user's request is understood by the Cloud-LLM, the request decomposition module breaks down the request into smaller, manageable sub-tasks. It identifies the necessary data points and APIs required to fulfill the request effectively.

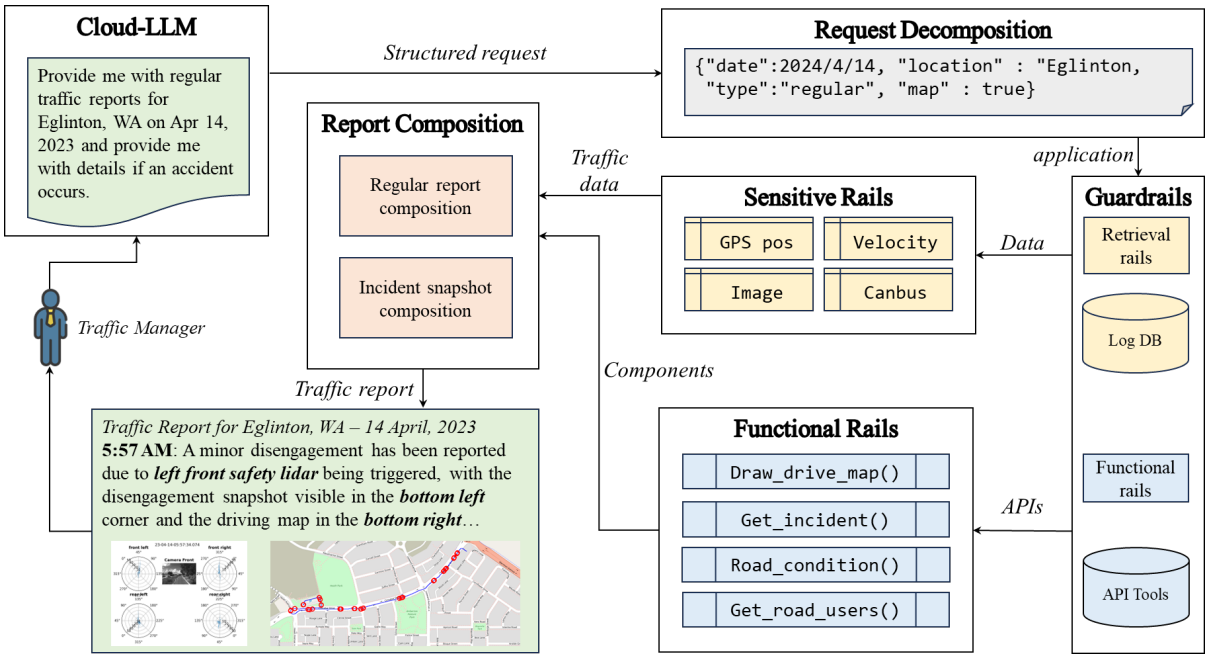
We design a system prompt to process user requests. The prompted LLM decomposes user requests into structured requests including report format including regular report and incident report. Regular report is to record the shuttle bus driving path in the map while the incident report is to capture all the sensor data and environment data to help users get to the situation.

3.2 Guardrails

We assume the cloud based LLM is not safe, and we don't want LLM to access any sensitive data including traffic, images, lidar, and precise location. The guardrails module ensures that the system operates within predefined boundaries and adheres to any legal or safety

constraints. It prevents the generation of inappropriate or potentially harmful content in the traffic report.

Figure 1: Overview of our ATSR reporter



We decompose the report components and drive log data as two databases. We want to build a guardrail between non-local LLM and our private databases. The retrieval rails designed for retrieval requested traffic data while the functional rails designed for breaking down customized traffic reports without providing any sensitive data. For content generation, the LLM uses placeholders for sensitive information retrieval.

The sensitive rails module handles sensitive information and applies necessary filters or anonymization techniques to protect privacy and comply with data protection regulations. It ensures that any personally identifiable information or sensitive data is handled appropriately. The functional rails module focuses on retrieving the required data and executing the necessary functions to generate the traffic report. It interacts with various data APIs, such as traffic data, GPS data, and incident reports, to gather the relevant information.

Table 1: Sensitive data

Components	Specification
Image	Front or rear camera captures
Safety Lidar	Four safety lidars for collision avoidance
CAN bus	Vehicle self status including velocity and steering angle
GPS location	Precise GPS location

Table 2: Report API tools

APIs	Specification
Get map	Get the traffic map of the specified road from map database
Get messages	Retrieve a specific topics for a given duration
Get data frame	Retrieve all topics at a given time
Draw snapshot	Visualize sensor data at a specified moment
Draw path	Visualize driving trajectories for a specified period of time

3.3 Report composition

Finally, the Report Composition module takes the processed data and composes the traffic report. It structures the information into two main components: The regular report component provides a general overview of the traffic conditions for the specified location and date. It includes details such as traffic flow, average speed, and any notable congestion points. If an accident or incident is detected, the incident snapshot component generates a detailed report of the event. It includes information such as the location of the incident, type of incident, involved parties, and any available updates or resolution status.

4 Experiments

We present a one-year autonomous shuttle bus service experiment conducted in Eglinton, Western Australia. As shown in Figure 3, the vehicle is equipped with a comprehensive set of sensors to ensure safe operation and accurate localization. The sensor suite includes front-facing and rear-facing cameras, detection radars, four safety radars, and a localization radar. Additionally, the vehicle employs a Global Positioning System (GPS) and an Inertial Measurement Unit (IMU) for attitude estimation and localization (Quirke-Brown et al., 2023).

Figure 2: nUWAy autonomous shuttle bus

As depicted in Figure 3 and Table 3, the autonomous shuttle bus service experiment follows a route that connects the Stockland Sales Center (Sales Office) to Amberton Beach and back, covering a total distance of approximately 4.1 kilometers including a 1 km stretch along Cinnabar Drive, followed by sections of Viridian Boulevard and Celeste Street, each measuring roughly 200m. The route also incorporates a segment of approximately 170 m within a seaside parking lot. On the return journey from the seaside to the sales center, the route further comprises an 800 m section along Mulberry Way and Indigo Street. The maximum allowable speed along the route is 50 kilometers per hour, which is equivalent to 16.7 meters per second. This diverse route composition allows for the evaluation of the autonomous shuttle bus's performance under different road conditions and environments.

The varying road types and lengths along the route present unique challenges for the autonomous shuttle bus, testing its ability to navigate and adapt to changing surroundings. The inclusion of both urban roads and a seaside parking lot enables the assessment of the vehicle's performance in distinct scenarios, such as handling intersections, pedestrian interactions, and navigating open spaces.

Figure 3: Proposed bus routes

The autonomous vehicle trials were conducted over a period of twelve months from April 2023 to April 2024 as shown in Figure 4. The trials took place on various dates and times, with durations ranging from one to seven hours per day. In total, nine trial days were reported, accumulating approximately 17 hours and 23 minutes of autonomous vehicle operation. The trials were conducted on a pre-defined route in Eglinton, Western Australia, covering segments of Cinnabar Drive, Viridian Boulevard, Celeste Street, Mulberry Way, and Indigo Street. Throughout the trials, the autonomous vehicle navigated through residential areas and a seaside parking lot. The paths taken by the vehicle during each trial were recorded and visualized using mapping software.

Table 3: Shuttle service route features

Criteria	Specification
Route length	East-West: 2km. West-East route: 2.1km
Destination	Amberton Beach Bar and Sales Centre
Lane widths	> 3.0m
Path gradient	< 2.3%
Network	4G (services of three telecom operators)
Number and Type of Intersection	East-West route: 3 roundabouts and 3 T-intersections West-East route: 4 roundabouts and 3 T-intersections
Number of Pedestrian Crossings	Four on each routes
Speed Limit	50km/h

Figure 4: Statistics of autonomous shuttle service

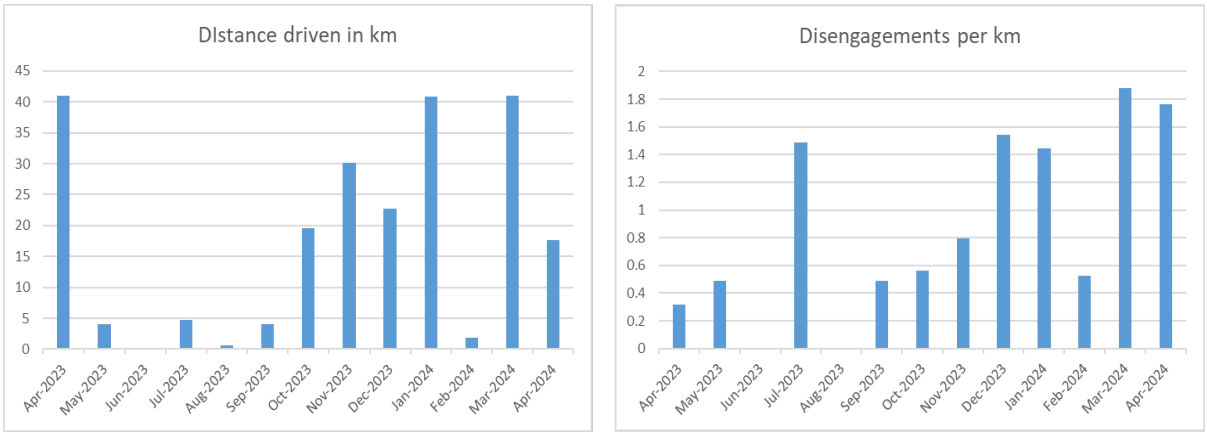
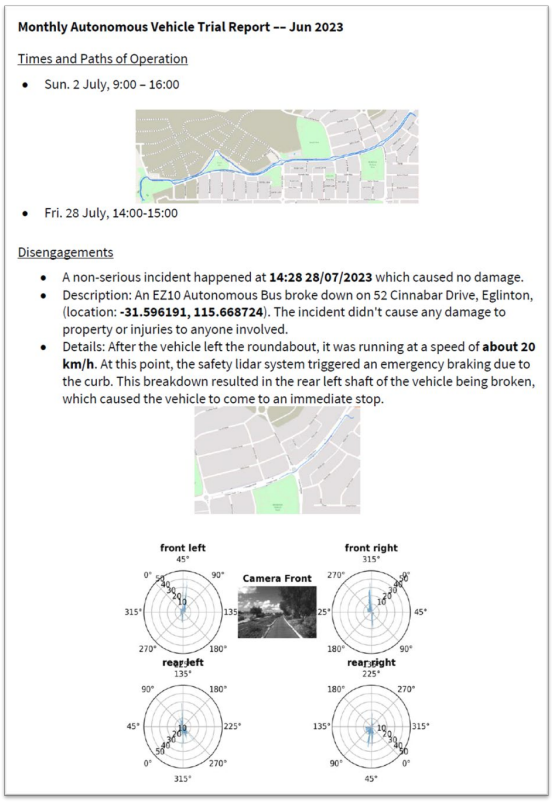


Figure 5: Demo report of the output



Our transportation data primarily utilizes the rosbag format, which is a widely used file format in the Robot Operating System (ROS) framework. Rosbag is designed to store serialized message data, enabling the recording and playback of topics from a ROS system. It encapsulates various types of sensor data, such as images, point clouds, and IMU readings, along with their respective timestamps, facilitating the synchronization and analysis of the collected data.

In our proposed pipeline, the user's request is initially processed by a LLM agent, which performs format standardization and structured configuration. The LLM plays a crucial role in interpreting the user's natural language request and converting it into a structured format that can be easily understood and processed by the subsequent modules.

The formatted request is then passed to the guardrails module, a safety mechanism responsible for securely retrieving the required sensor data from the log database. The guardrails module ensures that only authorized and relevant data is accessed, preventing any potential breaches or misuse of sensitive information. Upon receiving the instructions, the guardrails module interacts with an API toolkit to invoke pertinent APIs based on the desired report format. The API toolkit serves as a comprehensive library, offering a list of functions to process and visualize the retrieved sensor data. These APIs enable tasks such as data filtering, statistical analysis, and the generation of insightful visualizations. By leveraging the appropriate APIs, the guardrails module can extract meaningful information and transform the raw sensor data into a structured and informative representation.

Once the data has been processed, the report composition module assembles the various components and generates the final report. This module combines the outputs from the guardrails and API toolkit, organizing the information into a coherent and visually appealing format. The report may include text, tables, charts, and other visual elements to effectively convey the insights derived from the transportation data. To ensure a seamless user experience, the LLM is once again employed to communicate the generated report to the user and facilitate its display in the browser. The LLM converts the structured report data into a

natural language summary, making it easily comprehensible for the user. Additionally, the LLM can provide guidance on how to navigate and interpret the report within the browser interface.

It is important to note that the LLM is only invoked during the user request processing stage, ensuring that the actual report generation and data storage remain local. This architecture safeguards data privacy and security, as the sensitive transportation data never leaves the local environment. With the power of LLMs, guardrails, API toolkits, and report composition modules, our system creates a robust and secure pipeline for generating comprehensive reports from rosbag data. This workflow streamlines the process of extracting valuable insights from complex transportation data, enabling stakeholders to make informed decisions and optimize transportation systems. The localized nature of data storage and report generation further enhances the system's reliability and data protection measures.

5 Discussion

The proposed Autonomous Shuttle Traffic Reporter (ASTR) system, which integrates LLMs with a multi-agent architecture, demonstrates significant potential for enhancing the efficiency and granularity of traffic log generation in autonomous vehicle systems. By implementing guardrails and localized data storage, the ASTR system effectively maintains data security and privacy.

The case study was conducted using nearly a year's worth of service data from autonomous shuttle buses in Eglinton, Western Australia. The results demonstrate the system's capability to reduce log generation time and capture more traffic details compared to traditional methods, highlighting the potential for LLM-based systems to improve the efficiency and effectiveness of traffic management and decision-making processes in autonomous vehicle deployments.

5.1 Limitation

In our experiment, the results were constrained by several technical challenges. Firstly, there is no standardized metric for evaluating incident reports, making it difficult to quantify the improvements gained from incorporating map paths and sensor data. Secondly, the impact of filters and anonymization techniques on the generative content produced by LLMs remains uncertain and requires further discussion. Additionally, the rapid evolution of language models introduces new features with each release, some of which include built-in protections against data leakage. This continuous advancement complicates the assessment of how these models interact with the data and the effectiveness of the safeguards they provide.

5.2 Future Works

There are several areas we plan to explore in future research. Firstly, with the advent of multi-modality large language models, sensor data should be converted into higher-level representations beyond simple plot images. An ideal incident report should be easily understandable by a wide range of readers, regardless of their background in signal processing. Secondly, generative AI has demonstrated its potential to enhance perception in various scenarios. This technology could be leveraged to extend the vehicle's perception range during incidents, aiding in the reconstruction of scenarios for more comprehensive analysis.

5.3 Conclusions

In conclusion, the proposed report system represents a promising approach to leveraging LLMs for autonomous vehicle traffic report generation while maintaining data security and privacy. The case study results demonstrate the system's potential to enhance the efficiency and granularity of traffic reporting. By addressing the identified challenges and opportunities, the proposed system can contribute to the advancement of intelligent transportation systems and support the safe and efficient deployment of autonomous vehicles in urban environments.

6 Acknowledgments

The authors would like to thank all the Renewable Energy Vehicle Project (REV) sponsors for their support on this project, especially Stockland, Allkem and CD Dodd. The authors would like to thank Zhihui Lai, Kieran Quirke-Brown and Lee Le for participating in this project.

7 References

Colin, R 2020, 'Exploring the limits of transfer learning with a unified text-to-text transformer', Journal of Machine Learning Research, 140, vol. 21, pp. 1–67.

Cui, Y, Huang, S, Zhong, J, Liu, Z, Wang, Y, Sun, C, Li, B, Chen, L & Amir Khajepour 2023, 'DriveLLM: Charting the Path Toward Full Autonomous Driving with Large Language Models', IEEE transactions on intelligent vehicles, Institute of Electrical and Electronics Engineers, pp. 1–15.

de Zarzà, I, de Curtò, J, Roig, G & Calafate, CT 2023, 'LLM Multimodal Traffic Accident Forecasting', Sensors, vol. 23, no. 22, p. 9225, viewed 8 December 2023, <https://www.mdpi.com/1424-8220/23/22/9225>.

Dwork, C & Roth, A 2014, 'The algorithmic foundations of differential privacy', Foundations and Trends® in Theoretical Computer Science, 3–4, vol. 9, Now Publishers, Inc., pp. 211–407.

Greshake, K, Abdelnabi, S, Mishra, S, Endres, C, Holz, T & Fritz, M 2023, 'Not what you've signed up for: Compromising real-world LLM-Integrated applications with indirect prompt injection', in AISec '23: Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security, Association for Computing Machinery, pp. 79–90.

Huang, W, Abbeel, P, Pathak, D & Mordatch, I 2022, 'Language Models as Zero-Shot Planners: Extracting Actionable Knowledge for Embodied Agents', ArXiv (Cornell University), Cornell University.

Lai, S, Xu, Z, Zhang, W, Liu, H & Xiong, H 2024, LLMLight: Large Language Models as Traffic Signal Control Agents, arXiv.org, viewed 6 May 2024, <https://arxiv.org/abs/2312.16044>.

Lewis, P 2020, 'Retrieval-augmented generation for knowledge-intensive nlp tasks', Advances in Neural Information Processing Systems, vol. 33, pp. 9459–9474.

Mao, J, Ye, J, Qian, Y, Pavone, M & Wang, Y 2023, 'A Language Agent for Autonomous Driving', arXiv (Cornell University), Cornell University.

Martino, A, Iannelli, M & Truong, C 2023, 'Knowledge injection to counter large language model (LLM) hallucination', in Springer, pp. 182–185.

Namer, A, Miller, J, Vagts, H & Maltzman, B 2023, A cost-effective method to prevent data exfiltration from LLM prompt responses, Technical Disclosure Commons.

OpenAI, Achiam, J, Adler, S, Agarwal, S, Ahmad, L, Akkaya, I, Florencia Leoni Aleman, Almeida, D, Janko Altenschmidt, Altman, S, Shyamal Anadkat, Avila, R, Igor Babuschkin, Balaji, S, Balcom, V, Baltescu, P, Bao, H, Bavarian, M, Belgum, J & Bello, I 2024, GPT-4 technical report.

Ouyang, L 2022, 'Training language models to follow instructions with human feedback', Advances in neural information processing systems, vol. 35, pp. 27730–27744.

Quirke-Brown, K, Lai, Z, Kong, X, Tan, T, Du, Y & Bräunl, T 2023, 'Developing an Autonomous Shuttle Service', in Australasian Transport Research Forum 2023.

Sha, H, Mu, Y, Jiang, Y, Chen, L, Xu, C, Luo, P, Li, SE, Tomizuka, M, Zhan, W & Ding, M 2023, LanguageMPC: Large Language Models as Decision Makers for Autonomous Driving, arXiv.org, viewed 8 October 2023, <https://arxiv.org/abs/2310.03026>.

Smith, B 2023, How do we best govern AI?, microsoft on the issues, viewed May 2023, <https://blogs.microsoft.com/on-the-issues/2023/05/25/how-do-we-best-govern-ai/>.

Su, J, Jiang, C, Jin, X, Qiao, Y, Xiao, T, Ma, H, Wei, R, Jing, Z, Xu, J & Lin, J 2024, Large Language Models for Forecasting and Anomaly Detection: A Systematic Literature Review, arXiv.org, viewed 17 March 2024, <https://arxiv.org/abs/2402.10350>.

Tjondronegoro, D, Yuwono, E, Richards, B, Green, D & Siiri Hatakka 2022, Responsible AI implementation: A human-centered framework for accelerating the innovation process.

Vehicle Industry Services - California DMV 2023, California DMV, viewed 18 August 2024, <https://www.dmv.ca.gov/portal/vehicle-industry-services/>.

Wang, L, Ma, C, Feng, X, Zhang, Z, Yang, H, Zhang, J, Chen, Z, Tang, J, Chen, X, Lin, Y, Zhao, WX, Wei, Z & Wen, J 2024, 'A survey on large language model based autonomous agents', Frontiers of Computer Science, 6, vol. 18, Springer Science and Business Media LLC.

Wolf, Y, Wies, N, Oshri Avnery, Levine, Y & Amnon Shashua 2023, Fundamental limitations of alignment in large language models.

Yan, Y, Wen, H, Zhong, S, Chen, W, Chen, H, Wen, Q, Zimmermann, R & Liang, Y 2024, UrbanCLIP: Learning text-enhanced urban region profiling with contrastive language-image pretraining from the web.

Zheng, O, Abdel-Aty, M, Wang, D, Wang, Z & Ding, S 2023, ChatGPT is on the horizon: Could a large language model be suitable for intelligent traffic safety research and applications?

Zhou, Z, Lin, Y & Li, Y 2024, Large language model empowered participatory urban planning.