Large Language Models for Anomaly and Out-of-Distribution Detection: A Survey

Anonymous ACL submission

Abstract

Detecting anomalies or out-of-distribution (OOD) samples is critical for maintaining the reliability and trustworthiness of machine learning systems. Recently, Large Language Models (LLMs) have demonstrated their effectiveness not only in natural language processing but also in broader applications due to their advanced comprehension and generative capabilities. The integration of LLMs into anomaly and OOD detection marks a significant shift from the traditional paradigm in the field. This survey focuses on the problem of anomaly and OOD detection under the context of LLMs. We propose a new taxonomy to categorize existing approaches into three classes based on the role played by LLMs. Following our proposed taxonomy, we further discuss the related work under each of the categories and finally discuss potential challenges and directions for future research in this field.

1 Introduction

Most machine learning models operate under the closed-set assumption (Krizhevsky et al., 2012), where the test data is assumed to be drawn i.i.d. from the same distribution as the training data. However, in real-world applications, this assumption often cannot hold, as test examples can come from distributions not represented in the training data. These instances, known as anomalies or out-of-distribution (OOD) samples, can severely degrade the performance and reliability of existing models (Yang et al., 2024a). To build robust AI systems, methods including probabilistic approaches (Lee et al., 2018; Leys et al., 2018) and recent deep learning techniques (Pang et al., 2021; Yang et al., 2024a) have been explored to detect these unknown instances across various domains, such as fraud detection in finance and fault detection in industrial systems (Hilal et al., 2022; Liu et al., 2024b).

Large Language Models (LLMs), such as GPT-4 (Achiam et al., 2023) and LLaMA (Touvron et al., 2023), have recently demonstrated remarkable capabilities in language comprehension and summarization. To further harness the potential of LLMs beyond text data, there is also a growing interest in extending them to multi-modal tasks such as vision-language understanding and generation (Wang et al., 2024), evolving them into Multimodal LLMs (MLLMs) (Yin et al., 2023). Given the zero- and few-shot reasoning capabilities of LLMs and MLLMs, researchers try to apply these models to anomaly and out-of-distribution (OOD) detection, as illustrated in Figure 1, yielding promising results. However, the emergence of LLMs has fundamentally changed the learning paradigm in this field, highlighting the need for a comprehensive survey to analyze the emerging challenges and systematically review the rapidly expanding works.

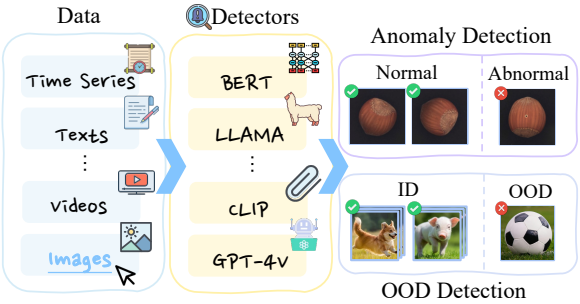

Figure 1: A simple illustration of leveraging LLMs for images anomaly and OOD detection.

While prior works have explored various aspects of anomaly and OOD detection, none have specifically focused on the utilization of LLMs on these problems across diverse data modalities. Yang et al. (2024a) and Salehi et al. (2021) present unified frameworks for OOD detection but do not delve into the utilization of LLMs. While Su et al. (2024) review some small-sized language models for forecasting and anomaly detection, they neither cover the usage of LLMs with emergent abilities nor address OOD detection. A recent survey by Miyai

et al. (2024a) summarizes works on anomaly and OOD detection in vision using vision-language models but neglects other data modalities. Therefore, we aim to conduct a systematic survey that covers both anomaly and OOD detection tasks across various data domains, concentrating on how LLMs are used in existing works.

In this survey, we propose a novel taxonomy that focuses on how LLMs can profoundly impact anomaly and OOD detection in three fundamental ways, as illustrated in Figure 2: ❶ **LLMs for Augmentation (§3):** LLMs are not used directly for detection, but their emergent abilities, advanced semantic understanding, and vast knowledge augment the detection process; ❷ **LLMs for Detection (§4):** LLMs are employed as a detector to identify anomalies and OOD instances; and ❸ **LLMs for Explanation (§5):** LLMs provide insightful explanatory analyses of detection results, aiding in further planning and problem-solving in real-world scenarios. At the end (**§6**), we also outline the challenges and future research directions, in order to provide a better understanding of anomaly and OOD detection in the era of LLMs and shed light on the following research.

2 Preliminaries

Large Language Models. Large language models (LLMs) generally refer to Transformer-based pre-trained language models with hundreds of billions of parameters or more. Early LLMs like BERT (Devlin et al., 2018) and RoBERTa (Liu et al., 2019) utilize an encoder-only architecture, excelling in text representation learning (Bengio et al., 2013). Recently, the focus has shifted toward models aimed at natural language generation, often using the "next token prediction" objective as their core task. Examples include T5 (Raffel et al., 2020) and BART (Lewis et al., 2019), which employ an encoder-decoder structure, as well as GPT-3 (Brown et al., 2020), PaLM (Chowdhery et al., 2023), and LLaMA (Touvron et al., 2023), which are based on decoder-only architectures. Advancements in these architectures and training methods have led to superior reasoning and emergent abilities, such as in-context learning(Brown et al., 2020) and chain-of-thought reasoning (Wei et al., 2022).

Multimodal Large Language Models. The remarkable abilities of Large Language Models (LLMs) have inspired efforts to integrate language with other modalities, with a particular focus on combining language and vision.Notable examples of Multimodal Large Language Models include CLIP (Radford et al., 2021), BLIP2 (Li et al., 2023a), and Flamingo (Alayrac et al., 2022), which were pre-trained on large-scale cross-modal datasets comprising images and text. Models like GPT-4(V) (OpenAI, 2023) and Gemini (Team et al., 2023) showcase the emergent abilities of Multimodal LLMs, significantly improving vision understanding. In light of the emergence of these MLLMs, researchers are increasingly using them as backbones to tackle tasks such as anomaly and OOD detection.

2.1 Problem Definition

With LLMs advancing in zero-shot and few-shot learning, the general pipeline of anomaly and out-of-distribution (OOD) detection methods shifts to adapt pre-trained LLMs for detection without extensive training. This shift challenges traditional definitions of anomaly and OOD detection, as the conventional train-test paradigm may not always apply. Following previous studies (Miyai et al., 2024a; Yang et al., 2024a), we propose to redefine anomaly and OOD detection under the context of LLMs and highlight the differences between the two problems as follows:

Definition 1 ***LLM-based Anomaly Detection:*** *Given a test dataset $\mathcal{D}_{test} = \{x_1, \cdots, x_n\}$, where each sample x_i is drawn from distribution $\mathbb{P}^{in}$ or $\mathbb{P}^{out}$. The objective of LLM-based Anomaly Detection is to use a pre-trained LLM as the backbone and develop a detection model $f_{LLM}(\cdot)$ to predict whether each sample $x' \in \mathcal{D}_{test}$ belongs to $\mathbb{P}^{out}$, where $\mathbb{P}^{out}$ has covariate shift with $\mathbb{P}^{in}$*

Definition 2 ***LLM-based OOD Detection:*** *Given a test dataset $\mathcal{D}_{test} = \{x_1, \cdots, x_n\}$, where each sample x_i is drawn from distribution $\mathbb{P}^{in}$ or $\mathbb{P}^{out}$, and a known ID class set $\mathcal{C} = \{c_1, \cdots, c_k\}$. The objective of LLM-based OOD Detection is to use a pre-trained LLM as backbone and develop detection model $f_{LLM}(\cdot)$ to predict whether each sample $x' \in \mathcal{D}_{test}$ belongs to $\mathbb{P}^{out}$, where $\mathbb{P}^{out}$ has semantic shift with $\mathbb{P}^{in}$. If not, x' will be classified into $x_i \in C$.*

Discussions. The distinction between anomaly detection and OOD detection in the context of LLMs highlights the unique challenges posed by covariate and semantic shifts. *Anomaly detection* aims to identify subtle deviations within the data that may not involve a complete change in the underlying

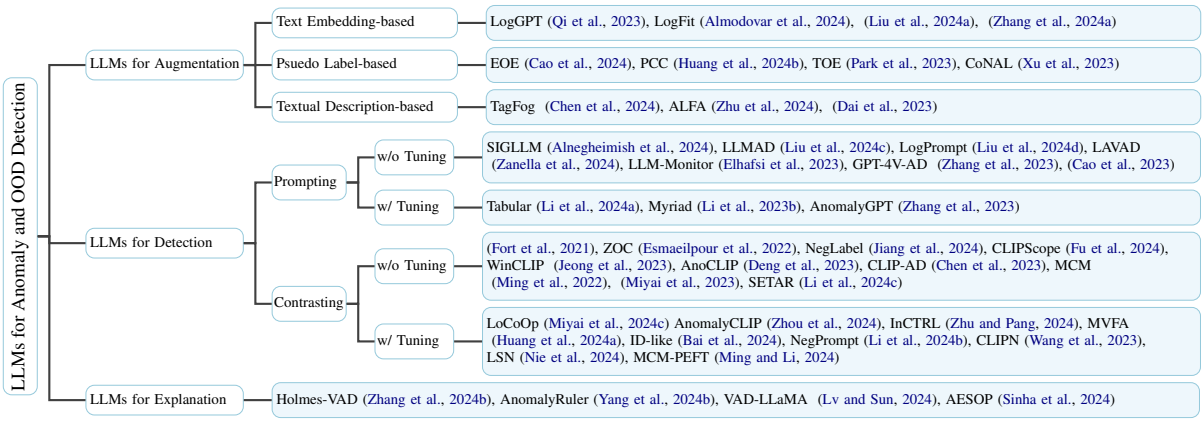

Figure 2: Taxonomy of methods utilizing LLMs for anomaly and OOD detection tasks.

class or concept, such as detecting defects or irregularities in industrial processes. In contrast, *OOD detection* focuses on identifying instances that do not belong to any of the known ID classes at the object level, such as recognizing a dog when the only provided ID class is cat. This differentiation underscores the need for tailored approaches for each detection task.

3 LLMs for Augmentation

In this section, we review methods that leverage LLM as a data augmenter, producing meaningful augmented knowledge that enhances the detection of anomalies or OOD samples. Such augmented information includes text embedding, pseudo labels, and textual descriptions derived from LLMs. Therefore, these approaches can be categorized into three types as shown in Figure 3.

3.1 Text Embedding-based Augmentation

LLMs are powerful feature extractor which can derive meaningful and effective embedding used for further detection tasks. For instance, in log data, Hadadi et al. (2024) and Qi et al. (2023) fine-tune pre-trained GPT models in a supervised manner and use the extracted semantic embeddings as an important component for future anomaly detection.

For OOD detection in text data, a standard pipeline involves using encoder-only LLMs to generate sentence representations, which are then used to derive OOD confidence scores. Typically, these models are fine-tuned on ID data, and OOD detectors are applied to the sentence representations they produce (Liu et al., 2024a). Recently, there has been a shift toward leveraging larger language models with decoder architectures, which offer enhanced capabilities in extracting and refining textual representations. Liu et al. (2024a) explore the use of decoder-only LLMs, such as LLaMa, incorporating fine-tuning techniques like LoRA to minimize additional parameter usage. Their findings demonstrate that fine-tuned LLMs, when combined with customized OOD scoring functions, can significantly improve OOD detection performance. A key advantage of recent LLMs with decoder architecture is their autoregressive ability, which allows for more effective handling of sequential data. Building on this, Zhang et al. (2024a) propose using the likelihood ratio between a pre-trained LLM and its fine-tuned variant as a criterion for OOD detection, effectively leveraging the deep, contextual knowledge embedded within LLMs for text data.

3.2 Pseudo Label-based Augmentation

The emergent abilities of LLMs offer a promising approach for generating high-quality synthetic datasets that, in some cases, can surpass those curated by humans (Ding et al., 2024). A significant challenge in using LLMs for OOD detection is the lack of OOD labels, which often hampers model performance. Traditional methods rely on extensive human effort and auxiliary datasets, but LLMs can overcome this by generating high-quality pseudo-OOD labels through suitable prompts. These pseudo labels can then be used as text prompts for contrasting-based OOD detection methods, augment existing ID data and enhance the distinction between ID and OOD samples during detection.

EOE (Cao et al., 2024) and PCC (Huang et al., 2024b) prompt LLMs to generate potential OOD class labels which are visually similar to known ID classes. Then, they define a new score function with penalty on these generated pseudo labels dur-

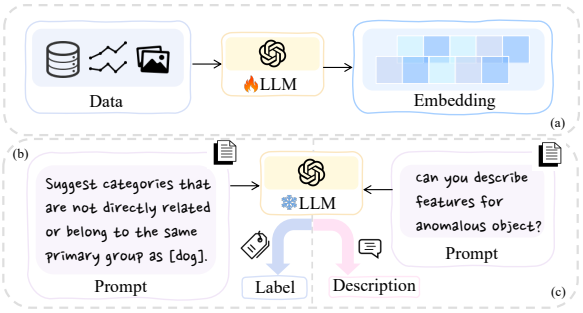

Figure 3: The illustration of three approaches in (§3): (a) Text Embedding-based Augmentation; (b) Pseudo Label-based Augmentation, and (c) Textual Description-based Augmentation.

ing inference stage, greatly outperforming methods with only known ID labels. Following the similar idea, TOE (Park et al., 2023) further evaluates generating pseudo OOD labels for OOD detection at three verbosity levels: word-level, description-level, and caption-level, using BERT, GPT-3 and BLIP-2 respectively. Results indicate that using caption-level pseudo OOD labels outperform other two approaches since BLIP-2 can leverage both semantic and visual underdtanding. For text data, CoNAL (Xu et al., 2023) prompts LLMs to extend closed-set labels with novel ones and generates new examples based on these labels, forming a comprehensive set of probable OOD samples. By utilizing a contrastive confidence loss for training, detection model achieves both high accuracy on the ID training set and lower relative confidence on the generated novel examples.

3.3 Textual Description-based Augmentation

In addition to generating pseudo lebels, other methods utilize LLMs to generate detailed textual descriptions about both known ID classes and potential unknown OOD samples. For example, TagFog (Chen et al., 2024) uses the Jigsaw strategy to generate fake OOD samples and prompts ChatGPT to create detailed descriptions for each ID class, guiding the training of the image encoder of CLIP for OOD detection. When using LLMs for anomaly detection, it is crucial to make LLMs recognize the close correlation between normal images and their respective normal prompts, while identifying a more distant association with abnormal prompts. Therefore, detailed and nuanced descriptions of normal and anomalous stages of an object are necessary. ALFA (Zhu et al., 2024) formulates prompts to query an LLM to describe normal and abnormal features for each class and then used these descriptions together as prompts for LLMs to better identify abnormal object. To avoid LLM hallucination issues, Dai et al. (2023) use LLMs to describe visual features for distinguishing categories in images and introduce a consistency-based uncertainty calibration method to estimate the confidence score of each generation.

4 LLMs for Detection

The primary objective of this section is to explore existing works that utilize LLMs to detect anomalies or OOD samples. Under this line of research, approaches can be categorized into two classes as illustrated in Figure 4: ❶ Prompting-based Detection methods, which involve directly prompting LLMs to generate language responses that include detection results; ❷ Contrasting-based Detection methods, which focus on multimodal scenarios, using MLLMs pre-trained with a contrastive objective as detectors.

4.1 Prompting-based Detection

The general pipeline for prompting-based detection methods consists of two primary stages: (i) constructing a structured prompt template with instruction prompt $\mathcal{P}$ and input data $\mathcal{X}$; and (ii) feeding the template-based prompt $\hat{X}$ into LLMs to generate a language response. The function $\texttt{Parse}(\cdot)$ is then applied to extract the detection results. Depending on the scenario, the LLM can either be frozen or fine-tuned, denoted as $f^{\heartsuit}_{\text{LLM}}$ or $f^{\clubsuit}_{\text{LLM}}$, respectively. This process can be summarized as follows:

$$\text{Prompt Construction:} \quad \hat{X} = \texttt{Template}(\mathcal{X}, \mathcal{P}),$$
$$\text{Detection:} \quad \tilde{Y} = \texttt{Parse}\left(f^{\heartsuit/\clubsuit}_{\text{LLM}}(\hat{X})\right)$$

4.1.1 Detection without LLM Tuning

Since some approaches do not require additional tuning, they mainly focus on employing various prompt engineering techniques (Sahoo et al., 2024) to guide LLMs to produce better detection results. To design suitable prompts for anomaly or OOD detection, researchers have employed a combination of various prompt techniques, such as *role-play prompting* (Wu et al., 2023), *in-context learning* (Brown et al., 2020), and *chain-of-thought (CoT) reasoning* (Wei et al., 2022), to create effective prompt templates. Studies such as SIGLLM (Alnegheimish et al., 2024), LLMAD (Liu et al., 2024c), and LogPrompt (Liu et al., 2024d) focus on

time series and log data. SIGLLM (Alnegheimish et al., 2024) investigates two distinct pipelines for using LLMs in time series anomaly detection: one directly prompts an LLM with specific role-play instructions to identify anomalous elements in given data, and the other uses the LLM's forecasting ability to detect anomalies by comparing original and forecasted signals, where discrepancies indicate anomalies. LLMAD (Liu et al., 2024c) incorporates in-context learning examples retrieved from a constructed database and CoT prompts that inject domain knowledge of time series. LogPrompt (Liu et al., 2024d) explores three prompting strategies for log data: self-prompt, CoT prompt, and in-context prompt, demonstrating that the prompt with CoT techniques outperforms other prompting strategies. The tailored CoT prompt for log data includes a specific task instruction, i.e. "classify the given log entries into normal and abnormal categories", and step-by-step rules for considering given data as anomalies.

Unlike time series and log data which can be directly converted into raw text data, other data modalities, such as videos and images, require additional processing to be transformed into a format that LLMs can understand. For instance, LAVAD (Zanella et al., 2024) first exploits a captioning model to generate a textual description for each video frame and further uses an LLM to summarize captions within a temporal window. This summary is then used to prompt the LLM to provide an anomaly score for each frame. LLM-Monitor (Elhafsi et al., 2023) uses an object detector to identify objects in video clips and then designs specific prompt templates incorporating CoT and in-context examples to query LLMs for anomaly detection.

With the integration of multimodal understanding into LLMs, these models are now capable of comprehending various modalities beyond text, enabling more direct applications for anomaly detection across a wide range of data types. Cao et al. (2023) conduct comprehensive experiments and analyses using GPT-4V(ision) for anomaly detection across various modality datasets and tasks. To enhance GPT-4V's performance, they also incorporate different types of additional cues such as class information, human expertise, and reference images as prompts. Similarly, GPT-4V-AD (Zhang et al., 2023) employs GPT-4V as the backbone, designing a general prompt description for all image categories and injecting specific image category information, resulting in a specific output format

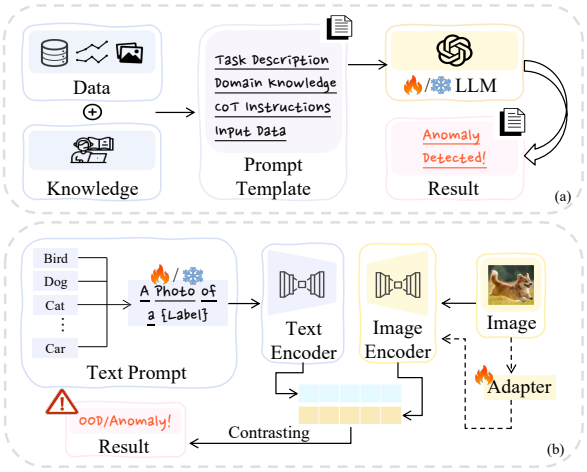

Figure 4: The illustration of two approaches in (§4): (a) Prompting-based Detection and (b) Contrasting-based Detection.

for each region with respective anomaly scores.

4.1.2 Detection with LLM Tuning

Directly prompting frozen LLMs for anomaly or OOD detection results across various data types often yields suboptimal performance due to the inherent modality gap between text and other data modalities. As a result, additional training and fine-tuning on LLMs for downstream detection tasks has become a prevalent research trend. Unfortunately, fine-tuning entire LLMs is often computationally expensive and poses significant challenges. Therefore, parameter-efficient fine-tuning (PEFT) has been extensively employed instead. For example, Tabular (Li et al., 2024a) designs a prompt template to query the LLM to output anomalies based on given converted tabular data. To better adapt the LLM for anomaly detection at the batch level, they apply Low-Rank Adaptation (LoRA), using a synthetic dataset with ground truth labels in a supervised manner.

To enhance LLMs for localization understanding and adapting to industrial tasks, AnomalyGPT (Zhang et al., 2023) first derives localization features from a frozen image encoder and image decoder and these features are then fed to a tunable prompt learner. Without fine-tuning the entire LLM, they fine-tune the prompt learner with LoRA to significantly reduce computational costs. Myriad (Li et al., 2023b) employs Mini-GPT-4 as the backbone and integrates a trainable encoder, referred to as Vision Expert Tokenizer, to embed the vision expert's segmentation output into tokens that the LLM can understand. With expert-driven

visual-language extraction, Myriad can generate accurate anomaly detection descriptions.

4.2 Contrasting-based Detection

In this section, we focus on MLLMs, such as CLIP, which are pre-trained with an image-text contrastive objective and learn by pulling the paired images and texts close and pushing others far away in the embedding space. The zero-shot classification ability of these models further builds the foundation for contrasting-based anomaly and OOD detection methods: (i) given an image x_i and a text prompt f with a target class set C, CLIP extracts image features $h \in \mathbb{R}^D$ using an image encoder f_{img}, and text features $e_j \in \mathbb{R}^D$ using a text encoder f_{text} with a prompt template for each class $c_j \in C$, and (ii) the similarity between h and each e_j is usually used as an important component in the score function f_{score} for deciding whether x_i is an anomaly or OOD sample. This process can be summarized as follows:

$$\begin{aligned} \text{Feature Extraction:} \quad & h = f_{\text{img}}(x_i), \\ \textit{and} \quad & e_j = f_{\text{text}}(\text{prompt}(c_j)), \\ \text{Detection:} \quad & \tilde{Y} = f_{\text{score}}\left(\cos(h, e_j)\right) \end{aligned}$$

We further categorize contrasting-based detection methods into two main classes depending on whether there exists additional training and fine-tuning.

4.2.1 Detection without LLM Tuning

Despite the promise, existing CLIP-like models perform zero-shot classification in a closed-world setting. That is, it will match an input into a fixed set of categories, even if it is irrelevant (Ming et al., 2022). To address this, one approach involves designing effective post-hoc score functions tailored for OOD detection that solely rely on ID class labels. Alternatively, some researchers incorporate anomaly or OOD class information into the text prompts, allowing the model to match OOD or abnormal images to paired prompts.

- *Without Anomaly/OOD Prompts.* To address the challenges of OOD detection using only in-distribution (ID) class information while avoiding the matching of OOD inputs to irrelevant ID classes, one notable approach is the Maximum Concept Matching (MCM) framework proposed by (Ming et al., 2022). This method is not limited to CLIP and can be generally applicable to other pre-trained models that promote multi-modal feature alignment. They view the textual embeddings of ID classes as a collection of concept prototypes and define the maximum concept matching (MCM) score based on the cosine similarity between the image feature and the textual feature. Following the idea of MCM, several subsequent works focus on improving OOD detection results by either adding a local MCM score or modifying weights in the original MCM framework, such as (Miyai et al., 2023) and (Li et al., 2024c).

- *With Anomaly/OOD Prompts.* Fort et al. (2021) first investigate using CLIP for OOD detection and demonstrate encouraging performance. However, in their setup, they include the candidate labels related to the actual OOD classes and utilize this knowledge as a very weak form of outlier exposure, which contradicts the open-world assumption. Therefore, after this work, researchers aim to leverage pseudo-OOD labels in the text prompt instead of using actual OOD labels. The earliest work under this idea is ZOC (Esmaeilpour et al., 2022) which trains a text description generator on top of CLIP's image encoder to dynamically generate candidate unseen labels for each test image. The similarity of the test image with seen and generated unseen labels is used as the OOD score. Instead of training an additional text decoder, NegLabel (Jiang et al., 2024) and CLIPScope (Fu et al., 2024) rely on auxiliary datasets to gather potential OOD labels. CLIPScope gathers nouns from open-world sources as potential OOD labels and uses them in designed prompts to ensure maximal coverage of potential OOD samples. NegLabel employs the NegMining algorithm to select high-quality negative labels with sufficient semantic differences from ID labels. Recent work utilizes the emergent abilities of LLMs to generate reliable OOD labels, such as (Cao et al., 2024), (Huang et al., 2024b), (Park et al., 2023), and (Xu et al., 2023).

For contrasting-based anomaly detection, WinCLIP (Jeong et al., 2023) initially investigates a one-class design by using only the normal prompt "normal [o]" where [o] represents object-level label, i.e "bottle", and defining an anomaly score as the similarity between vectors derived from the image encoder and normal prompts. However, this one-class design yields poorer results compared to a simple binary zero-shot framework, CLIP-AC (Jeong et al., 2023), which adapts CLIP

with two class prompts: "normal [o]" vs. "anomalous [o]". This framework sets the foundational pipeline for future work in contrasting-based anomaly detection and has inspired subsequent research.

While using the default prompt has demonstrated promising performance, similar to the prompt engineering discussion around GPT-3 (Brown et al., 2020), researchers have observed that performance can be significantly improved by customizing the prompt text. Models like WinCLIP (Jeong et al., 2023) and AnoCLIP (Deng et al., 2023) use a Prompt Ensemble technique to generate all combinations of pre-defined lists of state words per label and text templates. After generating all combinations of states and templates, they compute the average of text embeddings per label to represent the normal and anomalous classes. In practice, more descriptions in prompts do not always yield better performance. Therefore, CLIP-AD (Chen et al., 2023) proposes Representative Vector Selection (RVS), from a distributional perspective for the design of the text prompt, broadening research opportunities beyond merely crafting adjectives.

4.2.2 Detection with LLM Tuning

Following the similar detection pipeline of methods without LLM tuning, researchers propose to employ prompt tuning or adapter tuning techniques to eliminate the need for manually crafting prompts and enhance the understanding of local features of images. Additionally, by incorporating a few ID or normal images during training or inference phases, some methods transition into few-shot scenarios.

- *LLM Adapter-Tuning.* Adapter-tuning methods involve integrating additional components or layers into the model architecture to facilitate better alignment or localization (Hu et al., 2023). CLIP was originally designed for classifying the semantics of objects in the scene, which does not align well with the sensory anomaly detection task where both normal and abnormal samples are often from the same class of object. To reconcile this, InCTRL (Zhu and Pang, 2024) includes a tunable adapter layer to further adapt the image representations for anomaly detection. To better adapt to medical image anomaly detection, MVFA (Huang et al., 2024a) proposes a multi-level visual feature adaptation architecture to align CLIP's features with the requirements of anomaly detection in medical contexts. This is achieved by integrating multiple residual adapters into the pre-trained visual encoder, guided by multi-level, pixel-wise visual-language feature alignment loss functions.

- *LLM Prompt-Tuning.* Manually crafting suitable prompts always requires extensive human effort. Therefore, researchers employ the idea of prompt tuning, such as CoOp (Zhou et al., 2022), to learn a soft or differentiable context vector to replace the fixed text prompt. For OOD detection, most approaches rely on using auxiliary prompts to represent potential OOD textual information, and one crucial problem is to identify hard OOD data that is similar to ID samples. To solve this, Bai et al. (2024) first constructs outliers highly correlated with ID data and introduces a novel prompt learning framework for learning specific prompts for the most challenging OOD samples, which behave like ID classes. Additionally, LSN (Nie et al., 2024), NegPrompt (Li et al., 2024b), and CLIPN (Wang et al., 2023) all work on learning extra negative prompts to fully leverage the capabilities of CLIP for OOD detection. Unlike the other two approaches, CLIPN requires training an additional "no" text encoder using a large external dataset to get negative prompts for all classes. This auxiliary training is computationally expensive, limiting its application to generalized tasks. Also, LSN demonstrates that naive "no" logic prompts cannot fully leverage negative features. Therefore, both LSN and NegPrompt focus on training on more detailed negative prompts, while LSN also aims to develop class-specific positive and negative prompts, enabling more accurate detection.

  Instead of focusing on leveraging OOD information, some methods aim to perform prompt tuning to optimize word embeddings for ID labels and then use the MCM score as the detection criterion. MCM-PEFT (Ming and Li, 2024) demonstrates that simply applying prompt tuning for CLIP on few-shot ID datasets can significantly improve detection accuracy. However, a primary limitation of this approach is its exclusive reliance on the features of ID classes, leading to incorrect detection when input images share a high visual similarity with the class in the prompt. To address this, LoCoOp (Miyai et al., 2024c) treats such ID-irrelevant nuisances as OOD and learns to push them away from the ID class text embed-

dings, preventing the model from producing high ID confidence scores for the OOD features. Additionally, Lafon et al. (2024) enhances detection capabilities by learning a diverse set of prompts utilizing both global and local visual representations. To better adapt to learning local features, AnomalyCLIP (Zhou et al., 2024) aims to learn object-agnostic text prompts that capture generic normality and abnormality in images, allowing the model to focus on abnormal regions rather than object semantics.

5 LLMs for Explanation

Due to their remarkable capabilities in understanding and generating human-like text, LLMs have been explored for providing insightful explanations and analyses for anomaly or OOD detection results, thereby aiding in further planning and problem-solving.

For applications in safety-critical domains, such as autonomous driving, providing explanations to stakeholders of AI systems has become an ethical and regulatory requirement (Li et al., 2023c). Consequently, there is a growing interest in developing explainable video anomaly detection frameworks. Holmes-VAD (Zhang et al., 2024b), for instance, trains a lightweight temporal sampler to select frames with high anomaly scores and then employs an LLM to generate detailed explanatory analyses, offering clear insights into the detected anomalies. VAD-LLaMA (Lv and Sun, 2024) generates instruction-tuning data to train only the projection layer of Video-LLaMA, enabling more comprehensive explanations of anomalies. AnomalyRuler (Yang et al., 2024b) emphasizes rule-based reasoning with efficient few-normal-shot prompting, allowing for rapid adaptation to different VAD scenarios while providing interpretable, rule-driven explanations.

Moreover, with the powerful capabilities of LLMs in understanding instructions and self-planning to solve tasks, an emerging research direction is to build autonomous agents based on LLMs to guide decision-making after anomalies or OOD are detected. For instance, AESOP (Sinha et al., 2024) employs the autoregressive generation of an LLM to provide a zero-shot assessment of whether interventions are needed for the robotic system after an anomaly is detected.

6 Challenges and Future Directions

In this section, we briefly summarize challenges and future directions within the anomaly and OOD detection research field in the era of LLMs.

Explainability and Trustworthiness. There is an increasing trend of utilizing LLMs to build explainable anomaly or OOD detection frameworks. Future research should focus on developing methods to enhance the explainability of LLMs for anomaly and OOD detection, increasing the trustworthiness of LLM-based systems and facilitating their adoption in critical domains such as healthcare, finance, and security (Holzinger et al., 2019; Guidotti et al., 2019; Ribeiro et al., 2016).

Unsolvable Problem Detection. Miyai et al. (2024b) propose Unsolvable Problem Detection (UPD), which evaluates the LLMs' ability to recognize and abstain from answering unexpected or unsolvable input questions, aiding in preventing incorrect or misleading outputs in critical applications where the consequences of errors can be significant. Future work should focus on developing effective solutions for this problem.

Handling Multimodal Data. The emergence of MLLMs capable of processing and understanding multiple data types offers significant potential (Alayrac et al., 2022; Li et al., 2023a). Future research should explore methods to better adapt LLMs to comprehend and integrate various multimodal data, thereby enhancing their ability to detect anomalies and OOD instances across diverse datasets.

7 Conclusion

In this survey, we examined the use of Large Language Models (LLMs) and multimodal LLMs (MLLMs) in anomaly and out-of-distribution (OOD) detection. We introduced a novel taxonomy categorizing methods into three approaches: augmentation, detection, and explanation. This taxonomy clarifies how LLMs can augment data, detect anomalies or OOD, and build explainable systems. We also discussed limitations and future research directions, aiming to highlight advancements and challenges in the field and encourage further progress.

Limitations

While this survey provides a comprehensive overview of the utilization of Large Language Models (LLMs) for anomaly and out-of-distribution (OOD) detection, several limitations should be acknowledged:

- **Scope of Coverage**: Although we endeavored to include the latest research, the rapid pace of advancements in the field means that some recent developments may not be covered.
- **Depth of Analysis**: Given the broad range of topics discussed, certain methods may not be explored in the depth they deserve.
- **Evaluations and Benchmarks**: Due to space constraints, we did not include a detailed summary of common evaluation metrics and benchmark datasets used in this area.

By acknowledging these limitations, we aim to provide a balanced perspective and encourage further research to address these gaps and build on the foundations laid by this survey.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, et al. 2022. Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems*, 35:23716–23736.

Crispin Almodovar, Fariza Sabrina, Sarvnaz Karimi, and Salahuddin Azad. 2024. Logfit: Log anomaly detection using fine-tuned language models. *IEEE Transactions on Network and Service Management*, 21(2):1715–1723.

Sarah Alnegheimish, Linh Nguyen, Laure Berti-Equille, and Kalyan Veeramachaneni. 2024. Large language models can be zero-shot anomaly detectors for time series? *arXiv preprint arXiv:2405.14755*.

Yichen Bai, Zongbo Han, Bing Cao, Xiaoheng Jiang, Qinghua Hu, and Changqing Zhang. 2024. Id-like prompt learning for few-shot out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17480–17489.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Chentao Cao, Zhun Zhong, Zhanke Zhou, Yang Liu, Tongliang Liu, and Bo Han. 2024. Envisioning outlier exposure by large language models for out-of-distribution detection. In *ICML*.

Yunkang Cao, Xiaohao Xu, Chen Sun, Xiaonan Huang, and Weiming Shen. 2023. Towards generic anomaly detection and understanding: Large-scale visual-linguistic model (gpt-4v) takes the lead. *arXiv preprint arXiv:2311.02782*.

Jiankang Chen, Tong Zhang, Wei-Shi Zheng, and Ruixuan Wang. 2024. Tagfog: Textual anchor guidance and fake outlier generation for visual out-of-distribution detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 1100–1109.

Xuhai Chen, Jiangning Zhang, Guanzhong Tian, Haoyang He, Wuhao Zhang, Yabiao Wang, Chengjie Wang, Yunsheng Wu, and Yong Liu. 2023. Clip-ad: A language-guided staged dual-path model for zero-shot anomaly detection. *arXiv preprint arXiv:2311.00453*.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. 2023. Palm: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240):1–113.

Yi Dai, Hao Lang, Kaisheng Zeng, Fei Huang, and Yongbin Li. 2023. Exploring large language models for multi-modal out-of-distribution detection. *arXiv preprint arXiv:2310.08027*.

Hanqiu Deng, Zhaoxiang Zhang, Jinan Bao, and Xingyu Li. 2023. Anovl: Adapting vision-language models for unified zero-shot anomaly localization. *arXiv preprint arXiv:2308.15939*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Bosheng Ding, Chengwei Qin, Ruochen Zhao, Tianze Luo, Xinze Li, Guizhen Chen, Wenhan Xia, Junjie Hu, Anh Tuan Luu, and Shafiq Joty. 2024. Data augmentation using llms: Data perspectives, learning paradigms and challenges. *arXiv preprint arXiv:2403.02990*.

Amine Elhafsi, Rohan Sinha, Christopher Agia, Edward Schmerling, Issa AD Nesnas, and Marco Pavone. 2023. Semantic anomaly detection with large language models. *Autonomous Robots*, 47(8).

Sepideh Esmaeilpour, Bing Liu, Eric Robertson, and Lei Shu. 2022. Zero-shot out-of-distribution detection based on the pre-trained model clip. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 6568–6576.

Stanislav Fort, Jie Ren, and Balaji Lakshminarayanan. 2021. Exploring the limits of out-of-distribution detection. *Advances in Neural Information Processing Systems*, 34.

Hao Fu, Naman Patel, Prashanth Krishnamurthy, and Farshad Khorrami. 2024. Clipscope: Enhancing zero-shot ood detection with bayesian scoring. *arXiv preprint arXiv:2405.14737*.

Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2019. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):1–42.

Fatemeh Hadadi, Qinghua Xu, Domenico Bianculli, and Lionel Briand. 2024. Anomaly detection on unstable logs with gpt models. *arXiv preprint arXiv:2406.07467*.

Waleed Hilal, S. Andrew Gadsden, and John Yawney. 2022. Financial fraud: A review of anomaly detection techniques and recent advances. *Expert Systems with Applications*, 193:116429.

Andreas Holzinger, Georg Langs, Daniel Denk, Kurt Zatloukal, and Henning Müller. 2019. Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4):e1312.

Zhiqiang Hu, Lei Wang, Yihuai Lan, Wanyu Xu, Ee-Peng Lim, Lidong Bing, Xing Xu, Soujanya Poria, and Roy Ka-Wei Lee. 2023. Llm-adapters: An adapter family for parameter-efficient fine-tuning of large language models. *arXiv preprint arXiv:2304.01933*.

Chaoqin Huang, Aofan Jiang, Jinghao Feng, Ya Zhang, Xinchao Wang, and Yanfeng Wang. 2024a. Adapting visual-language models for generalizable anomaly detection in medical images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11375–11385.

K Huang, G Song, Hanwen Su, and Jiyan Wang. 2024b. Out-of-distribution detection using peer-class generated by large language model. *arXiv preprint arXiv:2403.13324*.

Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. 2023. Winclip: Zero-/few-shot anomaly classification and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19606–19616.

Xue Jiang, Feng Liu, Zhen Fang, Hong Chen, Tongliang Liu, Feng Zheng, and Bo Han. 2024. Negative label guided ood detection with pretrained vision-language models. *arXiv preprint arXiv:2403.20078*.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, page 1097–1105, Red Hook, NY, USA. Curran Associates Inc.

Marc Lafon, Elias Ramzi, Clément Rambour, Nicolas Audebert, and Nicolas Thome. 2024. Gallop: Learning global and local prompts for vision-language models. *arXiv preprint arXiv:2407.01400*.

Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. 2018. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in neural information processing systems*, 31.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Christophe Leys, Olivier Klein, Yves Dominicy, and Christophe Ley. 2018. Detecting multivariate outliers: Use a robust variant of the mahalanobis distance. *Journal of Experimental Social Psychology*, 74:150–156.

Aodong Li, Yunhan Zhao, Chen Qiu, Marius Kloft, Padhraic Smyth, Maja Rudolph, and Stephan Mandt. 2024a. Anomaly detection of tabular data using llms. *arXiv preprint arXiv:2406.16308*.

Junnan Li, Dongxu Li, Silvio Savarese, and Li Fei-Fei. 2023a. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *arXiv preprint arXiv:2301.12597*.

Tianqi Li, Guansong Pang, Xiao Bai, Wenjun Miao, and Jin Zheng. 2024b. Learning transferable negative prompts for out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17584–17594.

Yixia Li, Boya Xiong, Guanhua Chen, and Yun Chen. 2024c. Setar: Out-of-distribution detection with selective low-rank approximation. *arXiv preprint arXiv:2406.12629*.

Yuanze Li, Haolin Wang, Shihao Yuan, Ming Liu, Debin Zhao, Yiwen Guo, Chen Xu, Guangming Shi, and Wangmeng Zuo. 2023b. Myriad: Large multimodal model by applying vision experts for industrial anomaly detection. *arXiv preprint arXiv:2310.19070*.

Zhong Li, Yuxuan Zhu, and Matthijs Van Leeuwen. 2023c. A survey on explainable anomaly detection. *ACM Transactions on Knowledge Discovery from Data*, 18(1):1–54.

Bo Liu, Li-Ming Zhan, Zexin Lu, Yujie Feng, Lei Xue, and Xiao-Ming Wu. 2024a. How good are LLMs at out-of-distribution detection? In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 8211–8222, Torino, Italia. ELRA and ICCL.

Jiaqi Liu, Guoyang Xie, Jinbao Wang, Shangnian Li, Chengjie Wang, Feng Zheng, and Yaochu Jin. 2024b. Deep industrial image anomaly detection: A survey. *Machine Intelligence Research*, 21(1):104–135.

Jun Liu, Chaoyun Zhang, Jiaxu Qian, Minghua Ma, Si Qin, Chetan Bansal, Qingwei Lin, Saravan Rajmohan, and Dongmei Zhang. 2024c. Large language models can deliver accurate and interpretable time series anomaly detection. *arXiv preprint arXiv:2405.15370*.

Yilun Liu, Shimin Tao, Weibin Meng, Feiyu Yao, Xiaofeng Zhao, and Hao Yang. 2024d. Logprompt: Prompt engineering towards zero-shot and interpretable log analysis. In *Proceedings of the 2024 IEEE/ACM 46th International Conference on Software Engineering: Companion Proceedings*, pages 364–365.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Hui Lv and Qianru Sun. 2024. Video anomaly detection and explanation via large language models. *arXiv preprint arXiv:2401.05702*.

Yifei Ming, Ziyang Cai, Jiuxiang Gu, Yiyou Sun, Wei Li, and Yixuan Li. 2022. Delving into out-of-distribution detection with vision-language representations. *Advances in neural information processing systems*, 35:35087–35102.

Yifei Ming and Yixuan Li. 2024. How does fine-tuning impact out-of-distribution detection for vision-language models? *International Journal of Computer Vision*, 132(2):596–609.

Atsuyuki Miyai, Jingkang Yang, Jingyang Zhang, Yifei Ming, Yueqian Lin, Qing Yu, Go Irie, Shafiq Joty, Yixuan Li, Hai Li, et al. 2024a. Generalized out-of-distribution detection and beyond in vision language model era: A survey. *arXiv preprint arXiv:2407.21794*.

Atsuyuki Miyai, Jingkang Yang, Jingyang Zhang, Yifei Ming, Qing Yu, Go Irie, Yixuan Li, Hai Li, Ziwei Liu, and Kiyoharu Aizawa. 2024b. Unsolvable problem detection: Evaluating trustworthiness of vision language models. *arXiv preprint arXiv:2403.20331*.

Atsuyuki Miyai, Qing Yu, Go Irie, and Kiyoharu Aizawa. 2023. Zero-shot in-distribution detection in multi-object settings using vision-language foundation models. *arXiv preprint arXiv:2304.04521*.

Atsuyuki Miyai, Qing Yu, Go Irie, and Kiyoharu Aizawa. 2024c. Locoop: Few-shot out-of-distribution detection via prompt learning. *Advances in Neural Information Processing Systems*, 36.

Jun Nie, Yonggang Zhang, Zhen Fang, Tongliang Liu, Bo Han, and Xinmei Tian. 2024. Out-of-distribution detection with negative prompts. In *The Twelfth International Conference on Learning Representations*.

OpenAI. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. 2021. Deep learning for anomaly detection: A review. *ACM computing surveys (CSUR)*, 54(2).

Sangha Park, Jisoo Mok, Dahuin Jung, Saehyung Lee, and Sungroh Yoon. 2023. On the powerfulness of textual outlier exposure for visual ood detection. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Jiaxing Qi, Shaohan Huang, Zhongzhi Luan, Shu Yang, Carol Fung, Hailong Yang, Depei Qian, Jing Shang, Zhiwen Xiao, and Zhihui Wu. 2023. Loggpt: Exploring chatgpt for log-based anomaly detection. In *2023 IEEE International Conference on High Performance Computing & Communications, Data Science & Systems, Smart City & Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*. IEEE.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.

Pranab Sahoo, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal, and Aman Chadha. 2024. A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv preprint arXiv:2402.07927*.

Mohammadreza Salehi, Hossein Mirzaei, Dan Hendrycks, Yixuan Li, Mohammad Hossein Rohban, and Mohammad Sabokrou. 2021. A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges. *arXiv preprint arXiv:2110.14051*.

Rohan Sinha, Amine Elhafsi, Christopher Agia, Matthew Foutter, Edward Schmerling, and Marco Pavone. 2024. Real-time anomaly detection and reactive planning with large language models. *arXiv preprint arXiv:2407.08735*.

Jing Su, Chufeng Jiang, Xin Jin, Yuxin Qiao, Tingsong Xiao, Hongda Ma, Rong Wei, Zhi Jing, Jiajun Xu, and Junhong Lin. 2024. Large language models for forecasting and anomaly detection: A systematic literature review. *arXiv preprint arXiv:2402.10350*.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Thomas Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Hualiang Wang, Yi Li, Huifeng Yao, and Xiaomeng Li. 2023. Clipn for zero-shot ood detection: Teaching clip to say no. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1802–1812.

Wenhai Wang, Zhe Chen, Xiaokang Chen, Jiannan Wu, Xizhou Zhu, Gang Zeng, Ping Luo, Tong Lu, Jie Zhou, Yu Qiao, et al. 2024. Visionllm: Large language model is also an open-ended decoder for vision-centric tasks. *Advances in Neural Information Processing Systems*, 36.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Ning Wu, Ming Gong, Linjun Shou, Shining Liang, and Daxin Jiang. 2023. Large language models are diverse role-players for summarization evaluation. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 695–707. Springer.

Albert Xu, Xiang Ren, and Robin Jia. 2023. Contrastive novelty-augmented learning: Anticipating outliers with large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11778–11801, Toronto, Canada. Association for Computational Linguistics.

Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. 2024a. Generalized out-of-distribution detection: A survey. *International Journal of Computer Vision*.

Yuchen Yang, Kwonjoon Lee, Behzad Dariush, Yinzhi Cao, and Shao-Yuan Lo. 2024b. Follow the rules: Reasoning for video anomaly detection with large language models. *arXiv preprint arXiv:2407.10299*.

Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. 2023. A survey on multimodal large language models. *arXiv preprint arXiv:2306.13549*.

Luca Zanella, Willi Menapace, Massimiliano Mancini, Yiming Wang, and Elisa Ricci. 2024. Harnessing large language models for training-free video anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Andi Zhang, Tim Z Xiao, Weiyang Liu, Robert Bamler, and Damon Wischik. 2024a. Your finetuned large language model is already a powerful out-of-distribution detector. *arXiv preprint arXiv:2404.08679*.

Huaxin Zhang, Xiaohao Xu, Xiang Wang, Jialong Zuo, Chuchu Han, Xiaonan Huang, Changxin Gao, Yuehuan Wang, and Nong Sang. 2024b. Holmes-vad: Towards unbiased and explainable video anomaly detection via multi-modal llm. *arXiv preprint arXiv:2406.12235*.

Jiangning Zhang, Xuhai Chen, Zhucun Xue, Yabiao Wang, Chengjie Wang, and Yong Liu. 2023. Exploring grounding potential of vqa-oriented gpt-4v for zero-shot anomaly detection. *arXiv preprint arXiv:2311.02612*.

Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. 2022. Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9).

Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. 2024. AnomalyCLIP: Object-agnostic prompt learning for zero-shot anomaly detection. In *The Twelfth International Conference on Learning Representations*.

Jiaqi Zhu, Shaofeng Cai, Fang Deng, and Junran Wu. 2024. Do llms understand visual anomalies? uncovering llm capabilities in zero-shot anomaly detection. *arXiv preprint arXiv:2404.09654*.

Jiawen Zhu and Guansong Pang. 2024. Toward generalist anomaly detection via in-context residual learning with few-shot sample prompts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17826–17836.