# A Survey of Out-of-domain Detection, New Intent Discovery, Generalized Category Discovery, and Conversational AI

www.surveyx.cn


## Abstract

This survey paper explores the interrelated fields of Out-of-domain Detection, New Intent Discovery, Generalized Category Discovery, Conversational AI, Intent Recognition, and Natural Language Processing, emphasizing their collective role in advancing artificial intelligence systems. The survey synthesizes state-of-the-art techniques to address challenges such as handling out-of-domain inputs, discovering new user intents, and categorizing data into evolving groups. It highlights the integration of methodologies like multimodal intent understanding and joint models for intent detection and slot filling, which enhance AI's ability to process and analyze human language. The paper systematically reviews recent advancements, including the use of transformer-based architectures, self-supervised learning, and domain adaptation strategies, which have significantly improved the robustness and adaptability of AI systems across various applications. Furthermore, it discusses the implications of these advancements for improving human-computer interaction, particularly in task-oriented dialogue systems and abuse detection. The survey concludes by identifying emerging trends and future research directions, such as enhancing model robustness against domain shifts, refining uncertainty quantification techniques, and developing more efficient frameworks for new intent discovery. These insights aim to advance the field of AI, ultimately enhancing machine understanding and interaction capabilities in dynamic and complex environments.


## 1 Introduction

### 1.1 Importance of Interrelated Fields

The interconnectedness of Out-of-domain Detection, New Intent Discovery, Generalized Category Discovery, Conversational AI, Intent Recognition, and Natural Language Processing is crucial for enhancing artificial intelligence systems. This synergy fosters the creation of robust models capable of comprehending and interacting with complex human language across various contexts. For example, integrating Out-of-domain Detection with Conversational AI is vital for developing AI agents that autonomously learn and adapt in open environments, thereby improving their engagement in meaningful dialogues [1]. Additionally, the interplay among New Intent Discovery, Slot-Value Discovery, and Joint Ontology Expansion significantly enhances conversational agents' performance by enabling them to manage evolving user queries more effectively [2]. The challenges posed by Out-of-domain and Out-of-scope (OOD/OOS) inputs in task-oriented dialogue systems necessitate a comprehensive approach that merges Intent Detection and Discovery for seamless interactions [3]. Furthermore, the relationship between intent detection and slot filling is essential for improving the efficacy of virtual assistants, as accurate interpretation of user intents is fundamental for delivering relevant responses. The development of socialbots capable of coherent conversations in open-domain dialogue systems further emphasizes the importance of these interconnected fields [4]. Moreover, integrating multimodal intent understanding, which utilizes various modalities for analyzing human

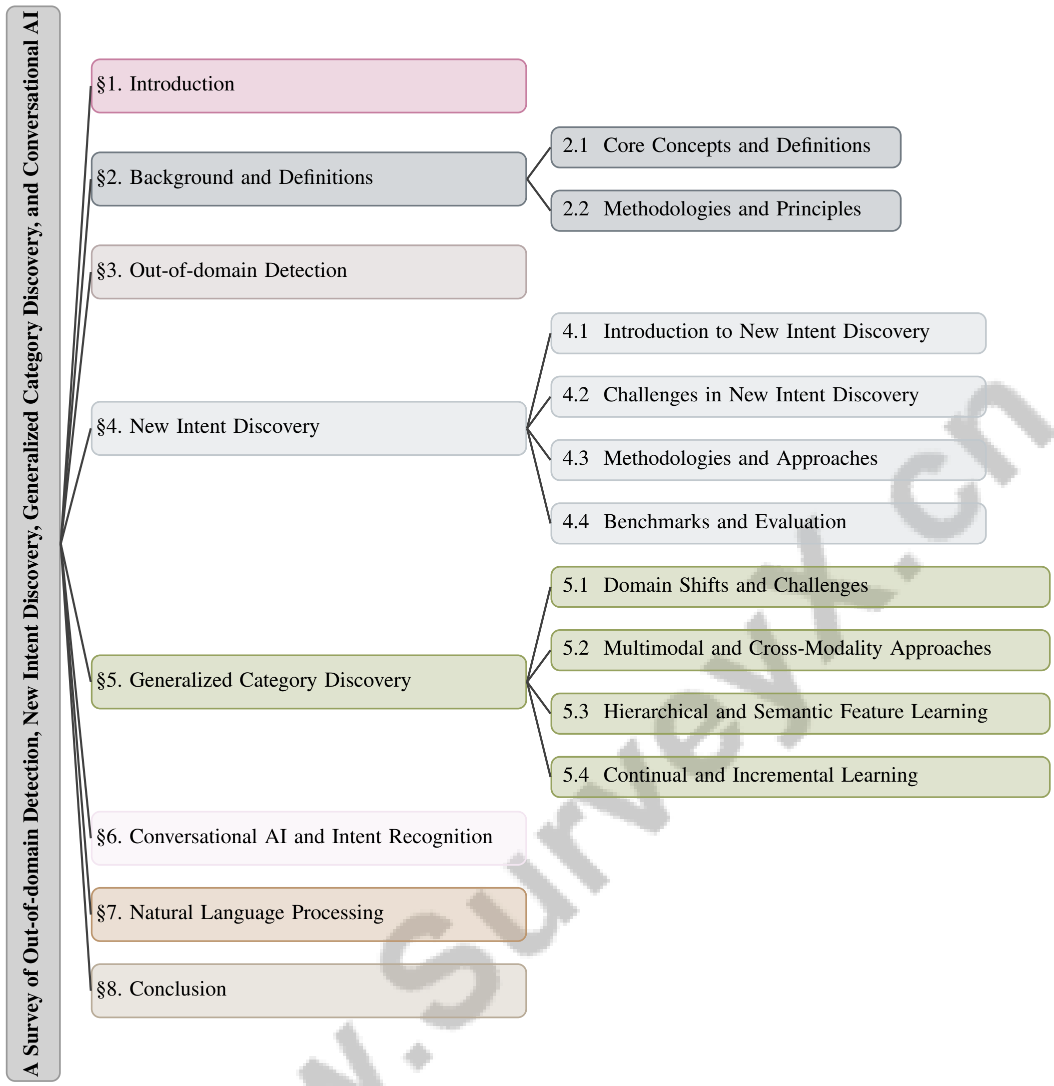


Figure 1: chapter structure

language, underscores the necessity for a holistic approach to processing user inputs [5]. The interconnectedness of user input intent classification, out-of-domain input detection, new intent discovery, and continual learning is critical for enhancing QA systems [6]. This interconnectedness is essential for developing AI systems that generalize to unseen distributions and adapt to new information in real-time, ultimately improving human-robot interactions and the overall user experience.

### 1.2 Objectives of the Survey

This survey aims to provide a comprehensive review of the interrelated fields of Out-of-domain Detection, New Intent Discovery, Generalized Category Discovery, Conversational AI, Intent Recognition, and Natural Language Processing. By synthesizing state-of-the-art techniques and methodologies, it addresses inherent challenges and limitations within these domains. Specifically, the survey explores advancements in joint multiple intent detection and slot filling in spoken language understanding, focusing on the limitations of existing autoregressive models [7]. It also proposes methods for detecting out-of-domain intents and discovering new intents from user inputs that fall outside predefined domains, thereby enhancing chatbot system robustness [3]. The evolution of joint models for intent detection and slot filling is highlighted, emphasizing their interdependencies and the necessity for integrated approaches [8]. Additionally, the survey reviews ontology expansion techniques to improve conversational understanding [2] and addresses deficiencies in QA systems by enhancing



AI-generated, for reference only.

their robustness and efficiency through advanced methodologies [6]. Another objective is to examine domain generalization, focusing on developing models capable of generalizing to unseen distributions [9]. Lastly, it tackles challenges related to the inability of existing methods to learn discriminative representations for fine-grained intent classes and their struggles with out-of-domain data from unseen open classes [5]. By achieving these objectives, the survey aims to improve the understanding and interaction capabilities of AI systems, thereby advancing the field of artificial intelligence.

### 1.3 Overview of Main Topics

This survey investigates several interconnected domains that collectively enhance artificial intelligence systems. Key topics include Out-of-domain Detection, which identifies inputs outside predefined categories to ensure robust AI performance in open environments. New Intent Discovery focuses on identifying previously undefined user intents to adapt to evolving user needs, closely related to New Slot-Value Discovery and Joint Ontology Expansion, pivotal for managing dynamic user queries [2]. Generalized Category Discovery addresses categorizing data into evolving groups, facilitating the handling of both known and novel categories. The survey also covers Conversational AI, essential for enabling human-like dialogue, and Intent Recognition, which seeks to understand user intentions from spoken or written language. A significant aspect is the integration of intent classification and slot filling, traditionally treated independently, and the exploration of joint models that enhance conversational agents' interaction capabilities [8]. Finally, the survey examines Natural Language Processing techniques crucial for processing and analyzing human language, ultimately enhancing machine understanding and interaction. These topics aim to advance the development of AI systems capable of generalizing to unseen distributions and adapting to new information in real-time.

### 1.4 Structure of the Survey

This survey is systematically organized into key sections to provide a comprehensive understanding of the interconnected fields of Out-of-domain Detection, New Intent Discovery, Generalized Category Discovery, Conversational AI, Intent Recognition, and Natural Language Processing. It begins with an **Introduction** (Section 1), highlighting the importance of these interrelated fields, defining the survey's objectives, and offering an overview of the main topics discussed.

Following the introduction, **Section 2** delves into the **Background and Definitions**, elucidating core concepts and methodologies pertinent to the survey topics, thereby establishing foundational principles for subsequent discussions.

**Section 3** focuses on **Out-of-domain Detection**, exploring techniques, challenges, advancements, and real-world applications related to detecting inputs outside predefined categories. Ensuring effective AI operation in open environments is essential for robust out-of-domain intent classification and new intent discovery, allowing systems to identify and adapt to previously unseen user intents and inputs, thus enhancing performance in real-world applications [10, 11, 12, 13, 14].

In **Section 4**, the survey examines **New Intent Discovery**, discussing its significance, challenges, methodologies, and evaluation metrics, which are vital for understanding how AI systems can adapt to evolving user needs by identifying previously undefined user intents.

**Section 5** addresses **Generalized Category Discovery**, highlighting methods for categorizing data into evolving groups and the importance of handling dynamic and complex datasets, including domain shifts, multimodal approaches, hierarchical feature learning, and continual learning.

The role of **Conversational AI and Intent Recognition** is explored in **Section 6**, discussing the integration of modalities, ontology expansion, advancements in intent recognition, uncertainty detection, and innovative frameworks for enhanced interaction. This section underscores the importance of facilitating human-like dialogue and understanding user intentions.

**Section 7** provides an overview of **Natural Language Processing**, discussing recent advancements and applications in abuse detection, sentiment analysis, intent recognition, task-oriented dialogue systems, uncertainty quantification, and cross-domain applications, emphasizing NLP's impact on enhancing machine understanding and interaction.

The survey culminates in , which offers a comprehensive overview of key findings and insights derived from the research, delving into the interconnections between various fields, highlighting the significance of new intent discovery in natural language processing and its applications, such as clustering frameworks and zero-shot data generation. It also emphasizes emerging trends, such as addressing long-tailed distributions in intent classification, and outlines future research directions to enhance the robustness and adaptability of models in real-world scenarios [15, 16, 10, 17, 18]. This comprehensive structure ensures a thorough exploration of the topics, advancing the field of artificial intelligence.The following sections are organized as shown in Figure 1.

## 2 Background and Definitions

### 2.1 Core Concepts and Definitions

Ontology Expansion is crucial in conversational AI, allowing systems to adapt by incorporating new user intents and values, enhancing flexibility in dynamic environments [2]. Intent Recognition, particularly Out-of-Domain (OOD) detection, utilizes labeled datasets to train models that process inputs beyond predefined categories [3]. Domain Generalization (DG) develops models that generalize across unseen domains without accessing target data during training, ensuring effective operation in diverse settings. Novel Category Discovery, especially within Generalized Category Discovery (GCD), leverages labeled data from similar classes to identify categories within unlabeled data [19]. Multimodal Generalized Category Discovery (MM-GCD) integrates multimodal inputs to classify known and unknown categories, emphasizing feature and output space alignment for comprehensive recognition [20].

Traditional novel category discovery methods often fail in GCD due to assumptions about unlabeled data, highlighting the need for advanced approaches [21]. The Cityscapes-GCD dataset, with 15 base and 4 novel classes, simulates real-world scenarios to evaluate clustering capabilities using partial labeled category information [22]. Detecting out-of-domain utterances in dialog systems prevents erroneous actions by agents [23], and benchmarks for out-of-distribution performance analyze model generalization to unseen data [10]. Benchmarks evaluate models' handling of user inputs that do not conform to predefined categories or intents in task-oriented dialog systems [24].

In spoken language understanding, intent detection and slot filling are key for interpreting user inputs [25]. DOMAIN-INTENT-SLOT schemas and automatic intent-slot induction (AISI) enhance these processes, improving recognition accuracy [26]. Uncertainty quantification, out-of-distribution inputs, and distance awareness are crucial for high-quality uncertainty estimation, allowing AI systems to manage unexpected inputs [27]. Novelty detection, the ability to identify and characterize new situations, is essential for autonomous learning in open environments [1]. Additionally, enabling robots to navigate and locate object categories in unfamiliar environments while generalizing across contexts highlights AI's complexity in interacting with diverse datasets [28]. Explainable Artificial Intelligence (XAI) methods are necessary for transparency in deep learning models, particularly in drug discovery [29]. These core concepts form the foundation of the survey, emphasizing AI's multifaceted engagement with complex datasets.

### 2.2 Methodologies and Principles

The methodologies underpinning the survey topics encompass strategies to enhance AI systems' adaptability and robustness across various domains. Hybrid architectures, such as integrating Variational Autoencoders (VAE) with kernel-PCA and HDBSCAN, are pivotal for OOD intent detection and discovery, combining dimensionality reduction and clustering to improve recognition [3]. These methods are evaluated using metrics assessing the model's ability to differentiate in-domain from out-of-domain examples, focusing on false positive rates and the ROC curve area.

In uncertainty quantification, Bayesian Neural Networks (BNNs) face challenges in estimating aleatoric and epistemic uncertainties, critical for predictive models [30]. The minimax learning problem formalizes uncertainty quantification, addressing difficulties in estimating predictive uncertainty in deep learning models [27]. The variational nested dropout (VND) framework employs a probabilistic approach to adjust dropout rates dynamically during training, enhancing the model's capacity to learn the importance of network components [31].

Domain Generalization (DG) methodologies, including domain alignment, meta-learning, data augmentation, and ensemble learning, are essential for developing models that generalize across unseen domains. These techniques enable AI systems to operate in diverse settings by aligning



AI-generated, for reference only.

feature distributions and leveraging meta-learning for adaptability. The challenge of maintaining performance amid domain shifts, particularly in wireless communications, highlights the necessity for robust models [32].

Novelty detection is addressed through frameworks like Self-Initiated Open World Learning (SOL), emphasizing self-initiated learning for continuous adaptation post-deployment without human intervention [1]. ProtoInfoMax enhances discriminative capabilities by maximizing mutual information between in-domain and out-of-domain samples, mitigating overconfidence errors [33]. Unsupervised methods leveraging rich linguistic information across transformer layers have shown effectiveness in distinguishing in-domain from out-of-domain samples, bolstering AI systems' robustness.

In intent recognition, existing spoken language understanding (SLU) models often struggle to incorporate intent information effectively to enhance slot filling performance [25]. The Global Locally Graph Interaction Network (GL-GIN) exemplifies a sophisticated methodology that models local dependencies between slots and global interactions between intents and slots, enhancing performance through parallel decoding [7]. The AGIF principle focuses on dynamically filtering and integrating relevant intent information for each token, improving slot prediction accuracy [34].

These methodologies underscore the necessity for adaptive, robust, and equitable AI systems capable of responding effectively to the dynamic nature of real-world environments. This is crucial for applications like abusive language detection and conversational agents, where emerging forms of abusive language and user intents necessitate continuous model updates and retraining. Techniques like the Testing Concept Activation Vector (TCAV) for interpretability and the Multiple Novel Intent Detection (MNID) framework for efficient annotation enhance generalizability and accuracy, driving advancements in AI [35, 36].

In recent years, out-of-domain (OOD) detection has garnered significant attention due to its critical role in enhancing the robustness of machine learning systems. As illustrated in Figure 2, the hierarchical structure of OOD detection is meticulously categorized, showcasing various techniques, challenges, advancements, and real-world applications. This figure not only delineates traditional and innovative methods employed in OOD detection but also emphasizes the challenges faced in the field, recent advancements, and the implications of these developments across diverse applications. Such a comprehensive overview facilitates a deeper understanding of the multifaceted nature of OOD detection and its relevance in contemporary research and practical scenarios.

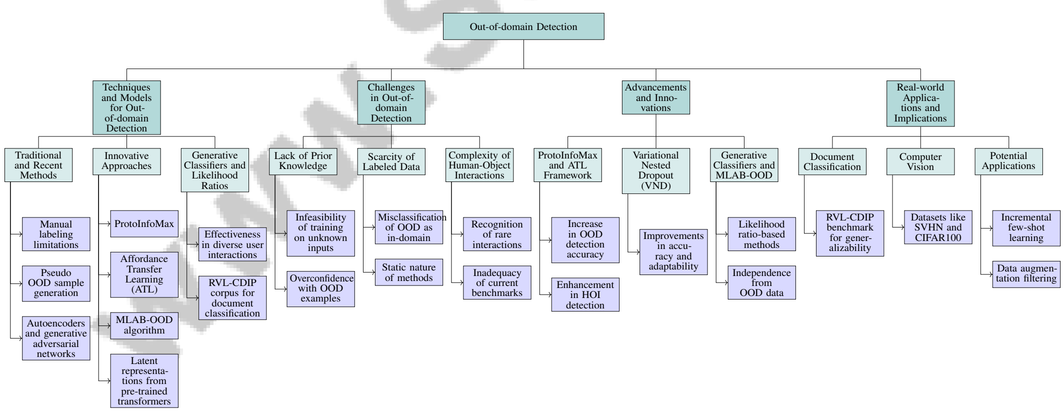


Figure 2: This figure illustrates the hierarchical structure of out-of-domain (OOD) detection, categorizing techniques, challenges, advancements, and real-world applications. It highlights traditional and innovative methods, challenges faced in OOD detection, recent advancements in the field, and its implications across various applications.

## 3 Out-of-domain Detection

### 3.1 Techniques and Models for Out-of-domain Detection

Out-of-domain (OOD) detection is essential for enhancing AI systems' robustness, especially in task-oriented dialogue frameworks where user inputs may exceed predefined intents. This is critical in

| Method Name | Detection Techniques | Data Utilization | Model Generalization |
|---|---|---|---|
| PIM[33] | Mutual Information Maximization | Ood Data | Low-resource Settings |
| ATL[37] | Affordance Transfer Learning | Novel Hoi Samples | Novel Objects |
| MLAB-OOD[38] | Metric Learning | Pseudo Ood Samples | Unseen Categories |
| MDF[39] | Mahalanobis Distance Features | Unsupervised In-domain Data | Unseen Input Categories |

Table 1: This table presents a selection of advanced techniques and models for out-of-domain (OOD) detection, highlighting their detection techniques, data utilization strategies, and model generalization capabilities. The methods include ProtoInfoMax, Affordance Transfer Learning (ATL), MLAB-OOD, and Mahalanobis Distance Features (MDF), each contributing to enhanced OOD detection performance in various contexts.

natural language understanding (NLU) for dialogue systems, where failing to recognize unsupported OOD inputs can lead to operational failures. Traditional OOD detection methods often rely on manually labeled samples, limiting their applicability due to scarce in-domain training data. Recent advances, such as models generating high-quality pseudo OOD samples that resemble in-domain utterances, show promise in improving detection performance. Techniques like autoencoders and generative adversarial networks leverage both labeled and unlabeled data, enhancing detection accuracy and resilience against unexpected inputs [11, 40].

Innovative approaches address traditional methods' limitations. Prototypical Networks with Mutual Information Maximization (ProtoInfoMax) optimize alignment between in-domain and OOD samples, enhancing discriminative capacity in low-resource settings [33]. The Affordance Transfer Learning (ATL) framework improves detection by transferring affordance representations from known to novel objects, enhancing generalization to unseen categories [37]. The MLAB-OOD algorithm uses metric learning and adaptive boundary setting to delineate decision boundaries between in-domain and OOD samples, improving detection accuracy [38]. Leveraging latent representations from all layers of pre-trained transformers has proven effective in detecting OOD samples, utilizing deep contextual understanding [39].

Generative classifiers and likelihood ratio-based methods have been validated for their effectiveness in detecting OOD inputs, showcasing their capability to manage diverse user interactions [24]. The RVL-CDIP corpus serves as a benchmark for document classification, highlighting the need for evaluation on out-of-distribution documents to ensure model generalizability [10].

As illustrated in Figure 3, the categorization of techniques and models for out-of-domain (OOD) detection encompasses traditional methods, innovative approaches, and generative and likelihood-based methods. Each category encapsulates specific techniques that contribute to the enhancement of OOD detection in AI systems. Advances in OOD detection methodologies underscore the need for robust systems capable of navigating dynamic user interactions, especially in low-resource settings where traditional methods may falter [5, 11, 41]. Table 1 provides a comprehensive overview of various innovative approaches to out-of-domain detection, detailing the methodologies employed, data utilization, and generalization capabilities of each model.

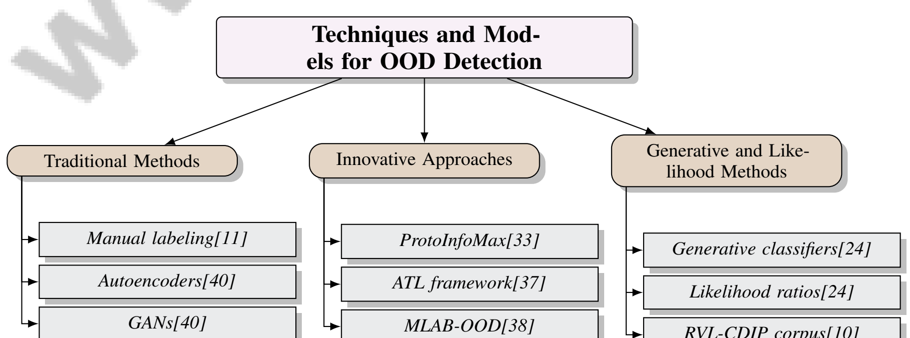


Figure 3: This figure illustrates the categorization of techniques and models for out-of-domain (OOD) detection, highlighting traditional methods, innovative approaches, and generative and likelihood-based methods. Each category encapsulates specific techniques that contribute to the enhancement of OOD detection in AI systems.



AI-generated, for reference only.

## 3.2 Challenges in Out-of-domain Detection

Out-of-domain (OOD) detection faces challenges that complicate identifying inputs outside predefined categories. A primary challenge is the lack of prior knowledge about OOD inputs, making it infeasible to train a domain classifier directly on unknown inputs [39]. Over-reliance on in-domain data during training often leads to models exhibiting overconfidence with OOD examples, as existing benchmarks fail to assess performance adequately on out-of-distribution inputs [10].

The scarcity of large-scale labeled data for OOD scenarios complicates matters, as models trained without sufficient OOD examples misclassify OOD inputs as in-domain. This challenge is intensified by the static nature of methods assuming fixed intents, limiting adaptability to unforeseen OOD inputs. Such rigidity poses challenges in applications like abusive language detection and document classification, where interactions are dynamic and influenced by social and political events. As new forms of abusive language emerge, models trained on static datasets may fail to adapt and accurately identify evolving threats. Classifier performance can deteriorate with out-of-distribution inputs, underscoring the need for continuous updates to maintain robustness [10, 36].

The complexity of human-object interactions necessitates models recognizing both common and rare interactions that do not conform to established categories, akin to advanced activity prediction models in drug discovery [42, 43, 17]. Current benchmarks' inadequacy, often reliant on small datasets and synthetic OOD examples, complicates evaluation. The choice of loss functions is critical, as many approaches do not surpass a well-tuned empirical risk minimization baseline, highlighting the need for sophisticated loss function designs capturing OOD detection nuances.

The challenges in OOD detection underscore the need for innovative methodologies and comprehensive benchmarks to navigate these complexities. Given the prevalence of low-resource scenarios and traditional methods' limitations, developing robust systems capable of identifying OOD inputs without extensive labeled in-domain data is essential. Recent advancements, such as OOD-resistant Prototypical Networks and pre-trained language models, have shown promise in enhancing OOD detection accuracy, even amidst distributional shifts. These systems must handle diverse unforeseen inputs, particularly in dynamic environments with uncontrolled and potentially malicious data [44, 11, 41].

## 3.3 Advancements and Innovations

| Method Name | Detection Techniques | Adaptability and Generalization | Performance Metrics |
|---|---|---|---|
| PIM[33] | Protoinfomax | Novel Object Recognition | Equal Error Rate |
| ATL[37] | Affordance Transfer Learning | Novel Object Recognition | Mean Average Precision |
| VND[31] | Variational Nested Dropout | Dynamic Adaptation | Expected Calibration Error |
| MLAB-OOD[38] | Metric Learning | New And Unseen | Accuracy, Macro F1-score |
| MDF[39] | Mahalanobis Distance | Unsupervised In-domain | Detection Accuracy |

Table 2: Summary of recent advancements in out-of-domain (OOD) detection methodologies, highlighting the detection techniques, adaptability, generalization capabilities, and performance metrics of various innovative methods. The table provides a comparative analysis of five notable approaches: ProtoInfoMax (PIM), Affordance Transfer Learning (ATL), Variational Nested Dropout (VND), MLAB-OOD, and Mahalanobis Distance Feature (MDF) based methods.

Recent advancements in out-of-domain (OOD) detection have led to innovative frameworks and methodologies enhancing detection capabilities across applications. The ProtoInfoMax model, integrating a prototypical network with mutual information maximization, achieves up to a 20% increase in OOD detection accuracy compared to existing methods [33]. This approach optimizes alignment between in-domain and OOD samples, enhancing discriminative capacity.

The Affordance Transfer Learning (ATL) framework improves Human-Object Interaction (HOI) detection and object affordance recognition, particularly in long-tailed settings and with novel objects [37]. This framework transfers affordance representations from known to novel objects, enhancing generalization to unseen categories.

Variational Nested Dropout (VND) demonstrates substantial improvements in accuracy, calibration, and adaptability, confirming its effectiveness in learning ordered representations within neural networks [31]. This approach allows dynamic adaptation of dropout rates during training, enhancing the model's learning of important components.

Generative classifiers and likelihood ratio-based methods have been benchmarked for their effectiveness in detecting OOD inputs, demonstrating significant performance improvements over traditional approaches [24]. These models utilize likelihood ratios to distinguish between in-domain and OOD examples, showcasing their potential to handle diverse user interactions.

The MLAB-OOD algorithm introduces a novel approach by operating independently of OOD data, significantly improving detection performance in scenarios with fewer classes [38]. This method delineates decision boundaries between in-domain and OOD samples, enhancing detection accuracy.

Unsupervised methods leveraging Mahalanobis distance features from all transformer layers improve detection performance compared to methods relying on single-layer features [39]. This approach capitalizes on language models' deep contextual understanding to enhance OOD detection capabilities.

Table 2 presents a comprehensive overview of recent advancements in OOD detection methodologies, detailing the detection techniques, adaptability, generalization, and performance metrics of various innovative methods. Recent advancements in OOD detection research illustrate the field's evolution, emphasizing the need for methodologies improving AI systems' robustness and adaptability while addressing challenges like limited in-domain training data, multimodal intent understanding, and distribution uncertainty calibration. Studies suggest leveraging pre-trained language models without fine-tuning can yield near-perfect OOD detection performance, while novel approaches like OOD-resistant Prototypical Networks and Bayesian frameworks enhance detection capabilities across datasets and scenarios. These findings highlight the importance of developing adaptable AI systems capable of reliably managing unpredictable inputs [5, 11, 41, 44, 45].

### 3.4 Real-world Applications and Implications

Out-of-domain (OOD) detection is pivotal for enhancing AI systems' robustness across real-world applications. In document classification, the RVL-CDIP benchmark, with 4,417 out-of-distribution document images, is instrumental in evaluating models' generalizability to unseen document types [10]. This capability ensures AI systems accurately classify documents in dynamic environments.

In computer vision, datasets like SVHN, CIFAR100, and cropped versions of LSUN and TinyImageNet assess OOD detection models' efficacy, providing insights into their ability to differentiate between in-domain and out-of-domain visual inputs [46]. These datasets develop systems operating effectively in real-world scenarios where visual data may not conform to expected distributions.

The potential applications of OOD detection extend to incremental few-shot learning and data augmentation filtering, as outlined in future work exploring unsupervised OOD detection methods' real-world applicability [39]. These applications are significant for developing AI systems that continuously learn and adapt to new information, enhancing performance in dynamic environments.

The practical applications and implications of OOD detection impact numerous fields by ensuring AI systems effectively handle diverse and unforeseen inputs. This capability is essential for enhancing AI systems' robustness and adaptability, particularly in dynamic environments where new forms of input, such as emerging types of abusive language or novel user intents, continuously arise. By improving AI systems' interaction with their environments, we ensure these systems remain effective and relevant, contributing to their performance and reliability in real-world applications [35, 36].

## 4 New Intent Discovery

Exploring New Intent Discovery (NID) is vital in addressing the complexities of natural language understanding. The following subsection, "Introduction to New Intent Discovery," delves into NID's foundational concepts, underscoring its importance in identifying and classifying undefined user intents within dynamic environments.

### 4.1 Introduction to New Intent Discovery

New Intent Discovery (NID) is essential for natural language understanding, focusing on identifying and classifying previously undefined user intents, especially in dynamic open-world scenarios [15]. NID enhances AI systems' adaptability, enabling question-answering and dialogue systems to accommodate new interactions without extensive retraining [6]. In task-oriented dialogue systems, discovering new intents is crucial for interaction quality and user satisfaction due to the limitations of predefined intent sets [47].

The complexity of intent recognition is highlighted by joint multiple intent detection and slot filling in spoken language understanding, where users may express multiple intents in one utterance [25]. The Generalized Intent Discovery (GID) framework addresses this by classifying labeled in-domain intents while discovering new out-of-domain types [48]. The challenge of classifying user-generated sentences into intents in open-world contexts underscores the need for robust methodologies [49].

Innovative approaches, such as integrating new slot discovery within information extraction frameworks and active learning schemes, advance NID methodologies [50]. Frameworks for dynamically identifying and annotating new intents in user queries without prior knowledge are critical for improving dialogue systems and human-computer interaction in open-world contexts [35]. The importance of robust OOD detection in conversational agents, especially in unpredictable environments, emphasizes the need for reliable NID methods [38]. Benchmarks for evaluating OOD detection methods in natural language processing are essential for enhancing intent classification reliability in task-oriented dialogue systems [24].

As illustrated in Figure 4, the hierarchical structure of New Intent Discovery (NID) underscores its significance in enhancing AI adaptability and interaction quality. This figure also highlights the challenges faced in complex recognition and open-world classification, alongside innovative approaches such as clustering frameworks and active learning schemes.

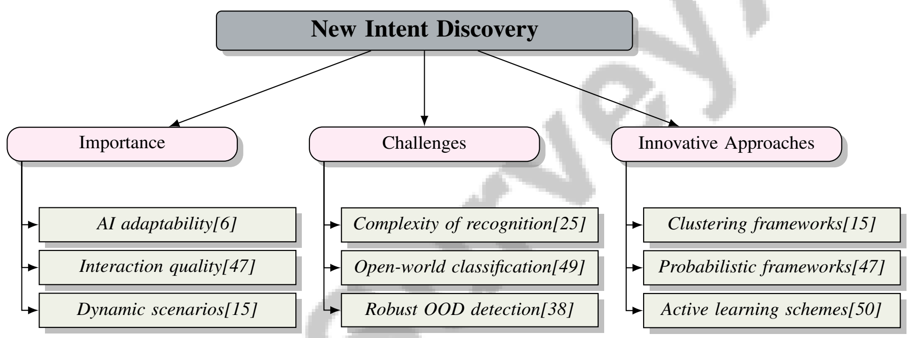


Figure 4: This figure illustrates the hierarchical structure of New Intent Discovery (NID), highlighting its importance in enhancing AI adaptability and interaction quality, the challenges faced in complex recognition and open-world classification, and innovative approaches such as clustering frameworks and active learning schemes.

## 4.2 Challenges in New Intent Discovery

Discovering new user intents presents challenges that complicate natural language understanding system development. A significant hurdle is the reliance on substantial labeled data and multi-stage training, which are impractical for dynamic user intents [49]. Methods often forget prior knowledge during discovery, leading to ineffective clustering and intent recognition [47]. This is exacerbated by domain knowledge dependency, affecting large language models' performance in recognizing OOD intents [51].

Existing gradient boosting methods inadequately estimate knowledge uncertainty in OOD detection scenarios, limiting their effectiveness in identifying new intents [52]. Adapting to new goals and tasks in real-time poses significant challenges, especially for systems like robots in open-world environments [53].

Frameworks that operate effectively without extensive human intervention in labeling are needed. IntentGPT's training-free nature and adaptability to new intents reduce human input dependency [49]. Future research could extend frameworks like OODGAT to more complex scenarios, such as few-shot and incremental learning, enhancing applicability in dynamic environments [54].

NID challenges underscore the need for innovative methodologies and robust evaluation frameworks to effectively navigate identifying both familiar and novel user intents within skewed and long-tailed distributions often present in real-world scenarios. This is underscored by the imbalanced new intent discovery (i-NID) task and the ImbaNID-Bench benchmark, which aim to enhance user intent detection using limited labeled data alongside extensive unlabeled datasets [15, 16, 55].

### 4.3 Methodologies and Approaches

NID methodologies focus on developing systems capable of identifying and categorizing previously undefined user intents. IntentGPT exemplifies a novel, training-free, model-agnostic approach leveraging large language models to classify and discover user intents using minimal labeled examples. It employs an In-Context Prompt Generator and an Intent Predictor to dynamically generate prompts and classify intents based on user utterances, showcasing prompt-based methodologies' potential in NID [49].

The Affordance Transfer Learning (ATL) method enhances detection capabilities by composing novel Human-Object Interaction samples through integrating affordance and object representations, improving generalization to novel categories [37].

Incorporating loss function design into NID methodologies significantly impacts model performance. Specialized loss functions could enhance domain generalization outcomes, suggesting a critical avenue for improving intent discovery frameworks' robustness [56].

The MNID framework utilizes clustering techniques to detect multiple novel intents and efficiently tag data points, facilitating classifier retraining. This underscores clustering's importance in managing dynamic user interactions. Similarly, the USNID framework combines unsupervised and semi-supervised learning techniques to improve intent discovery, demonstrating multiple learning paradigms' efficacy [15].

State-of-the-art models like ResNet50 and TextCNN explore NID methodologies, showcasing these models' versatility across domains. Prompt-based models have been evaluated alongside baseline models like Softmax and Siamese networks in few-shot learning scenarios, illustrating prompt-based approaches' potential in NID [57].

The Generalized Intent Discovery (GID) method offers a unified framework jointly classifying in-domain intents and discovering out-of-domain intents, emphasizing a holistic approach to intent classification and discovery [48]. The PFID framework, treating intent assignments as latent variables and utilizing the Expectation Maximization algorithm, optimizes intent discovery while preserving prior knowledge, enhancing intent recognition reliability [47].

These methodologies enhance NID by establishing comprehensive frameworks and techniques that effectively identify and categorize both known and novel user intents, particularly within imbalanced and long-tailed data distributions. By addressing prior approaches' limitations, which often assume uniform intent class distributions, these advanced techniques improve AI systems' adaptability and robustness, enabling them to better manage evolving user interactions' complexities in real-world scenarios. The i-NID task and ImbaNID benchmark support these advancements, providing a solid foundation for evaluating and refining intent recognition capabilities across diverse applications [16, 55, 58].

### 4.4 Benchmarks and Evaluation

Assessing NID methods relies on robust benchmarks and evaluation metrics to ensure effectiveness in identifying and categorizing previously undefined user intents. Table 3 presents an overview of key benchmarks utilized for evaluating New Intent Discovery (NID) methods, highlighting their significance in diverse domains and task formats. Benchmarks for evaluating NID methods include datasets like SNIPS, ATIS, BANKING, CLINC, and StackOverflow, representing various intent categories. These datasets are vital for assessing NID approaches' efficacy in real-world applications where user intents may exhibit imbalanced and long-tailed distributions. The ImbaNID-Bench benchmark enhances this evaluation landscape by simulating practical scenarios reflecting user intent classification complexities, allowing comprehensive model performance analysis across diverse intent categories [16, 10, 55, 58, 18].

Evaluation metrics like accuracy (ACC), normalized mutual information (NMI), adjusted rand index (ARI), and macro F1-score measure alignment between predicted and ground truth intents, offering insights into NID methods' clustering and classification capabilities. The SLMM method, for instance,

| Benchmark | Size | Domain | Task Format | Metric |
|---|---|---|---|---|
| DRAC[59] | 1,000,000 | Image Classification | Class Discovery | Dataset Reconstruction Accuracy, Cluster Accuracy |
| ImbaNID-Bench[16] | 9,995 | Intent Discovery | Clustering | NMI, ARI |
| NovelCraft[60] | 4,500,000 | Novelty Detection | Visual Novelty Detection | AUROC, AUPRC |
| NSD[61] | 13,084 | Dialogue Systems | Slot Detection | Span F1, Token F1 |
| ZOG[18] | 75,593 | Dialogue Systems | Utterance Generation | Accuracy, BERTScore |
| Clevr-4[42] | 8,400 | Computer Vision | Clustering | ACC |
| ImbaNID-Bench[55] | 10,000 | Intent Recognition | Intent Classification | NMI, ARI |
| OOD[23] | 30,000 | Intent Classification | Out-of-Domain Detection | AUROC, AUPROOD |

Table 3: Table illustrating various benchmarks used for evaluating New Intent Discovery (NID) methods, detailing their size, domain, task format, and evaluation metrics. These benchmarks encompass a range of applications including image classification, novelty detection, and dialogue systems, providing a comprehensive framework for assessing the efficacy of NID approaches across different domains.

uses accuracy and macro F1-score to assess performance, highlighting these metrics' importance in evaluating intent discovery frameworks [62].

Methodologies like SLMM and IntentGPT are validated through experiments on benchmark datasets, demonstrating their potential for effective intent discovery in dynamic environments [49]. Evaluating methods like PFID on datasets such as CLINC and BANKING underscores robust evaluation metrics' significance in comparing performance against various baselines [47].

Metrics like Dataset Reconstruction Accuracy and Cluster Accuracy provide additional performance measures, indicating how well new intent discovery methods can reconstruct and cluster data points [59]. These metrics are crucial for assessing NID methods' efficacy in dynamic and evolving user interaction contexts.

Combining diverse benchmarks and comprehensive evaluation metrics provides a solid framework for assessing and advancing NID methods. Future research should prioritize enhancing NID methods' interpretability, integrating them with large language models to leverage powerful capabilities, and reducing computational complexity. This focus aims to improve NID techniques' robustness and generalizability in real-world applications, particularly in handling imbalanced and long-tailed user intent distributions and language and intent categories' evolving nature in dynamic environments [16, 41, 58, 36, 63].

# 5 Generalized Category Discovery

## 5.1 Domain Shifts and Challenges

Generalized Category Discovery (GCD) encounters significant challenges due to domain shifts, particularly in differentiating known from novel categories within unlabeled datasets. A key issue is imbalanced classification performance, where models typically excel with previously learned classes over new ones due to limited labeled data for novel categories [64]. This problem is exacerbated by the intertwined training of known and novel categories, hindering the learning of discriminative features and the transfer of category-specific knowledge [65].

A core challenge is balancing new class discovery with old class retention in rehearsal-free settings, where catastrophic forgetting is common [66]. Performance degradation may occur when new unlabeled data includes previously seen categories [67]. Class distribution mismatches in unlabeled datasets, containing both seen and novel classes, further complicate learning as unreliable pseudo-labels bias predictions towards seen classes, particularly in long-tailed distributions [68].

Continual Category Discovery (CCD) involves training models on initial labeled data and subsequently on unlabeled streams to discover new classes while retaining previous knowledge [69]. This is complicated by the need to learn representations for novel categories without labeled examples, often causing performance drops [70]. Traditional methods struggle with classifying old categories and discovering new ones from unlabeled data due to overfitting and representation challenges [71]. Innovative loss functions, like the Implicit Taylor Loss (ITL), offer solutions by enabling effective loss function learning through implicit gradients [56].

These challenges highlight the need for innovative GCD methodologies to navigate domain shifts. Recent advancements, including adapter-tuning techniques and novel loss strategies, show promise in balancing generalization and adaptability, emphasizing the development of models that can classify known categories while discovering and adapting to new ones from unlabeled datasets [71, 72, 56, 9, 64].

### 5.2 Multimodal and Cross-Modality Approaches

Multimodal integration in category discovery draws from human cognitive processes using varied sensory inputs for recognition. By combining textual and visual data, models enhance category discovery capabilities, improving AI systems' robustness and adaptability [17]. This integration fosters a comprehensive understanding of complex datasets, leveraging rich information from both text and images.

Active Generalized Category Discovery (AGCD) employs active learning strategies to improve category discovery. Through iterative active labeling, models query labels for the most informative samples, refining their understanding of known and novel categories [64]. This process enhances generalization across diverse datasets by continuously updating the knowledge base.

Catastrophic forgetting remains a significant challenge in multimodal and cross-modality approaches, as models often lose previously acquired knowledge when exposed to new, unlabeled data [69]. This issue is pronounced when learning robust representations for both old and new classes, as existing methods may disrupt pretrained knowledge during fine-tuning [71].

Innovative methodologies that integrate diverse modalities while retaining existing knowledge are crucial for addressing challenges in multimodal data analysis. This integration can enhance classification of complex, unlabeled data through frameworks like TextGCD and approaches that synthesize pseudo text embeddings, improving performance in tasks such as Generalized Category Discovery and multimodal intent understanding [73, 5, 2, 36, 17]. By leveraging cross-modality approaches, models achieve a more holistic understanding of data, enhancing their ability to discover and classify categories in dynamic environments.

### 5.3 Hierarchical and Semantic Feature Learning

Hierarchical and semantic feature learning is crucial for Generalized Category Discovery (GCD), focusing on structured and meaningful representations that enhance category discernment and classification accuracy. The Decoupled Prototypical Network (DPN) addresses GCD challenges by decoupling the training of known and novel categories, improving generalization and discriminative ability [65]. This approach highlights the importance of separating learning processes to effectively handle novel category discovery complexities.

The Happy framework introduces a debiased learning approach with clustering-guided initialization and soft entropy regularization, enhancing clustering performance for new classes [66]. This method underscores the significance of reducing bias in clustering to improve novel category discovery.

The Contextual Information Guided Semi-Supervised Learning Model (CIG-SL) integrates contextual information at both instance-level and cluster-level, facilitating the learning of discriminative features for novel categories in GCD [70]. By leveraging contextual cues, this model enhances the representation learning process, making it more adaptable to dynamic data environments.

The SimGCD method employs self-distillation and entropy regularization to boost classification performance, addressing class-balance bias that can hinder performance in imbalanced scenarios [68]. This underscores the importance of balancing class distributions for robust category discovery.

AdaptGCD incorporates adapter tuning into the GCD framework, enhancing adaptability to new tasks while preserving pretrained knowledge [71]. This integration is vital for maintaining existing knowledge integrity while adapting to novel categories.

The introduction of an adaptive sampling strategy in the Active Generalized Category Discovery (AGCD) method considers novelty, informativeness, and diversity, alongside a stable label mapping algorithm to address GCD's clustering nature [64]. This ensures that the most informative samples are selected for training, enhancing the model's ability to discover new categories.

The Parametric Information Maximization approach addresses class-balance bias in traditional methods, optimizing information maximization to improve the model's capacity to handle diverse category distributions [74].

Collectively, these methodologies enhance hierarchical and semantic feature learning, implementing innovative frameworks that significantly improve the model's capacity to accurately classify both known and novel categories in complex environments. The introduction of 'self-expertise' in fine-grained category discovery refines the model's ability to discern subtle differences among categories, while dual-context approaches leverage instance-level and cluster-level contextuality to bolster classification accuracy in unlabeled datasets. Additionally, the HiLo framework addresses domain shifts by extracting high-level semantic and low-level domain features, ensuring robust categorization amidst diverse data sources. Empirical results across various benchmarks demonstrate that these approaches surpass existing state-of-the-art techniques in their respective tasks [75, 70, 76].

### 5.4 Continual and Incremental Learning

Continual and incremental learning are pivotal for advancing Generalized Category Discovery (GCD), enabling models to effectively adapt to new categories while retaining knowledge of previously learned classes. The PromptCCD framework exemplifies this by utilizing Gaussian Mixture Models to facilitate continual learning and category discovery from unlabeled data streams, highlighting the importance of robust frameworks in managing dynamic datasets [69]. This approach emphasizes the necessity for models to dynamically integrate new information without compromising existing knowledge.

Decoupling training objectives for known and novel categories, as demonstrated in recent methodologies, significantly enhances targeted learning and knowledge transfer, improving adaptability in GCD tasks [65]. This decoupling allows for focused learning strategies that effectively balance discovering new categories with retaining previously acquired knowledge. Future research could refine these processes and explore integrating additional semantic information to enhance model robustness [65].

The Parametric Information Maximization (PIM) model offers state-of-the-art performance in GCD tasks by adaptively controlling the influence of the marginal entropy term, addressing class imbalance issues during optimization [74]. Its robustness in scenarios with unknown class numbers highlights its effectiveness in handling diverse category distributions. Future work could focus on refining contextual mining strategies and applying proposed methods to other complex datasets and tasks [70].

In exploring efficient fine-tuning strategies, future research could investigate approaches such as LoRA or combinations of different strategies to enhance GCD capabilities [71]. Additionally, refining density-based selection processes and exploring alternative classification strategies remain promising avenues for expanding incremental learning methods' applicability to other domains or datasets with similar challenges [67].

The methodologies discussed across these references underscore the importance of continual and incremental learning in the context of GCD. They advocate for scalable and adaptable strategies capable of effectively handling larger datasets while ensuring strong performance. Approaches such as MetaGCD and AdaptGCD address the challenges of continually discovering novel classes amidst existing known categories, utilizing meta-learning frameworks and adapter-tuning techniques to enhance adaptability and minimize forgetting. Additionally, Incremental Generalized Category Discovery (IGCD) and Online Continuous Generalized Category Discovery (OCGCD) highlight the necessity of managing dynamic data streams and mitigating catastrophic forgetting through innovative sampling and categorization methods. Finally, Active Generalized Category Discovery (AGCD) emphasizes the importance of selectively labeling new classes to improve classification performance, illustrating the multifaceted nature of GCD and the critical need for robust, adaptable methodologies in this evolving field [71, 77, 67, 78, 64]. Future research should continue to explore innovative strategies that enhance AI systems' effectiveness in dynamic and evolving environments.

# 6 Conversational AI and Intent Recognition

## 6.1 Integration of Modalities for Enhanced Dialogue

Enhancing dialogue capabilities in conversational AI systems necessitates the integration of multiple modalities, such as text, speech, and visual data, to achieve a nuanced understanding of user interactions. This multimodal integration allows systems to dynamically adapt to user preferences and contexts, improving response quality and relevance. The CML framework exemplifies this by autonomously learning commands through user engagement, showcasing the potential of combining conversational AI with autonomous learning [1]. Such integration advances human-computer interaction, enhancing user satisfaction across various applications.

## 6.2 Ontology Expansion and New Intent Discovery

Ontology expansion is crucial for the adaptability of conversational AI systems, enabling the discovery of new intents by extending existing frameworks to accommodate evolving user interactions [6]. Addressing the imbalanced nature of new intent discovery, frameworks like ADVIN leverage pre-trained language models and knowledge transfer to identify new intents from unlabeled corpora [79]. Zero-shot and one-shot generation methods enhance dialogue systems' adaptability by reducing reliance on extensive labeled datasets, while automated processes like RCAP streamline intent-slot schema extraction [26]. Active learning strategies, such as MNID's dual tagging, improve training sample quality and intent discovery efficiency [35]. Integrating implicit and explicit debiasing processes, as in the FER-GCD framework, enhances bias management in intent discovery [80]. These advancements underscore the necessity for dynamic methodologies that empower AI systems to adapt to real-time user intents, with techniques like multi-task pre-training and contrastive learning improving out-of-scope detection and intent recognition accuracy [2, 81, 82].

## 6.3 Advancements in Intent Recognition Techniques

Recent advancements in intent recognition have significantly enhanced conversational AI's ability to accurately interpret user intentions. Hybrid systems, like SetFit, integrate large language models with contrastively fine-tuned sentence transformers, achieving near-native LLM accuracy while reducing latency [83, 41, 84, 82, 85]. The Stack-Propagation framework, leveraging BERT, has improved intent detection and slot filling accuracy in spoken language understanding tasks [25]. DIARC architecture advances intent recognition in human-robot interactions, emphasizing robust system architectures [53]. The SNGP model enhances uncertainty estimation, improving intent recognition reliability [27]. These techniques expand AI capabilities across diverse applications, emphasizing the broader impact of intent recognition advancements beyond traditional conversational AI [80, 79, 81].

## 6.4 Uncertainty and Out-of-Domain Detection

Uncertainty quantification is critical for improving out-of-domain (OOD) detection in conversational AI, providing a structured approach to managing unpredictable interactions. Bayesian Evidential Deep Learning (BEDL) improves uncertainty estimation, addressing overfitting issues [30]. Spectral-normalized Neural Gaussian Process (SNGP) enhances uncertainty estimation by enforcing distance-awareness, improving OOD input management [27]. Recent advancements in OOD detection, such as those by Jin et al., outperform traditional methods, emphasizing the need for robust evaluation metrics [86]. Methods like SLMM and RCAP enhance dialogue system adaptability and manage OOD inputs effectively [62, 26]. Integrating periodic activation functions, as explored by Meronen et al., enhances model interpretability and robustness [87]. These approaches address traditional models' limitations, significantly enhancing conversational AI's utility in real-world applications [2, 27, 88, 89, 90].

## 6.5 Innovative Frameworks for Enhanced Interaction

Innovative frameworks are essential for advancing conversational AI's interaction capabilities. Self-supervised learning techniques allow AI systems to autonomously learn from unlabeled data, refining models without extensive manual annotation [1]. Hybrid models combining rule-based and machine learning approaches optimize dialogue management, ensuring coherence and contextual relevance



AI-generated, for reference only.

in complex interactions [53]. Reinforcement learning techniques enhance interaction quality by allowing AI systems to learn optimal dialogue strategies through trial and error [1]. Knowledge graphs and ontologies provide structured representations, facilitating accurate intent recognition and response generation [6]. Modular and scalable architectures enhance interaction capabilities by enabling seamless integration of new functionalities [53]. These frameworks underscore the importance of integrating advanced learning techniques, hybrid models, and structured knowledge representations to enhance AI interaction capabilities, enabling dynamic identification and adaptation to emerging user intents [2, 35].

# 7 Natural Language Processing

## 7.1 Advancements in NLP Techniques

Recent advancements in Natural Language Processing (NLP) have notably enhanced AI systems, particularly in out-of-domain detection, intent recognition, and conversational AI. Self-supervised learning frameworks have been pivotal, leveraging vast unlabeled datasets to refine language models without extensive manual annotation, thereby enhancing linguistic feature extraction and text generation [15]. The advent of transformer-based architectures, such as BERT and its derivatives, has revolutionized the field by capturing deep contextual relationships within text, excelling in tasks like sentiment analysis and intent recognition across multilingual and cross-domain contexts [25, 72].

Transfer learning has further boosted NLP adaptability across domains and tasks, with fine-tuning pre-trained models on specific datasets enhancing task-specific performance and reducing the need for large labeled datasets [9]. Additionally, novel loss functions and optimization strategies, such as contrastive learning and domain adaptation, have refined model training, mitigating domain shift effects and bolstering model robustness in dynamic environments [56].

## 7.2 NLP in Abuse Detection and Sentiment Analysis

NLP plays a crucial role in detecting abusive language and conducting sentiment analysis, essential for maintaining safe online environments. As online discourse evolves, especially post-COVID-19, detection systems must adapt to new abusive language forms. Advanced techniques, including Hierarchical LSTM models and interpretability methods like TCAV, enhance classifiers' abilities to recognize explicit and implicit abuse, improving content moderation [41, 91, 36].

In abuse detection, transformer-based architectures like BERT significantly enhance the capture of contextual information and semantic nuances, improving accuracy [72]. Sentiment analysis leverages deep contextual embeddings to discern emotional tones in text, crucial for understanding public opinion and customer feedback [25]. Transfer learning and domain adaptation further enhance these models' versatility across domains and languages, enabling adaptation to new data distributions without extensive retraining [9]. These advancements underscore the necessity for sophisticated language models to accurately identify evolving abuse forms and improve detection systems' robustness [91, 36].

## 7.3 NLP in Intent Recognition and Task-Oriented Dialogue Systems

NLP is integral to enhancing intent recognition and optimizing task-oriented dialogue systems, crucial for efficient conversational agents. Techniques like IntentGPT improve dialogue systems' ability to decipher diverse user intents with minimal labeled data, enhancing user interactions in applications such as customer service [82, 92, 3, 49]. The SlotRefine model addresses slot uncoordination, improving natural language understanding [93]. Transformer-based architectures provide deep contextual embeddings, facilitating automatic identification of new intents and slots, allowing systems to manage diverse interactions accurately [94, 49].

Transfer learning techniques enable the discovery of new intent categories and improve representation learning, with multi-task pre-training strategies optimizing decision boundaries and detecting open intents [58, 84, 81, 13]. Fine-tuning pre-trained models facilitates domain adaptation, crucial for maintaining relevance and effectiveness. Domain adaptation strategies enhance generalization across domains and languages, improving performance in real-world applications [79, 3, 84]. These advancements highlight the critical role of advanced language models in improving human language interpretation, enabling systems to handle diverse and evolving user intents [92, 47, 3, 49, 82].

### 7.4 NLP for Uncertainty Quantification and Model Performance

NLP techniques are pivotal for uncertainty quantification and model performance evaluation, enhancing AI systems' robustness. Bayesian Neural Networks (BNNs) and ensemble methods provide probabilistic estimates, addressing aleatoric and epistemic uncertainties [30]. Integrating NLP with uncertainty frameworks enables handling of out-of-domain (OOD) inputs, with models like SNGP enhancing uncertainty estimation through distance-awareness, reducing overconfident predictions [27].

NLP techniques evaluate model performance using metrics such as accuracy and F1-score, providing comprehensive capability assessments. This integration enhances understanding of model behavior, optimizing performance across tasks and domains. Techniques like SNGP improve uncertainty estimation in single deep neural networks, particularly in OOD scenarios. Addressing ambiguous emotions through uncertainty quantification transforms tasks from simple classification to robust distribution estimation [88, 95, 27, 90].

Innovative methodologies like contrastive learning enhance model performance amid domain shifts, improving NLP models' robustness and adaptability [56]. Integrating NLP with uncertainty quantification is essential for enhancing AI systems' reliability, ensuring robustness across diverse applications [35, 90].

### 7.5 NLP in Multimodal and Cross-Domain Applications

NLP's application in multimodal and cross-domain contexts has expanded AI capabilities, enabling processing of diverse data types and domain adaptation. Multimodal NLP integrates text with images, audio, and video, enhancing understanding of complex data inputs, beneficial in applications like video captioning [17]. Cross-domain applications facilitate knowledge transfer, improving adaptability and generalization, with domain adaptation strategies aligning feature spaces and mitigating domain shift impacts [72].

Transformer-based architectures have been instrumental in advancing these applications, capturing contextual relationships and semantic nuances for processing complex multimodal data and adapting to new domains [25]. Contrastive learning techniques enhance cross-domain applications by improving discriminative power and aligning feature spaces, reducing reliance on large labeled datasets [56].

NLP's integration in multimodal and cross-domain applications underscores the necessity for sophisticated language models, facilitating novel category and intent discovery within complex datasets. Frameworks like TextGCD leverage visual-language models for category discovery, while methods like MIntOOD address multimodal intent understanding challenges [5, 17, 79, 41]. These advancements contribute to developing adaptable AI systems capable of high-quality performance across applications.

## 8 Conclusion

### 8.1 Emerging Trends and Future Directions

The expanding domains of out-of-domain detection, new intent discovery, generalized category discovery, conversational AI, and natural language processing offer numerous avenues for advancing artificial intelligence systems. In out-of-domain detection, future research should aim to develop training algorithms that enhance model stability and reduce overfitting, thereby increasing robustness across varied environments. Improvements to frameworks like the Spectral-normalized Neural Gaussian Process could further enhance uncertainty estimation and model reliability.

In intent recognition, there is a growing focus on detecting multiple intents within single utterances and incorporating additional knowledge to refine models. Future work may explore interactive intent design using large language models and hybrid strategies to improve detection capabilities. Additionally, adapting advanced models for better performance in few-shot out-of-scope intent classification is a promising area of research. Refining intent estimation and exploring alternative methods for intent assignments are also critical areas of interest.

Ontology expansion remains crucial for advancing conversational agents, necessitating integrated approaches to evaluate the impacts comprehensively. Future research should extend methods for multi-intent induction, improving dialogue systems' adaptability and accuracy. In new intent discovery, enhancing cluster interpretability and adapting methods for imbalanced datasets are key objectives. Research should also focus on automatically determining cluster counts and developing online strategies for efficient discovery.

In generalized category discovery, future studies could address domain shifts and enhance model robustness in real-world scenarios. Promising research avenues include refining adaptive sampling strategies and applying active generalized category discovery in complex situations. Integrating additional knowledge sources to improve class discovery precision and model reliability is another potential area of exploration. Moreover, enhancing robustness against out-of-domain noise and class imbalance, as well as developing strategies for estimating out-of-domain category counts, remains vital.

In specialized domains, improving model robustness for detecting rare and novel objects is essential, especially in challenging scenarios. Future research may explore methods that eliminate the need for retaining previous samples, thus enhancing learning efficiency in continual learning contexts. Additionally, alternative label extension methods and improvements in detection accuracy could be investigated, along with applying frameworks to other generative models. Research should also prioritize improving sampling techniques from inverted out-of-domain priors and exploring domain-specific applications to enhance the proposed paradigm's effectiveness.

In the pharmaceutical sector, future research should focus on developing interpretable molecular representations and enhancing existing explainable AI methods to better meet industry needs. In domain generalization, priorities should include continuous domain generalization and the integration of large-scale pre-training techniques, which may inform future conversational AI directions. Furthermore, advancing the semantic understanding of large language models in specific domains and enhancing their cross-domain in-context learning capabilities are critical for the progression of NLP applications.

These emerging trends and future research directions highlight the dynamic nature of the fields addressed in this survey, emphasizing the need for ongoing innovation and exploration to advance AI systems' capabilities in understanding and interacting with complex human language and environments.





AI-generated, for reference only.

## References


[1] Bing Liu, Eric Robertson, Scott Grigsby, and Sahisnu Mazumder. Self-initiated open world learning for autonomous ai agents. *arXiv preprint arXiv:2110.11385*, 2021.

[2] Jinggui Liang, Yuxia Wu, Yuan Fang, Hao Fei, and Lizi Liao. A survey of ontology expansion for conversational understanding. *arXiv preprint arXiv:2410.15019*, 2024.

[3] Masoud Akbari, Ali Mohades, and M Hassan Shirali-Shahreza. A hybrid architecture for out of domain intent detection and intent discovery. *arXiv preprint arXiv:2303.04134*, 2023.

[4] Jakub Konrád, Jan Pichl, Petr Marek, Petr Lorenc, Van Duy Ta, Ondřej Kobza, Lenka Hýlová, and Jan Šedivý. Alquist 4.0: Towards social intelligence using generative models and dialogue personalization, 2021.

[5] Hanlei Zhang, Qianrui Zhou, Hua Xu, Jianhua Su, Roberto Evans, and Kai Gao. Multimodal classification and out-of-distribution detection for multimodal intent understanding. *arXiv preprint arXiv:2412.12453*, 2024.

[6] Bo Wang and Tsunenori Mine. One stone, four birds: A comprehensive solution for qa system using supervised contrastive learning, 2024.

[7] Libo Qin, Fuxuan Wei, Tianbao Xie, Xiao Xu, Wanxiang Che, and Ting Liu. Gl-gin: Fast and accurate non-autoregressive model for joint multiple intent detection and slot filling. *arXiv preprint arXiv:2106.01925*, 2021.

[8] Henry Weld, Xiaoqi Huang, Siqu Long, Josiah Poon, and Soyeon Caren Han. A survey of joint intent detection and slot filling models in natural language understanding. *ACM Computing Surveys*, 55(8):1–38, 2022.

[9] Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and S Yu Philip. Generalizing to unseen domains: A survey on domain generalization. *IEEE transactions on knowledge and data engineering*, 35(8):8052–8072, 2022.

[10] Stefan Larson, Yi Yang Gordon Lim, Yutong Ai, David Kuang, and Kevin Leach. Evaluating out-of-distribution performance on document image classifiers. *Advances in Neural Information Processing Systems*, 35:11673–11685, 2022.

[11] Ming Tan, Yang Yu, Haoyu Wang, Dakuo Wang, Saloni Potdar, Shiyu Chang, and Mo Yu. Out-of-domain detection for low-resource text classification tasks, 2019.

[12] Yunhua Zhou, Jiawei Hong, and Xipeng Qiu. Towards open environment intent prediction. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 2226–2240, 2023.

[13] Hanlei Zhang, Hua Xu, Shaojie Zhao, and Qianrui Zhou. Learning discriminative representations and decision boundaries for open intent detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31:1611–1623, 2023.

[14] Paulo Cavalin and Claudio Santos Pinhanez. Theoretical and empirical advantages of dense-vector to one-hot encoding of intent classes in open-world scenarios. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 16000–16013, 2024.

[15] Hanlei Zhang, Hua Xu, Xin Wang, Fei Long, and Kai Gao. A clustering framework for unsupervised and semi-supervised new intent discovery, 2023.

[16] Shun Zhang, Chaoran Yan, Jian Yang, Jiaheng Liu, Ying Mo, Jiaqi Bai, Tongliang Li, and Zhoujun Li. Towards real-world scenario: Imbalanced new intent discovery. *arXiv preprint arXiv:2406.03127*, 2024.

[17] Haiyang Zheng, Nan Pu, Wenjing Li, Nicu Sebe, and Zhun Zhong. Textual knowledge matters: Cross-modality co-teaching for generalized visual class discovery, 2024.

[18] Pavel Burnyshev, Valentin Malykh, Andrey Bout, Ekaterina Artemova, and Irina Piontkovskaya. A single example can improve zero-shot data generation, 2021.

[19] Yixin Fei, Zhongkai Zhao, Siwei Yang, and Bingchen Zhao. Xcon: Learning with experts for fine-grained category discovery, 2022.

[20] Yuchang Su, Renping Zhou, Siyu Huang, Xingjian Li, Tianyang Wang, Ziyue Wang, and Min Xu. Multimodal generalized category discovery, 2024.

[21] Nan Pu, Zhun Zhong, and Nicu Sebe. Dynamic conceptional contrastive learning for generalized category discovery, 2023.

[22] Zhengyuan Peng, Qijian Tian, Jianqing Xu, Yizhang Jin, Xuequan Lu, Xin Tan, Yuan Xie, and Lizhuang Ma. Generalized category discovery in semantic segmentation, 2023.

[23] Alexander Podolskiy, Dmitry Lipin, Andrey Bout, Ekaterina Artemova, and Irina Piontkovskaya. Revisiting mahalanobis distance for transformer-based out-of-domain detection, 2022.

[24] Varun Gangal, Abhinav Arora, Arash Einolghozati, and Sonal Gupta. Likelihood ratios and generative classifiers for unsupervised out-of-domain detection in task oriented dialog, 2019.

[25] Libo Qin, Wanxiang Che, Yangming Li, Haoyang Wen, and Ting Liu. A stack-propagation framework with token-level intent detection for spoken language understanding. *arXiv preprint arXiv:1909.02188*, 2019.

[26] Zengfeng Zeng, Dan Ma, Haiqin Yang, Zhen Gou, and Jianping Shen. Automatic intent-slot induction for dialogue systems. In *Proceedings of the Web Conference 2021*, pages 2578–2589, 2021.

[27] Jeremiah Zhe Liu, Shreyas Padhy, Jie Ren, Zi Lin, Yeming Wen, Ghassen Jerfel, Zack Nado, Jasper Snoek, Dustin Tran, and Balaji Lakshminarayanan. A simple approach to improve single-model deep uncertainty via distance-awareness, 2022.

[28] Joel Loo, Zhanxin Wu, and David Hsu. Open scene graphs for open world object-goal navigation. *arXiv preprint arXiv:2407.02473*, 2024.

[29] José Jiménez-Luna, Francesca Grisoni, and Gisbert Schneider. Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2(10):573–584, 2020.

[30] Manuel Haussmann, Sebastian Gerwinn, and Melih Kandemir. Bayesian evidential deep learning with pac regularization, 2021.

[31] Yufei Cui, Yu Mao, Ziquan Liu, Qiao Li, Antoni B. Chan, Xue Liu, Tei-Wei Kuo, and Chun Jason Xue. Variational nested dropout, 2022.

[32] Mohamed Akrout, Amal Feriani, Faouzi Bellili, Amine Mezghani, and Ekram Hossain. Domain generalization in machine learning models for wireless communications: Concepts, state-of-the-art, and open issues. *IEEE Communications Surveys & Tutorials*, 2023.

[33] Iftitahu Ni'mah, Meng Fang, Vlado Menkovski, and Mykola Pechenizkiy. Protoinfomax: Prototypical networks with mutual information maximization for out-of-domain detection, 2021.

[34] Libo Qin, Xiao Xu, Wanxiang Che, and Ting Liu. Agif: An adaptive graph-interactive framework for joint multiple intent detection and slot filling. *arXiv preprint arXiv:2004.10087*, 2020.

[35] A framework to generate high-qua.

[36] Isar Nejadgholi, Kathleen C Fraser, and Svetlana Kiritchenko. Improving generalizability in implicitly abusive language detection with concept activation vectors. *arXiv preprint arXiv:2204.02261*, 2022.

[37] Zhi Hou, Baosheng Yu, Yu Qiao, Xiaojiang Peng, and Dacheng Tao. Affordance transfer learning for human-object interaction detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 495–504, 2021.

[38] Petr Lorenc, Tommaso Gargiani, Jan Pichl, Jakub Konrád, Petr Marek, Ondřej Kobza, and Jan Šedivý. Metric learning and adaptive boundary for out-of-domain detection, 2022.

[39] Keyang Xu, Tongzheng Ren, Shikun Zhang, Yihao Feng, and Caiming Xiong. Unsupervised out-of-domain detection via pre-trained transformers, 2022.

[40] Yinhe Zheng, Guanyi Chen, and Minlie Huang. Out-of-domain detection for natural language understanding in dialog systems, 2022.

[41] Rheeya Uppaal, Junjie Hu, and Yixuan Li. Is fine-tuning needed? pre-trained language models are near perfect for out-of-domain detection, 2023.

[42] Sagar Vaze, Andrea Vedaldi, and Andrew Zisserman. No representation rules them all in category discovery, 2023.

[43] Philipp Seidl, Andreu Vall, Sepp Hochreiter, and Günter Klambauer. Enhancing activity prediction models in drug discovery with the ability to understand human language. In *International Conference on Machine Learning*, pages 30458–30490. PMLR, 2023.

[44] Keyang Xu, Tongzheng Ren, Shikun Zhang, Yihao Feng, and Caiming Xiong. Unsupervised out-of-domain detection via pre-trained transformers. *arXiv preprint arXiv:2106.00948*, 2021.

[45] Yanan Wu, Zhiyuan Zeng, Keqing He, Yutao Mou, Pei Wang, and Weiran Xu. Distribution calibration for out-of-domain detection with bayesian approximation, 2022.

[46] Grégor Jouet, Clément Duhart, Francis Rousseaux, Julio Laborde, and Cyril de Runz. Combining pre-trained vision transformers and cider for out of domain detection, 2023.

[47] Yunhua Zhou, Guofeng Quan, and Xipeng Qiu. A probabilistic framework for discovering new intents. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3771–3784, 2023.

[48] Yutao Mou, Keqing He, Yanan Wu, Pei Wang, Jingang Wang, Wei Wu, Yi Huang, Junlan Feng, and Weiran Xu. Generalized intent discovery: Learning from open world dialogue system. *arXiv preprint arXiv:2209.06030*, 2022.

[49] Juan A Rodriguez, Nicholas Botzer, David Vazquez, Christopher Pal, Marco Pedersoli, and Issam Laradji. Intentgpt: Few-shot intent discovery with large language models. *arXiv preprint arXiv:2411.10670*, 2024.

[50] Dhruv Shah, Benjamin Eysenbach, Gregory Kahn, Nicholas Rhinehart, and Sergey Levine. Rapid exploration for open-world navigation with latent goal models. *arXiv preprint arXiv:2104.05859*, 2021.

[51] Xiaoshuai Song, Keqing He, Pei Wang, Guanting Dong, Yutao Mou, Jingang Wang, Yunsen Xian, Xunliang Cai, and Weiran Xu. Large language models meet open-world intent discovery and recognition: An evaluation of chatgpt. *arXiv preprint arXiv:2310.10176*, 2023.

[52] Aleksei Ustimenko, Artem Beliakov, and Liudmila Prokhorenkova. Gradient boosting performs gaussian process inference, 2023.

[53] Kartik Talamadupula, Gordon Briggs, Matthias Scheutz, and Subbarao Kambhampti. Architectural mechanisms for handling human instructions for open-world mixed-initiative team tasks and goals. *Advances in Cognitive Systems*, 5:37–56, 2017.

[54] Yu Song and Donglin Wang. Learning on graphs with out-of-distribution nodes. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1635–1645, 2022.

[55] Shun Zhang, Chaoran Yan, Jian Yang, Jiaheng Liu, Ying Mo, Jiaqi Bai, Tongliang Li, and Zhoujun Li. Towards real-world scenario: Imbalanced new intent discovery, 2024.

[56] Boyan Gao, Henry Gouk, Yongxin Yang, and Timothy Hospedales. Loss function learning for domain generalization by implicit gradient. In *International Conference on Machine Learning*, pages 7002–7016. PMLR, 2022.

[57] Yiwei Jiang, Maarten De Raedt, Johannes Deleu, Thomas Demeester, and Chris Develder. Few-shot out-of-scope intent classification: analyzing the robustness of prompt-based learning. *Applied Intelligence*, 54(2):1474–1496, 2024.

[58] Shun Zhang, Jian Yang, Jiaqi Bai, Chaoran Yan, Tongliang Li, Zhao Yan, and Zhoujun Li. New intent discovery with attracting and dispersing prototype, 2024.

[59] Semi-supervised class discovery.

[60] Patrick Feeney, Sarah Schneider, Panagiotis Lymperopoulos, Li-Ping Liu, Matthias Scheutz, and Michael C. Hughes. Novelcraft: A dataset for novelty detection and discovery in open worlds, 2023.

[61] Yanan Wu, Zhiyuan Zeng, Keqing He, Hong Xu, Yuanmeng Yan, Huixing Jiang, and Weiran Xu. Novel slot detection: A benchmark for discovering unknown slot types in the task-oriented dialogue system. *arXiv preprint arXiv:2105.14313*, 2021.

[62] Zifeng Cheng, Zhiwei Jiang, Yafeng Yin, Cong Wang, and Qing Gu. Learning to classify open intent via soft labeling and manifold mixup. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:635–645, 2022.

[63] Shun Zhang, Jian Yang, Jiaqi Bai, Chaoran Yan, Tongliang Li, Zhao Yan, and Zhoujun Li. New intent discovery with attracting and dispersing prototype. *arXiv preprint arXiv:2403.16913*, 2024.

[64] Shijie Ma, Fei Zhu, Zhun Zhong, Xu-Yao Zhang, and Cheng-Lin Liu. Active generalized category discovery, 2024.

[65] Wenbin An, Feng Tian, Qinghua Zheng, Wei Ding, QianYing Wang, and Ping Chen. Generalized category discovery with decoupled prototypical network, 2023.

[66] Shijie Ma, Fei Zhu, Zhun Zhong, Wenzhuo Liu, Xu-Yao Zhang, and Cheng-Lin Liu. Happy: A debiased learning framework for continual generalized category discovery, 2024.

[67] Bingchen Zhao and Oisin Mac Aodha. Incremental generalized category discovery, 2023.

[68] Xin Wen, Bingchen Zhao, and Xiaojuan Qi. Parametric classification for generalized category discovery: A baseline study, 2023.

[69] Fernando Julio Cendra, Bingchen Zhao, and Kai Han. Promptccd: Learning gaussian mixture prompt pool for continual category discovery, 2024.

[70] Tingzhang Luo, Mingxuan Du, Jiatao Shi, Xinxiang Chen, Bingchen Zhao, and Shaoguang Huang. Contextuality helps representation learning for generalized category discovery, 2024.

[71] Yuxun Qu, Yongqiang Tang, Chenyang Zhang, and Wensheng Zhang. Adaptgcd: Multi-expert adapter tuning for generalized category discovery, 2024.

[72] Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4396–4415, 2022.

[73] Enguang Wang, Zhimao Peng, Zhengyuan Xie, Fei Yang, Xialei Liu, and Ming-Ming Cheng. Unlocking the multi-modal potential of clip for generalized category discovery, 2024.

[74] Florent Chiaroni, Jose Dolz, Ziko Imtiaz Masud, Amar Mitiche, and Ismail Ben Ayed. Parametric information maximization for generalized category discovery, 2023.

[75] Sarah Rastegar, Mohammadreza Salehi, Yuki M. Asano, Hazel Doughty, and Cees G. M. Snoek. Selex: Self-expertise in fine-grained generalized category discovery, 2024.

[76] Hongjun Wang, Sagar Vaze, and Kai Han. Hilo: A learning framework for generalized category discovery robust to domain shifts, 2024.

[77] Keon-Hee Park, Hakyung Lee, Kyungwoo Song, and Gyeong-Moon Park. Online continuous generalized category discovery, 2024.

[78] Yanan Wu, Zhixiang Chi, Yang Wang, and Songhe Feng. Metagcd: Learning to continually learn in generalized category discovery, 2023.

[79] Nikhita Vedula, Rahul Gupta, Aman Alok, Mukund Sridhar, and Shankar Ananthakrishnan. Advin: Automatically discovering novel domains and intents from user text utterances. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7627–7631. IEEE, 2022.

[80] Unknown intent detection using m.

[81] Yuwei Zhang, Haode Zhang, Li-Ming Zhan, Xiao-Ming Wu, and Albert Y. S. Lam. New intent discovery with pre-training and contrastive learning, 2022.

[82] Gaurav Arora, Shreya Jain, and Srujana Merugu. Intent detection in the age of llms. *arXiv preprint arXiv:2410.01627*, 2024.

[83] Pei Wang, Keqing He, Yejie Wang, Xiaoshuai Song, Yutao Mou, Jingang Wang, Yunsen Xian, Xunliang Cai, and Weiran Xu. Beyond the known: Investigating llms performance on out-of-domain intent detection. *arXiv preprint arXiv:2402.17256*, 2024.

[84] Yilin Shen, Yen-Chang Hsu, Avik Ray, and Hongxia Jin. Enhancing the generalization for intent classification and out-of-domain detection in slu, 2021.

[85] Jinggui Liang, Lizi Liao, Hao Fei, and Jing Jiang. Synergizing large language models and pre-trained smaller models for conversational intent discovery. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 14133–14147, 2024.

[86] Di Jin, Shuyang Gao, Seokhwan Kim, Yang Liu, and Dilek Hakkani-Tur. Towards textual out-of-domain detection without in-domain labels, 2022.

[87] Lassi Meronen, Martin Trapp, and Arno Solin. Periodic activation functions induce stationarity, 2021.

[88] Jeremiah Zhe Liu, Zi Lin, Shreyas Padhy, Dustin Tran, Tania Bedrax-Weiss, and Balaji Lakshminarayanan. Simple and principled uncertainty estimation with deterministic deep learning via distance awareness, 2020.

[89] Yawen Ouyang, Jiasheng Ye, Yu Chen, Xinyu Dai, Shujian Huang, and Jiajun Chen. Energy-based unknown intent detection with data manipulation. *arXiv preprint arXiv:2107.12542*, 2021.

[90] Hui Wang, Shiwan Zhao, Jiaming Zhou, Xiguang Zheng, Haoqin Sun, Xuechen Wang, and Yong Qin. Uncertainty-aware mean opinion score prediction, 2024.

[91] Shashi Shekhar, Hitendra Garg, Rohit Agrawal, Shivendra Shivani, and Bhisham Sharma. Hatred and trolling detection transliteration framework using hierarchical lstm in code-mixed social media text. *Complex & Intelligent Systems*, 9(3):2813–2826, 2023.

[92] Intentgpt: Few-shot intent disco.

[93] Di Wu, Liang Ding, Fan Lu, and Jian Xie. Slotrefine: A fast non-autoregressive model for joint intent detection and slot filling. *arXiv preprint arXiv:2010.02693*, 2020.

[94] Yuxia Wu, Tianhao Dai, Zhedong Zheng, and Lizi Liao. Active discovering new slots for task-oriented conversation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024.

[95] Wen Wu, Bo Li, Chao Zhang, Chung-Cheng Chiu, Qiujia Li, Junwen Bai, Tara N. Sainath, and Philip C. Woodland. Handling ambiguity in emotion: From out-of-domain detection to distribution estimation, 2024.

**Disclaimer:**

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.