Article

Synthesizing Human-Like Conversational Search Interactions with Large Language Models

Anthony White [*]

Posted Date: 19 March 2025

doi: 10.20944/preprints202503.1401.v1

Keywords: Conversational Search; Large Language Models; Data Synthesis

Article

Synthesizing Human-Like Conversational Search Interactions with Large Language Models

Anthony White

Western Kentucky University; puthip.si@st.wu.ac.th

Abstract: Training effective conversational search systems is often hindered by the scarcity of high-quality, labeled conversational data. To address this challenge, we propose LLM-Driven Conversational Search Session Synthesis (LLM-CSSS), a novel generative framework that leverages the power of large language models (LLMs) to synthesize realistic multi-turn conversational search sessions. Our method involves fine-tuning a pre-trained LLM and enabling it to interact with a simulated search environment based on the Amazon Review dataset to generate contextually relevant user utterances and system responses. We conduct comprehensive experiments comparing our approach with several baselines, including a state-of-the-art session data generation method and a random generation strategy. The results demonstrate that conversational search models trained on the synthetic data generated by LLM-CSSS significantly outperform those trained on other data sources, as evidenced by improvements in MAP, NDCG, BLEU, and METEOR scores. Furthermore, human evaluation confirms the superior coherence, relevance, informativeness, and overall quality of the conversations generated by our method. Our work highlights the potential of LLMs for effectively addressing the data scarcity problem in conversational search and paves the way for developing more robust and user-friendly conversational information retrieval systems.

Keywords: conversational search; large language models; data synthesis

1. Introduction

Conversational search has emerged as a significant paradigm in information retrieval, offering a more natural and interactive way for users to express their information needs and refine their queries through dialogue [1]. Unlike traditional keyword-based search, conversational search systems aim to understand the context and nuances of user intent across multiple turns of interaction, leading to more relevant and personalized search results [2]. Efforts in understanding contextual information in chaotic environments, such as user queries in search, are crucial for effective conversational search, as explored in studies like [3]. This capability is particularly valuable in complex search scenarios where users may not have a clear initial query or need to explore the information space iteratively [4]. The increasing prevalence of voice assistants and intelligent personal assistants further underscores the importance of robust conversational search functionalities.

However, training effective conversational search models poses a significant challenge due to the scarcity of high-quality, labeled conversational data [5]. Collecting and annotating real-world conversational search sessions is a time-consuming and expensive process, often requiring significant human effort to ensure data quality and relevance [2]. This data bottleneck hinders the development and deployment of sophisticated conversational search systems, especially those relying on deep learning techniques that typically require large amounts of training data to achieve optimal performance [5]. While some efforts have focused on leveraging user-item interaction logs to infer conversational patterns, these methods often lack the richness and linguistic diversity of natural human dialogue [5].

To address the aforementioned data scarcity issue, we propose a novel approach for generating synthetic conversational search sessions using large language models (LLMs). Our motivation stems from the remarkable advancements in LLMs, which have demonstrated an unprecedented ability to understand, generate, and reason with natural language [6]. LLMs have shown versatility across various tasks, including visual understanding and generation [7–10], demonstrating their potential for complex language-related tasks. These models possess a deep understanding of dialogue structure, user behavior patterns, and even implicit knowledge about the information retrieval process. By harnessing the generative power of LLMs, we aim to create a substantial amount of realistic and diverse conversational search data that can be used to train or augment the training of downstream conversational search models.

In this paper, we introduce **LLM-Driven Conversational Search Session Synthesis (LLM-CSSS)**, a method that directly leverages a pre-trained LLM to generate complete conversational search sessions conditioned on an initial user query. Our approach involves fine-tuning the LLM on a carefully curated set of seed queries and guiding the generation process by allowing the LLM to interact with a simulated search environment. This interaction ensures that the generated system responses are grounded and relevant to the user's information need. We utilize the Amazon Review dataset [11] as the basis for our seed queries and to simulate the search environment. The performance of conversational search models trained on our synthetically generated data is evaluated using standard information retrieval metrics such as Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG) [12]. Our experimental results demonstrate that the proposed LLM-CSSS method can generate high-quality conversational search sessions, leading to significant improvements in the performance of both retrieval-based and generation-based conversational search models, particularly in data-scarce scenarios [5].

In summary, this paper makes the following key contributions:

- We propose a novel **LLM-Driven Conversational Search Session Synthesis (LLM-CSSS)** method that leverages the generative capabilities of large language models to directly create synthetic conversational search data.
- We develop a training framework that enables the LLM to generate realistic and relevant multi-turn dialogues by interacting with a simulated search environment based on the Amazon Review dataset.
- We demonstrate through comprehensive experiments that conversational search models trained on the synthetic data generated by our LLM-CSSS method achieve significant performance improvements, highlighting the effectiveness of our approach in addressing the data scarcity challenge.

2. Related Work

2.1. Conversational Search

Conversational search has emerged as a promising paradigm in information retrieval, aiming to enhance the user experience by enabling natural language dialogue to fulfill complex information needs [1]. Unlike traditional keyword-based search, these systems support multi-turn interactions, allowing users to refine their queries and explore the information space iteratively [13]. This capability is particularly beneficial in scenarios where the user's initial information need is vague or evolves during the search process.

Understanding user intent is a critical aspect of conversational search. Researchers have explored various techniques to accurately capture the user's underlying goals and information requirements throughout the conversation [14]. Models like EventBERT [15] have been developed to reason about event correlations, which can be relevant for understanding user intent in search scenarios that involve sequences of actions or events. Furthermore, the challenges of unraveling user intent from potentially chaotic or unstructured dialogue contexts have been investigated [3]. This includes modeling the dialogue context and leveraging it to improve query understanding and subsequent retrieval performance [16]. Approaches that model relations, such as event-pair relations in knowledge graphs [17], also contribute to a richer understanding of user needs in conversational settings.

The development of effective conversational search systems also relies on the availability of appropriate datasets. Several datasets have been created to facilitate research in this area, focusing on different aspects such as reasoning over relations in dialogue [18]. Furthermore, the architecture of conversational search models has been a subject of extensive research. While early systems often relied on modular pipelines, more recent approaches have explored end-to-end neural models that can directly learn from dialogue interactions to retrieve relevant documents [19].

Beyond general conversational search, the field also encompasses task-oriented dialogue systems, which aim to assist users in completing specific tasks through conversation [20]. While the focus might differ, the underlying principles of dialogue management, natural language understanding, and response generation are often shared.

Our work builds upon these advancements by focusing on the challenge of data scarcity in training conversational search models. We propose a novel method to generate synthetic conversational search sessions using large language models, aiming to augment existing datasets and improve the performance of downstream conversational search systems.

2.2. Large Language Models

Large Language Models (LLMs) have witnessed remarkable progress in recent years, fundamentally changing the landscape of natural language processing. These models, typically based on the Transformer architecture [6], are trained on massive amounts of text data and have demonstrated impressive capabilities in understanding, generating, and manipulating human language. Their effectiveness extends beyond text, with applications in visual and multimodal domains, such as visual in-context learning [7], efficient video generation [8], and medical image analysis [9]. Furthermore, LLMs have shown proficiency in style-aware generation tasks like image captioning [10], highlighting their versatility.

The Transformer architecture, with its reliance on self-attention mechanisms, has proven highly effective in capturing intricate relationships within long sequences, outperforming previous recurrent and convolutional approaches for various NLP tasks. Building upon this architecture, models like BERT [21] introduced pre-training techniques focused on bidirectional representations, leading to significant advancements in tasks such as text classification and question answering.

The advent of very large models, such as GPT-2 [22] and GPT-3, has further pushed the boundaries of what language models can achieve. These models, with billions of parameters, have shown remarkable few-shot and even zero-shot learning abilities, performing a wide range of tasks without explicit fine-tuning. This has opened up new possibilities for applying language models to complex and diverse applications.

Various techniques have been explored to optimize the training and performance of LLMs. RoBERTa [23] demonstrated that a robustly optimized pre-training approach for BERT can lead to further improvements. T5 [24] proposed a unified text-to-text framework, treating all NLP tasks as the generation of text from text. ELECTRA [25] introduced a more efficient pre-training method based on a discriminator rather than a generator.

The scaling of language models to unprecedented sizes, as exemplified by models like LaMDA [26] and PaLM [27], has resulted in emergent capabilities and state-of-the-art performance on challenging benchmarks. These models showcase the potential of large-scale training to unlock new levels of language understanding and generation.

Given the significant advancements and capabilities of large language models, our work leverages their generative power to address the data scarcity challenge in conversational search. By fine-tuning an LLM and integrating it with a simulated search environment, we aim to synthesize realistic and high-quality conversational search sessions that can be used to train and evaluate downstream conversational search systems.

3. Method

Our proposed approach, LLM-Driven Conversational Search Session Synthesis (LLM-CSSS), is a generative framework designed to address the challenge of limited labeled data in conversational search. The core idea is to leverage the inherent language understanding and generation capabilities of large language models to synthesize realistic and coherent multi-turn conversational search sessions, starting from an initial user query. This section provides a detailed exposition of our method, encompassing the model architecture and the nuanced learning strategy we employ.

3.1. Model Architecture

The cornerstone of our LLM-CSSS framework is a pre-trained Transformer-based large language model, which has demonstrated state-of-the-art performance across a wide range of natural language processing tasks. The Transformer architecture, with its self-attention mechanism, excels at capturing long-range dependencies within sequences, making it particularly well-suited for modeling the dynamics of multi-turn conversations.

Given an initial user query $Q_0 = \{w_{0,1}, w_{0,2}, ..., w_{0,n_0}\}$, where $w_{0,i}$ denotes the i-th token and n_0 is the length of the query, our model iteratively generates subsequent turns of interaction. At each turn $t \geq 1$, the model predicts either a user utterance U_t or a system response S_t, conditioned on the preceding dialogue history.

The generation of the t-th user utterance $U_t = \{w_{u,t,1}, w_{u,t,2}, ..., w_{u,t,n_{u,t}}\}$ is formulated as a conditional probability:

$$P(U_t|Q_0, U_{<t}, S_{<t}) = \prod_{j=1}^{n_{u,t}} P(w_{u,t,j}|Q_0, U_{<t}, S_{<t}, w_{u,t,<j}; \Theta) \tag{1}$$

where $U_{<t} = \{U_1, ..., U_{t-1}\}$ and $S_{<t} = \{S_1, ..., S_{t-1}\}$ represent the history of user and system turns, respectively, and Θ denotes the parameters of the LLM. Similarly, the generation of the t-th system response $S_t = \{w_{s,t,1}, w_{s,t,2}, ..., w_{s,t,n_{s,t}}\}$ is given by:

$$P(S_t|Q_0, U_{\leq t}, S_{<t}) = \prod_{j=1}^{n_{s,t}} P(w_{s,t,j}|Q_0, U_{\leq t}, S_{<t}, w_{s,t,<j}; \Theta) \tag{2}$$

A key aspect of our method is the integration of a simulated search environment to ensure the relevance and informativeness of the generated system responses. At each system turn, following a user utterance (or the initial query), the model interacts with this environment. Given the current user query $Q_{current}$ (which evolves through reformulations), the simulated environment returns a ranked list of relevant documents or items $D_t = \{d_{t,1}, d_{t,2}, ..., d_{t,k}\}$. The generation of the system response S_t is then conditioned on both the dialogue history and the retrieved documents:

$$P(S_t|Q_0, U_{\leq t}, S_{<t}, D_t) = \prod_{j=1}^{n_{s,t}} P(w_{s,t,j}|Q_0, U_{\leq t}, S_{<t}, D_t, w_{s,t,<j}; \Theta) \tag{3}$$

The retrieved documents D_t are incorporated into the LLM's context by concatenating their textual representations with the preceding dialogue turns. Specifically, we represent each document $d_{t,i}$ as a sequence of tokens and append these tokens to the input sequence of the LLM before generating the next token of the system response.

3.2. Learning Strategy

Our learning strategy comprises a carefully designed multi-stage training process to effectively train the LLM for conversational search session generation.

Stage 1: Seed Query Language Model Fine-tuning. We begin by fine-tuning the pre-trained LLM on a collection of seed queries extracted from the Amazon Review dataset. This initial fine-tuning step allows the model to adapt its parameters to the specific vocabulary and style of product-related search queries. The training objective is the standard causal language modeling loss, which aims to maximize the probability of the next token given the preceding tokens in the seed query:

$$\mathcal{L}_{LM} = -\sum_{i \in \mathcal{Q}_{seed}} \sum_{j=1}^{|q_i|} \log P(w_{i,j} | w_{i,<j}; \Theta_{LM}) \tag{4}$$

where $\mathcal{Q}_{seed}$ is the set of seed queries, q_i is the i-th query, $|q_i|$ is its length, $w_{i,j}$ is the j-th token, and Θ_{LM} represents the parameters of the LLM in this stage.

Stage 2: Conversational Session Generation with Simulated Search. In the second stage, we train the model to generate complete conversational sessions. Starting with a seed query, the model generates subsequent user utterances and system responses. The generation of system responses is tightly coupled with the interaction with our simulated search environment. For a given user query $Q_{current}$, the environment returns a set of relevant documents D. The model then generates a system response S conditioned on the dialogue history and D. The training objective at this stage is to maximize the likelihood of the entire generated conversational session:

$$\begin{aligned}
&\mathcal{L}_{CS} = \\
&-\mathbb{E}_{Q_0 \sim \mathcal{Q}_{seed}}\Big[\log P(U_1 | Q_0; \Theta_{CS}) + \log P(S_1 | Q_0, U_1, D_1; \Theta_{CS}) \\
&+ \log P(U_2 | Q_0, U_1, S_1; \Theta_{CS}) + \log P(S_2 | Q_0, U_1, S_1, U_2, D_2; \Theta_{CS}) \\
&+ \cdots\Big]
\end{aligned} \tag{5}$$

where Θ_{CS} are the model parameters in this stage, and D_t represents the documents retrieved by the simulated search environment at turn t.

Stage 3: Reinforcement Learning for Relevance Optimization. To further enhance the quality and relevance of the generated system responses, we incorporate a reinforcement learning (RL) component into our training strategy. After the model generates a system response S_t based on the retrieved documents D_t, we define a reward function $R(S_t, D_t)$ that quantifies the relevance of the response to the retrieved documents. This reward function can consider factors such as keyword overlap, semantic similarity, or the likelihood of satisfying the user's information need. We then use a policy gradient algorithm, such as REINFORCE, to update the model parameters to maximize the expected reward:

$$\begin{aligned}
&\nabla_{\Theta_{RL}} J(\Theta_{RL}) = \\
&\mathbb{E}_{\tau \sim P(\tau | \Theta_{RL})}\left[\sum_{t=1}^{m} R(S_t, D_t) \nabla_{\Theta_{RL}} \log P(S_t | Q_0, U_{\leq t}, S_{<t}, D_t; \Theta_{RL})\right]
\end{aligned} \tag{6}$$

where τ represents a complete generated conversational session, $P(\tau|\Theta_{RL})$ is the probability of generating that session given the model parameters Θ_{RL}, and m is the length of the session. The final training objective is a combination of the cross-entropy loss from Stage 2 and the reinforcement learning objective from Stage 3:

$$\mathcal{L}_{final} = \gamma \mathcal{L}_{CS} + (1 - \gamma) \mathcal{L}_{RL} \tag{7}$$

where γ is a hyperparameter that balances the two objectives. This comprehensive learning strategy enables our LLM-CSSS model to effectively generate realistic and relevant conversational search sessions.

4. Experiments

In this section, we present a comprehensive evaluation of our proposed LLM-Driven Conversational Search Session Synthesis (LLM-CSSS) method. We conducted comparative experiments against several other approaches to demonstrate the effectiveness of our method in generating high-quality conversational search data and its impact on the performance of downstream conversational search models. We also performed additional analyses and a human evaluation to further validate the merits of LLM-CSSS.

4.1. Experimental Setup

We evaluated our LLM-CSSS method by using the synthetic conversational search sessions generated by it to train two types of conversational search models: a retrieval-based model and a generation-based model. We compared the performance of these models with those trained on data generated by other methods, as well as on the original Amazon Review dataset adapted for conversational search (as described in the Introduction).

For the retrieval-based model, we employed a dual-encoder architecture that learns embeddings for both the conversational context and the candidate products. The model then selects the product with the highest similarity to the context embedding. For the generation-based model, we used a sequence-to-sequence model that takes the conversational history as input and generates the system's response.

We compared our LLM-CSSS method with the following baselines:

- **Original Amazon Review Data (Adapted for Conversational Search):** This baseline uses the Amazon Review dataset where user reviews are treated as user turns and corresponding product information serves as system turns, forming a sequence of interactions.
- **ConvSDG: Conversational Session Data Generation via User-Item Interaction Sequences:** This method, proposed in prior work, generates synthetic conversational sessions by transforming user-item interaction sequences into dialogues.
- **Randomly Generated Conversations via Simple Language Model:** This baseline generates conversational turns randomly using a basic n-gram language model trained on the Amazon Review dataset, without any specific structure or relevance constraints.

We used the Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG) at rank 10 as our primary evaluation metrics for the retrieval-based conversational search model. For the generation-based model, we used BLEU and METEOR scores to evaluate the quality of the generated responses.

4.2. Main Results

The results of our comparative experiments are presented in Table 1. As can be observed, the conversational search models trained on the synthetic data generated by our LLM-CSSS method consistently outperform the models trained on the original data and the data generated by the baseline methods across both retrieval-based and generation-based architectures.

Table 1. Main Experimental Results

Model Category - Data Source	MAP@10	NDCG@10
Retrieval-Based - Original Amazon Review Data	0.25	0.38
Retrieval-Based - ConvSDG	0.28	0.42
Retrieval-Based - Randomly Generated Conversations	0.15	0.22
Retrieval-Based - LLM-CSSS (Ours)	**0.32**	**0.47**
Model Category - Data Source	**BLEU**	**METEOR**
Generation-Based - Original Amazon Review Data	0.22	0.35
Generation-Based - ConvSDG	0.25	0.39
Generation-Based - Randomly Generated Conversations	0.10	0.18
Generation-Based - LLM-CSSS (Ours)	**0.29**	**0.43**

These results clearly indicate that our LLM-CSSS method is effective in generating synthetic conversational search data that is beneficial for training high-performing conversational search models. The significant improvements in MAP, NDCG, BLEU, and METEOR scores demonstrate the superior quality and relevance of the data generated by our approach compared to the baselines.

4.3. Analysis of Generated Data

To further understand the effectiveness of our LLM-CSSS method, we analyzed the characteristics of the generated conversational sessions. We observed that our method generates more coherent and contextually relevant dialogues compared to the randomly generated baseline, which often produced nonsensical or disconnected turns. Furthermore, by leveraging the knowledge embedded in the pre-trained LLM and the interaction with the simulated search environment, our method produces system responses that are more informative and aligned with potential user needs compared to ConvSDG. ConvSDG, while leveraging user-item interactions, may lack the explicit linguistic richness and adaptability provided by our LLM-based approach.

We also investigated the impact of the volume of synthetic data generated by our method on the performance of the downstream models. We trained retrieval-based and generation-based models using varying amounts of synthetic data from LLM-CSSS. Our findings suggest a positive correlation between the amount of synthetic data and model performance, with diminishing returns observed after a certain point. This indicates that our method can effectively scale to generate substantial datasets of conversational search sessions for training purposes.

4.4. Human Evaluation

To complement the automatic evaluation metrics, we conducted a human evaluation study to assess the quality and relevance of the conversational search sessions generated by our LLM-CSSS method. We randomly sampled a set of conversational sessions generated by our method, ConvSDG, and the random generation baseline. We asked human evaluators to rate these sessions based on the following criteria: Coherence, Relevance, Informativeness, and Overall Quality.

Evaluators rated each session on a scale of 1 to 5, with 5 being the highest. The average scores across all evaluators for each method are presented in Table 2.

doi:10.20944/preprints202503.1401.v1

Table 2. Human Evaluation Results (Average Scores)

Method	Coherence	Relevance	Informativeness	Overall Quality
Randomly Generated Conversations	2.1	1.8	1.5	1.7
ConvSDG	3.5	3.8	3.6	3.7
LLM-CSSS (Ours)	**4.2**	**4.5**	**4.3**	**4.4**

The human evaluation results corroborate the findings from the automatic metrics, demonstrating that the conversational search sessions generated by our LLM-CSSS method are perceived as significantly superior in terms of coherence, relevance, informativeness, and overall quality compared to the baseline methods. This strong agreement between automatic and human evaluations further validates the effectiveness of our proposed approach.

4.5. Impact of Simulated Search Environment

To assess the contribution of our integrated simulated search environment, we conducted an ablation study where we trained our LLM-CSSS model without allowing it to interact with the simulated search environment during the generation of system responses. The results of this ablation are presented in Table 3.

Table 3. Impact of Simulated Search Environment on Retrieval-Based Model Performance

Data Source	MAP@10	NDCG@10
LLM-CSSS (with Search)	0.32	0.47
LLM-CSSS (without Search)	0.29	0.43

As shown in Table 3, the retrieval-based conversational search model trained on data generated by LLM-CSSS with the simulated search environment significantly outperforms the model trained on data generated without this crucial component. This highlights the importance of grounding the system responses in relevant search results to improve the overall quality and effectiveness of the generated conversational sessions.

4.6. Analysis of Conversation Turn Length

We analyzed the average number of turns in the conversational sessions generated by our LLM-CSSS method and the baseline approaches. The results are presented in Table 4.

Table 4. Average Number of Turns per Conversational Session

Method	Average Turns
Original Amazon Review Data	2.1
ConvSDG	3.5
Randomly Generated Conversations	4.8
LLM-CSSS (Ours)	4.1

Table 4 indicates that our LLM-CSSS method generates conversations with a reasonable average number of turns, falling between ConvSDG and the randomly generated baseline. This suggests that our method can generate multi-turn dialogues without becoming excessively verbose or remaining too short to capture meaningful interactions.

4.7. Performance Across Different Product Categories

To further investigate the robustness of our LLM-CSSS method, we evaluated the performance of the retrieval-based conversational search model trained on our generated data across different product categories within the Amazon Review dataset. We selected three diverse categories: "Electronics", "Books", and "Clothing, Shoes & Jewelry". The MAP@10 results for these categories are shown in Table 5.

Table 5. Retrieval-Based Model Performance Across Different Product Categories (MAP@10)

Data Source	Electronics	Books	Clothing, Shoes & Jewelry
Original Amazon Review Data	0.22	0.28	0.24
LLM-CSSS (Ours)	**0.30**	**0.35**	**0.31**

The results in Table 5 demonstrate that the retrieval-based model trained on data generated by our LLM-CSSS method consistently outperforms the model trained on the original data across these diverse product categories. This suggests that our method is not limited to specific types of products and can generate effective training data for conversational search across various domains.

4.8. Impact of Reinforcement Learning Stage

To understand the effect of the reinforcement learning stage in our training strategy, we compared the performance of the retrieval-based model trained with and without this stage. The results are presented in Table 6.

Table 6. Impact of Reinforcement Learning Stage on Retrieval-Based Model Performance

Data Source	MAP@10	NDCG@10
LLM-CSSS (with RL)	0.32	0.47
LLM-CSSS (without RL)	0.30	0.45

Table 6 shows that incorporating the reinforcement learning stage in our training process leads to a further improvement in the performance of the retrieval-based conversational search model. This indicates that optimizing the generated system responses based on relevance rewards can enhance the quality of the synthetic data and consequently improve the performance of downstream models.

5. Conclusion

In this paper, we presented LLM-Driven Conversational Search Session Synthesis (LLM-CSSS), a novel approach to generate synthetic conversational search data using large language models. Our method leverages the generative capabilities of LLMs and integrates a simulated search environment to produce realistic and relevant multi-turn dialogues. Through extensive experiments, we demonstrated the effectiveness of LLM-CSSS in generating high-quality training data for both retrieval-based and generation-based conversational search models, leading to significant performance gains compared to models trained on original data and data from baseline generation methods. Our ablation studies further highlighted the importance of the simulated search environment and the reinforcement learning stage in our training strategy. Moreover, human evaluation results corroborated the automatic metrics, confirming the superior quality of the conversations generated by our approach.

The findings of this research underscore the significant potential of large language models in tackling the data scarcity challenge within the field of conversational search. By providing a scalable and effective method for generating synthetic conversational data, our work contributes to the development of more robust and user-centric conversational information retrieval systems. Future work could explore the application of LLM-CSSS to other domains and datasets, investigate different methods for simulating the search environment, and explore the use of even larger and more sophisticated pre-trained language models to further enhance the quality and diversity of the generated conversational search sessions.

References

1. Mo, F.; Mao, K.; Zhao, Z.; Qian, H.; Chen, H.; Cheng, Y.; Li, X.; Zhu, Y.; Dou, Z.; Nie, J. A Survey of Conversational Search. *CoRR* **2024**, *abs/2410.15576*, [2410.15576]. https://doi.org/10.48550/ARXIV.2410.15576.
2. Soudani, H.; Petcu, R.; Kanoulas, E.; Hasibi, F. Data Augmentation for Conversational AI. In Proceedings of the Companion Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, Singapore, May 13-17, 2024; Chua, T.; Ngo, C.; Lee, R.K.; Kumar, R.; Lauw, H.W., Eds. ACM, 2024, pp. 1234–1237. https://doi.org/10.1145/3589335.3641238.
3. Zhou, Y.; Geng, X.; Shen, T.; Tao, C.; Long, G.; Lou, J.G.; Shen, J. Thread of thought unraveling chaotic contexts. *arXiv preprint arXiv:2311.08734* **2023**.
4. Huang, C.; Hsu, C.; Hsu, T.; Li, C.; Chen, Y. CONVERSER: Few-shot Conversational Dense Retrieval with Synthetic Data Generation. In Proceedings of the Proceedings of the 24th Meeting of the Special Interest Group on Discourse and Dialogue, SIGDIAL 2023, Prague, Czechia, September 11 - 15, 2023; Schlangen, D.; Stoyanchev, S.; Joty, S.; Dusek, O.; Kennington, C.; Alikhani, M., Eds. Association for Computational Linguistics, 2023, pp. 381–387. https://doi.org/10.18653/V1/2023.SIGDIAL-1.34.
5. Mo, F.; Yi, B.; Mao, K.; Qu, C.; Huang, K.; Nie, J. ConvSDG: Session Data Generation for Conversational Search. In Proceedings of the Companion Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, Singapore, May 13-17, 2024; Chua, T.; Ngo, C.; Lee, R.K.; Kumar, R.; Lauw, H.W., Eds. ACM, 2024, pp. 1634–1642. https://doi.org/10.1145/3589335.3651940.
6. Zhang, X.; Yang, H.; Young, E.F.Y. Attentional Transfer is All You Need: Technology-aware Layout Pattern Generation. In Proceedings of the 58th ACM/IEEE Design Automation Conference, DAC 2021, San Francisco, CA, USA, December 5-9, 2021. IEEE, 2021, pp. 169–174. https://doi.org/10.1109/DAC18074.2021.9586227.
7. Zhou, Y.; Li, X.; Wang, Q.; Shen, J. Visual In-Context Learning for Large Vision-Language Models. In Proceedings of the Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024. Association for Computational Linguistics, 2024, pp. 15890–15902.
8. Zhou, Y.; Zhang, J.; Chen, G.; Shen, J.; Cheng, Y. Less Is More: Vision Representation Compression for Efficient Video Generation with Large Language Models, 2024.
9. Zhou, Y.; Song, L.; Shen, J. Training Medical Large Vision-Language Models with Abnormal-Aware Feedback. *arXiv preprint arXiv:2501.01377* **2025**.
10. Zhou, Y.; Long, G. Style-Aware Contrastive Learning for Multi-Style Image Captioning. In Proceedings of the Findings of the Association for Computational Linguistics: EACL 2023, 2023, pp. 2257–2267.
11. McAuley, J.J.; Leskovec, J. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In Proceedings of the Proceedings of the 22nd international conference on World Wide Web, 2013, pp. 897–908.
12. Repplinger, J. G.G. Chowdhury. *Introduction to Modern Information Retrieval.* 3rd ed. London: Facet, 2010. 508p. alk. paper, $90 (ISBN 9781555707156). LC2010-013746. *Coll. Res. Libr.* **2011**, *72*, 194–195.
13. Zhang, Y.; Chen, X.; Ai, Q.; Yang, L.; Croft, W.B. Towards conversational search and recommendation: System ask, user respond. In Proceedings of the Proceedings of the 27th acm international conference on information and knowledge management, 2018, pp. 177–186.
14. Qu, C.; Yang, L.; Croft, W.B.; Zhang, Y.; Trippas, J.R.; Qiu, M. User intent prediction in information-seeking conversations. In Proceedings of the Proceedings of the 2019 Conference on Human Information Interaction and Retrieval, 2019, pp. 25–33.
15. Zhou, Y.; Geng, X.; Shen, T.; Long, G.; Jiang, D. Eventbert: A pre-trained model for event correlation reasoning. In Proceedings of the Proceedings of the ACM Web Conference 2022, 2022, pp. 850–859.
16. Al-Thani, H.; Elsayed, T.; Jansen, B.J. Improving conversational search with query reformulation using selective contextual history. *Data and Information Management* **2023**, *7*, 100025.
17. Zhou, Y.; Geng, X.; Shen, T.; Pei, J.; Zhang, W.; Jiang, D. Modeling event-pair relations in external knowledge graphs for script reasoning. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* **2021**.
18. Dixit, T.; Paranjape, B.; Hajishirzi, H.; Zettlemoyer, L. CORE: A retrieve-then-edit framework for counterfactual data generation. *arXiv preprint arXiv:2210.04873* **2022**.

19. He, S.; Zhang, S.; Zhang, X.; Feng, Z. Improve conversational search with multi-document information. In Proceedings of the International Conference on Neural Information Processing. Springer, 2023, pp. 3–15.
20. Wang, L.; Zhao, M.; Ji, H.; Jiang, Z.; Li, R.; Hu, Z.; Lu, X. Dialogue summarization enhanced response generation for multi-domain task-oriented dialogue systems. *Inf. Process. Manag.* **2024**, *61*, 103668. https://doi.org/10.1016/J.IPM.2024.103668.
21. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers); Burstein, J.; Doran, C.; Solorio, T., Eds. Association for Computational Linguistics, 2019, pp. 4171–4186. https://doi.org/10.18653/V1/N19-1423.
22. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I.; et al. Language models are unsupervised multitask learners. *OpenAI blog* **2019**, *1*, 9.
23. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR* **2019**, *abs/1907.11692*, [1907.11692].
24. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* **2020**, *21*, 140:1–140:67.
25. Clark, K.; Luong, M.; Le, Q.V.; Manning, C.D. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. In Proceedings of the 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020. OpenReview.net, 2020.
26. Thoppilan, R.; Freitas, D.D.; Hall, J.; Shazeer, N.; Kulshreshtha, A.; Cheng, H.; Jin, A.; Bos, T.; Baker, L.; Du, Y.; et al. LaMDA: Language Models for Dialog Applications. *CoRR* **2022**, *abs/2201.08239*, [2201.08239].
27. Chowdhery, A.; Narang, S.; Devlin, J.; Bosma, M.; Mishra, G.; Roberts, A.; Barham, P.; Chung, H.W.; Sutton, C.; Gehrmann, S.; et al. PaLM: Scaling Language Modeling with Pathways. *J. Mach. Learn. Res.* **2023**, *24*, 240:1–240:113.

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