Personalized Text Generation with LLMs: A Systematic Literature Review with a Focus on Healthcare

André Felipe dos S. Caraíba[1], **Sávio S. T. de Oliveira**[1], **Deborah S. A. Fernandes**[1]

[1]Instituto de Informática – Universidade Federal de Goiás
Goiania, GO.

andrecaraiba@discente.ufg.br, savioteles@ufg.br, deborah.fernandes@ufg.br

Abstract. *Large Language Models (LLMs) produce generic responses that fail to account for individual preferences and contexts. Personalized LLMs address this by leveraging user-specific data to generate tailored outputs. This paper presents a systematic literature review of 52 studies on personalized text generation, published between 2022 and 2025, with focus on healthcare applications.This study follows Kitchenham's guidelines for planning and conducting systematic literature reviews (SLRs), while the reporting of the review process adheres to the PRISMA framework. The analysis focuses on techniques, models, metrics, and challenges identified in the literature. The results indicate that prompting is the most adopted technique (55.8%), followed by fine-tuning (46.2%) and RAG (26.9%). Health applications remain underrepresented (9.6%) but are rapidly emerging, covering tasks such as personalized medical responses and mental health assistance. Key challenges include the lack of personalization-aware metrics, static preference assumptions, and privacy concerns.*

1. Introduction

Healthcare represents a particularly promising yet challenging domain for personalized text generation. Unlike other domains such as e-commerce or news, where personalization errors lead to irrelevant recommendations, in healthcare inaccurate outputs can compromise patient safety, clinical decisions, or cause psychological harm [Neupane et al. 2025, Shafi et al. 2025]. Personalization in this context must account for individual patient histories, medical terminology conventions, evolving clinical conditions, and even the therapeutic style of healthcare professionals [Gupta et al. 2025, Xie et al. 2025, Hu et al. 2024]. Additionally, privacy regulations severely limit the availability of real-world clinical data, motivating the use of synthetic datasets [Shafi et al. 2025, Xie et al. 2025]. Despite these unique requirements, it remains unclear how personalization techniques, metrics, and challenges have been specifically addressed in the healthcare context, and which gaps hinder the safe advancement of this area.

Large Language Models (LLMs) achieve state-of-the-art performance across NLP tasks but produce identical responses regardless of who is asking — the "one-size-fits-all flaw" [Liu et al. 2025b]. This motivates Personalized LLMs (PLLMs), which leverage user-specific data such as profiles, historical dialogues, past content, interactions, and preferences to deliver tailored responses [Liu et al. 2025b].

Recent approaches to building PLLMs operate at three levels: at the input level through prompting techniques such as profile augmentation, RAG, and soft prompting; at the model level through parameter-efficient fine-tuning methods; and at the objective level through alignment methods such as Reinforcement Learning [Liu et al. 2025b]. Evaluation combines lexical metrics (ROUGE, BLEU), semantic metrics (BERTScore), human evaluation, and LLM-as-a-judge.

The objective of this work is to systematically analyze the literature on personalized text generation with LLMs, with a specific focus on healthcare applications, identifying techniques, models, metrics, challenges, applications, and research gaps to orient future research, regulation, validation, and deployment in this domain. The remainder of this paper presents the method (Section 2), results (Section 3), discussion (Section 4), and conclusions (Section 5).

2. Method

According to [Kitchenham et al. 2004], a systematic literature review is a method used to identify, evaluate, and interpret relevant research studies related to a particular research question, research area, or phenomenon of interest. It represents a form of secondary study. This process involves three phases: planning, conducting, and reporting the review. In the planning phase, the need for the study is assessed and the research questions are defined, resulting in the development of a review protocol. In the conducting phase, the review is performed according to the procedures defined in the protocol. Finally, the reporting phase involves synthesizing and presenting the results obtained from the review. In this work, the review planning followed the guidelines proposed by Kitchenham for systematic literature reviews. The reporting of the review process, including the identification, screening, and selection of studies, followed the PRISMA framework.

2.1. Planning

In this work, a comprehensive literature review was conducted in November 2025. As part of the planning phase, the following information was defined:

- Research questions: (Q1) What are the main techniques and models for personalized text generation? (Q2) What evaluation metrics are used? (Q3) What are the main applications? (Q4) Which health domains and clinical tasks are addressed? (Q5) What techniques, models, and metrics are employed in health-related applications? (Q6) What challenges and future perspectives were identified?
- Digital libraries: Semantic Scholar, ACM Digital Library, IEEE Xplore, and Elsevier (ScienceDirect);
- Employed keywords: The search strings were adapted to the syntax of each digital library, as shown in Table 1.
- Inclusion criteria: We included original articles published between 2021 and 2025, written in English or Portuguese, with full text available. Studies had to: (i) address personalized text generation tasks, and (ii) explicitly use language models or large language models (LLMs), describing evaluation criteria for the model or proposed system.
- Exclusion criteria: We excluded: (i) duplicate records, (ii) secondary studies such as reviews, editorials, and letters, (iii) papers without full-text access, (iv) papers without evaluation parameters, (v) papers with insufficient methodological quality,

(vi) studies that did not address text generation tasks, (vii) studies that did not use text generation models, and (viii) studies focused exclusively on vision-language models (VLM).

Table 1. Search strings used in each digital library.

Database	Search String
IEEE Xplore	(”Abstract”:personal*) AND (”Abstract”:large language model* OR ”Abstract”:llm* OR ”Abstract”:text generation) NOT (”Abstract”:image) NOT (”Abstract”:recommend*)
ACM	[Abstract: personal*] AND [[Abstract: large language models] OR [Abstract: llm]] AND NOT [Abstract: images] AND NOT [Abstract: recommend*] AND [E-Publication Date: (01/01/2021 TO 12/31/2025)]
Elsevier	(personalization OR personalized) AND (large language models OR llm) AND (generation OR generate) AND NOT (recommendations OR image)
Semantic Scholar	(personalized text generation language) OR (personalization text generation language) OR (LLMs personalization language) OR (Large Language Models Personalization) OR (personalization llm). Filtered by ACL and EMNLP venues.

2.2. Conducting

The initial search identified several articles, which were deduplicated and screened for relevance. Articles that did not meet the inclusion criteria were systematically excluded following PRISMA guidelines [Page et al. 2021]. The PRISMA flow diagram was used to illustrate the selection process (Figure 1) and the detailed description of each selection stage is provided in Section 3.1.

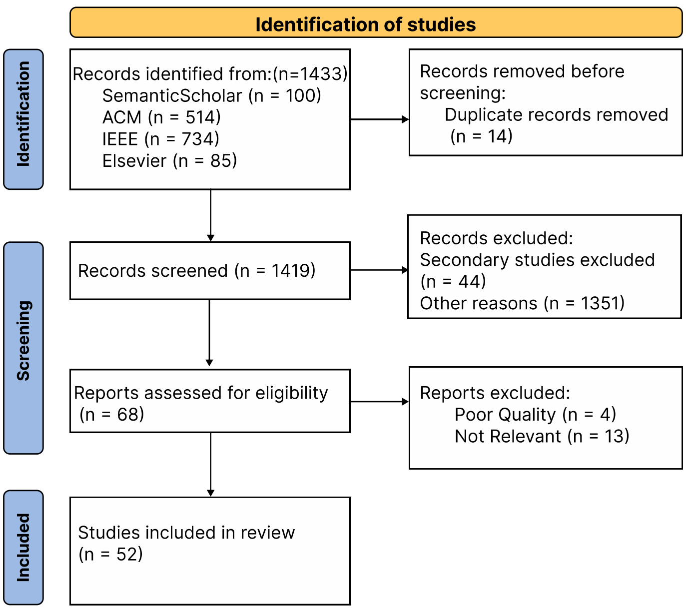

Figure 1. PRISMA flow diagram illustrating the study selection process.

3. Results

3.1. Study selection and data extraction

As illustrated in Figure 1, a comprehensive search was conducted across Semantic Scholar, ACM Digital Library, IEEE Xplore, and Elsevier by applying the defined search strings to each database (2.1). For Semantic Scholar, target conferences in the NLP field—namely ACL and EMNLP—were also selected. The search yielded 1433 records (Semantic Scholar $n = 100$, ACM $n = 514$, IEEE $n = 734$, Elsevier $n = 85$), from which 14 duplicates were removed. The remaining 1419 records were screened by title and abstract, excluding 44 secondary studies and 1351 for other reasons, leaving 68 articles for full-text assessment. Of these, 17 were excluded (4 for poor quality and 13 for lack of relevance), resulting in 52 included studies. The complete selection flow is illustrated in Figure 1.

The distribution of publications shows marked growth: no relevant article was found in 2021, 2 were published in 2022, 5 in 2023, 21 in 2024, and 24 in 2025, confirming that personalized text generation with LLMs is a rapidly emerging field.

3.2. Summary of articles evaluated

Among the 52 included studies, 34 (65.4%) focused on personalized text generation and 18 (34.6%) on personalized dialogue generation. General-purpose applications dominated (63.5%), while health-related domains accounted for 9.6%, followed by recipe generation (5.8%), e-commerce and product reviews (3.8% each), news (3.8%), and persuasion (3.8%). Historical content was the most common source of personalization (20 studies, 38.5%), followed by historical dialogues (16, 30.8%), user profiles (7, 13.5%), historical interactions (7, 13.5%), and human preferences (6, 11.5%). Table 2 presents the distribution of studies by publication venue, and Table 3 summarizes the five health-focused studies.

Table 2. Distribution of included studies by publication venue.

Venue	N	References
ACL	16	[Xie et al. 2025, Ryan et al. 2025, Liu et al. 2025a, Zhang et al. 2025b, Padmakumar et al. 2025, Qiu et al. 2025, Balepur et al. 2025, Zhang et al. 2025a, Oh et al. 2025, Liu et al. 2025c, Ji et al. 2025, Zhang et al. 2025c, Salemi et al. 2024b, Dai et al. 2024, Huang et al. 2024, Xu et al. 2023]
EMNLP	7	[Zhang 2024, Tan et al. 2024a, Tan et al. 2024b, Kirstein et al. 2024, Tang et al. 2024, Wang et al. 2024, Lee et al. 2024a]
WWW	4	[Lian et al. 2025, Prahlad et al. 2025, Li et al. 2024, Baek et al. 2024]
SIGIR	2	[Joko et al. 2024, Salemi et al. 2024a]
IEEE	17	[Shafi et al. 2025, Huang et al. 2025, Wang 2025, Nezhad et al. 2025, Wang 2024, Hu et al. 2024, Gui and Wang 2024, Hang et al. 2024, Ju and Wang 2024, Nezhad and Kangavari 2024, Lee et al. 2024b, Zhang et al. 2023, Han et al. 2024, Ji 2023, Zhan et al. 2023, Srivastava et al. 2022, Ke and Chen 2021]
Other	6	[Kim et al. 2025, Wasilewski 2025, Gupta et al. 2025, Neupane et al. 2025, Huq et al. 2025, Fendji et al. 2025]

3.3. Research Questions

The studies were conducted with the aim of addressing the established research questions. Therefore, in this section, these questions are revisited and discussed in light of the analyzed studies, which provide the basis for the answers presented.

Q1 – Techniques and models for personalized text generation

Table 3. Health-focused studies on personalized text generation. PT = Prompting, FT = Fine-Tuning, RAG = Retrieval-Augmented Generation, PA = Profile Augmentation, DS = Dataset Construction.

Reference	Year	Clinical Task	Techniques
[Gupta et al. 2025]	2025	Radiology reports	PA, PT
[Shafi et al. 2025]	2025	Mental health assistance	FT, PT, RAG
[Xie et al. 2025]	2025	Psychological counseling	DS, FT, PT
[Neupane et al. 2025]	2025	Medical responses	PA, PT, RAG
[Hu et al. 2024]	2024	Medical dialogues	FT

The analysis of techniques reveals that prompting was the most widely adopted approach, employed in 29 studies (55.8%), followed by fine-tuning in 24 (46.2%) and retrieval-augmented generation (RAG) in 14 (26.9%). Profile augmentation was used in 12 studies (23.1%), while alignment-based methods appeared in 5 (9.6%). Other techniques included reinforcement learning (6 studies), soft prompting (5), direct preference optimization (DPO) (3), knowledge graphs (2), multi-agent systems (2), prompt optimization (2), and steering (1). Notably, 3 studies integrated memory-augmented retrieval strategies, combining external memory modules with RAG pipelines to maintain personalized context across extended interactions [Wang et al. 2024, Baek et al. 2024, Zhang et al. 2023]. Additionally, 3 studies invested in the construction of dedicated personalization datasets, particularly for domains with limited user-specific data such as psychological counseling [Xie et al. 2025] and preference-aware dialogue [Balepur et al. 2025, Joko et al. 2024]. Many studies combined multiple techniques, reflecting a trend toward hybrid approaches for personalization.

In terms of models, GPT-4 was the most frequently used generator, appearing in 9 studies, closely followed by LLaMA-2-7B and GPT-3.5-turbo, each present in 7 studies. Other commonly used models included GPT-4o and Flan-T5-XXL-11B (4 studies each), LLaMA-3.1-8B, GPT-3.5, and GPT-4o-mini (3–4 studies each), as well as Mistral-7B and FlanT5-base (3 each). The results indicate a strong preference for both proprietary (GPT family) and open-source (LLaMA, Flan-T5, Mistral) models across the literature.

Q2 – Evaluation metrics for personalized text generation

The evaluation of personalized text generation relied on a diverse set of metrics. ROUGE-L was the most commonly used (20 studies, 38.5%), followed by human evaluation (19, 36.5%), ROUGE-1 (17, 32.7%), and BLEU (16, 30.8%). LLM-as-a-judge emerged as a notable trend (15 studies, 28.8%). Additional metrics included ROUGE-2 (8), BERTScore (5), perplexity and Dist-1/2 (4 each), win rate (4), METEOR, F1, Jaccard, and domain-specific measures.

Q3 – Applications for personalized text generation

General-purpose personalized text generation was the dominant focus (63.5%), encompassing tasks such as summarization, review generation, note-taking, and headline generation. Personalized dialogue generation accounted for 34.6%, including role-playing and conversational agents. Health applications represented 9.6%, with other domains including recipe generation (5.8%), persuasion (3.8%), e-commerce (3.8%), and code generation (1.9%).

Q4 – Health domains and clinical tasks in personalized text generation

Among the 52 studies, 5 (9.6%) specifically addressed health-related domains. The clinical tasks covered included personalized medical response generation (2 studies), personalized mental health assistance (1 study), radiology report generation (1 study), and psychological counseling through digital twins (1 study). Mental health emerged as a prominent area of interest, with 2 of the 5 health studies focusing on psychological support and counseling. The remaining health studies targeted clinical report generation and patient-facing medical dialogue. All health-focused studies were published between 2024 and 2025, suggesting that the application of personalized text generation in healthcare is a recent and growing research direction.

Q5 – Techniques, models, and metrics in health applications

In the health domain, prompting was the most frequently adopted technique (4 out of 5 studies, 80%), followed by fine-tuning (3 studies, 60%) and RAG (2 studies, 40%). Profile augmentation was also used in 2 studies to incorporate patient-specific information into the generation pipeline.

Regarding models, GPT-4 appeared in 2 health-focused studies, while other models included Claude Opus, Gemini, LLaMA 3.2-3B, Qwen2-7B, GPT-3.5-turbo, Mistral-7B, and domain-specific models such as BenTsao, HuatuoGPT2-7B, and Zhongjing. The use of domain-specific medical models in one study highlights the potential of specialized architectures for clinical applications.

The evaluation of health-related studies relied heavily on ROUGE variants (ROUGE-1 and ROUGE-L appearing in 4 studies each, ROUGE-2 in 3), BLEU (3 studies), and human evaluation (3 studies). Notably, some studies also employed domain-specific metrics such as RAGAS, TruLens, and clinical adequacy measures, reflecting the need for evaluation criteria tailored to the healthcare context.

Q6 – Challenges and future perspectives

The reviewed studies reported recurring challenges grouped into six themes:

- Evaluation limitations: Conventional metrics such as ROUGE and BLEU are insufficient for capturing personalization quality, particularly for long-form text [Salemi et al. 2024b]. Standardized evaluation protocols remain absent [Zhang et al. 2025c], and LLM-as-a-judge, while scalable, introduces its own biases.
- Static user preferences: Several methods assume fixed preferences over time [Oh et al. 2025, Zhang et al. 2025b, Zhang et al. 2025a], yet in practice preferences evolve with context and task. Future work should explore continual learning and dynamic preference modeling.
- Privacy and data security: Models fine-tuned on private data are susceptible to extraction attacks [Salemi et al. 2024b, Liu et al. 2025b]. Per-user PEFT strategies improve privacy but reduce performance [Liu et al. 2025b], highlighting the need for federated learning and differential privacy approaches.
- Data scarcity and single-source personalization: Many approaches rely on a single data type, limiting their ability to capture complex preferences [Zhang 2024]. Methods requiring pairwise feedback face additional constraints in low-resource scenarios [Ryan et al. 2025].

- Scalability and computational cost: Fixed top-k retrieval may miss relevant content [Wang et al. 2024], while fine-tuning is resource-intensive and difficult to scale [Zhang et al. 2025b, Lee et al. 2024a].
- Cross-domain generalization: Most studies evaluate within a single domain, raising generalizability concerns. StyleVector [Zhang et al. 2025a] and DPL [Qiu et al. 2025] noted that personalization may not transfer across tasks. Future benchmarks should incorporate cross-domain scenarios.

Promising future directions include: personalization-aware metrics, continual learning frameworks, hybrid approaches combining user embeddings with retrieval [Liu et al. 2025c], multimodal personalization [Zhang 2024], explainability methods [Neupane et al. 2025], and integration with real-time clinical data [Shafi et al. 2025].

4. Discussion

This study analyzed 52 articles on personalized text generation published between 2022 and 2025. The sharp rise from 2023 onward confirms that the convergence of LLMs with personalization techniques is a recent phenomenon. Prompting-based methods dominate, likely due to their compatibility with proprietary models and lower computational costs, though their insufficiency for deep personalization [Tan et al. 2024b, Zhang et al. 2025b] has driven a trend toward hybrid approaches. The coexistence of proprietary and open-source models reveals a dual dynamic: proprietary models are favored for evaluation and profile augmentation [Xie et al. 2025, Neupane et al. 2025], while open-source models are essential for fine-tuning, reproducibility, and privacy-sensitive deployments. The heavy reliance on lexical metrics such as ROUGE and BLEU is concerning, as these do not distinguish genuinely personalized outputs from high-quality generic ones [Salemi et al. 2024b, Zhang et al. 2025c], and the absence of personalization-specific metrics remains a critical gap.

Health applications, though underrepresented, introduce unique requirements that distinguish them from all other domains. In e-commerce, a personalization error results in an irrelevant suggestion; in healthcare, it may lead to a missed diagnosis, inappropriate treatment, or emotional harm. Clinical accuracy must not be sacrificed for stylistic adaptation [Neupane et al. 2025, Hu et al. 2024]. Privacy constraints are particularly stringent, with patient data regulated by frameworks such as HIPAA and models fine-tuned on private data susceptible to extraction attacks, motivating the use of synthetic data [Shafi et al. 2025, Xie et al. 2025]. Mental health applications further require emotionally sensitive responses and must contend with extreme data scarcity due to ethical restrictions [Xie et al. 2025, Shafi et al. 2025]. The need for transparency in clinical outputs [Neupane et al. 2025] and the evolving nature of patient conditions challenge the static preference assumptions prevalent in many methods [Oh et al. 2025, Zhang et al. 2025b, Zhang et al. 2025a]. The scarcity of health-focused studies (only 5 out of 52) reveals a significant gap between the potential of personalized text generation for healthcare and the current research effort, with impact across subdomains such as mental health, radiology, medical dialogues, and patient-facing clinical support.

5. Conclusions

This work presents a systematic review of personalized text generation using large language models, with a specific focus on healthcare applications. By analyzing 52 studies published between 2022 and 2025 following PRISMA guidelines, this review maps techniques, models, evaluation metrics, challenges, and application domains, being, to the best of our knowledge, the first to specifically investigate this intersection. The analysis reveals a field in rapid expansion, oriented toward hybrid techniques combining prompting, fine-tuning, and RAG, yet facing challenges such as the lack of personalization-aware metrics, static preference assumptions, privacy concerns, and limited cross-domain generalization.

Health-related studies represent only 9.6% of the reviewed literature, a significant gap given that healthcare imposes unique requirements not found in other domains: clinical accuracy that cannot be sacrificed for stylistic adaptation, strict privacy regulations, sensitivity to evolving patient conditions, and emotionally appropriate responses in mental health contexts.

Acknowledgments

The authors thank the Centro de Excelência em Inteligência Artificial (CEIA) at the Universidade Federal de Goiás (UFG) for the support provided to this work.

References

Baek, J., Chandrasekaran, N., Cucerzan, S., Herring, A., and Jauhar, S. K. (2024). Knowledge-augmented large language models for personalized contextual query suggestion. In *Proceedings of the ACM Web Conference 2024*, pages 3355–3366.

Balepur, N., Padmakumar, V., Yang, F., Feng, S., Rudinger, R., and Boyd-Graber, J. L. (2025). Whose boat does it float? improving personalization in preference tuning via inferred user personas. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3371–3393.

Dai, Z., Yao, C., Han, W., Yuanying, Y., Gao, Z., and Chen, J. (2024). Mpcoder: Multi-user personalized code generator with explicit and implicit style representation learning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3765–3780.

Fendji, J. L. K., Donatien, D., and Atemkeng, M. (2025). Hybrid profile based multi-document text summarisation. *Procedia Computer Science*, 252:862–872.

Gui, H. and Wang, Z. (2024). Response generation with personal attributes and act information. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Gupta, A., Hussain, M., Nikhileshwar, K., Rastogi, A., and Rangarajan, K. (2025). Integrating large language models into radiology workflow: Impact of generating personalized report templates from summary. *European Journal of Radiology*, 189:112198.

Han, A., Koushik, N., Bidarkundi, N., Shehata, M. N., Kunchipudi, V., Mammadli, T., Mehta, S., and Lerner, O. (2024). Enhancing llm conversational acuity using pragmatic measures. In *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*, pages 886–891. IEEE.

Hang, C. N., Tan, C. W., and Yu, P.-D. (2024). Mcqgen: A large language model-driven mcq generator for personalized learning. *IEEE Access*, 12:102261–102273.

Hu, Z., Zhao, H., Zhao, Y., Xu, S., and Xu, B. (2024). T-agent: A term-aware agent for medical dialogue generation. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Huang, B.-S., Kuo, T. T., Wang, L.-J., and Lin, C.-Y. (2025). Remind: Recall enhanced memory integration for natural language dialogue systems. In *2025 IEEE Conference on Artificial Intelligence (CAI)*, pages 1–6. IEEE.

Huang, Q., Liu, X., Ko, T., Wu, B., Wang, W., Zhang, Y., and Tang, L. (2024). Selective prompting tuning for personalized conversations with llms. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 16212–16226.

Huq, F., Samee, A., Lin, D. C.-E., Tang, A. X., and Bigham, J. P. (2025). Noteeline: Supporting real-time, personalized notetaking with llm-enhanced micronotes. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*, pages 1064–1081.

Ji, B. (2023). Based on text augmentation personalized dialog generation with persona-sparse data. In *2023 4th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, pages 717–720. IEEE.

Ji, K., Lian, Y., Li, L., Gao, J., Li, W., and Dai, B. (2025). Enhancing persona consistency for llms' role-playing using persona-aware contrastive learning. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 26221–26238.

Joko, H., Chatterjee, S., Ramsay, A., De Vries, A. P., Dalton, J., and Hasibi, F. (2024). Doing personal laps: Llm-augmented dialogue construction for personalized multi-session conversational search. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 796–806.

Ju, S. and Wang, C. (2024). Beyond simple text style transfer: Unveiling compound text style transfer with prompt-based pre-trained language models. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6850–6854. IEEE.

Ke, S.-W. and Chen, W.-L. (2021). Stylized dialogue generation. In *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 1456–1460. IEEE.

Kim, T., Agarwal, D., Ackerman, J., and Saha, M. (2025). Steering ai-driven personalization of scientific text for general audiences. *Proceedings of the ACM on Human-Computer Interaction*, 9(7):1–28.

Kirstein, F., Ruas, T., Kratel, R., and Gipp, B. (2024). Tell me what i need to know: Exploring llm-based (personalized) abstractive multi-source meeting summarization. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 920–939.

Kitchenham, B. et al. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26.

Lee, G., Jeong, M., Kim, Y., Jung, H., Oh, J., Kim, S., and Yun, S.-Y. (2024a). Bapo: Base-anchored preference optimization for overcoming forgetting in large language models personalization. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 6804–6820.

Lee, G. H., Lee, K. J., Jeong, B., and Kim, T. (2024b). Developing personalized marketing service using generative ai. *Ieee Access*, 12:22394–22402.

Li, C., Zhang, M., Mei, Q., Kong, W., and Bendersky, M. (2024). Learning to rewrite prompts for personalized text generation. In *Proceedings of the ACM Web Conference 2024*, pages 3367–3378.

Lian, J., Ao, X., Liu, X., Liu, Y., and He, Q. (2025). Panoramic interests: Stylistic-content aware personalized headline generation. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 1109–1112.

Liu, D., Wu, Z., Song, D., and Huang, H.-Y. (2025a). A persona-aware llm-enhanced framework for multi-session personalized dialogue generation. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 103–123.

Liu, J., Qiu, Z., Li, Z., Dai, Q., Yu, W., Zhu, J., Hu, M., Yang, M., Chua, T.-S., and King, I. (2025b). A survey of personalized large language models: Progress and future directions. *arXiv preprint arXiv:2502.11528*.

Liu, J., Zhu, Y., Wang, S., Wei, X., Min, E., Lu, Y., Wang, S., Yin, D., and Dou, Z. (2025c). Llms+ persona-plug= personalized llms. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9373–9385.

Neupane, S., Mitra, S., Mittal, S., Gaur, M., Golilarz, N. A., Rahimi, S., and Amirlatifi, A. (2025). Medinsight: A multi-source context augmentation framework for generating patient-centric medical responses using large language models. *ACM Transactions on Computing for Healthcare*, 6(2):1–19.

Nezhad, M. M. and Kangavari, M. (2024). Personalized persuasive text generation. In *2024 10th International Conference on Artificial Intelligence and Robotics (QICAR)*, pages 261–267. IEEE.

Nezhad, M. M., Kisomi, M. A., and Gholinezhad, F. (2025). Adaptive persuasion in conversational ai: An llm-driven framework for dynamic strategy switching via personality and sentiment analysis. In *2025 11th International Conference on Web Research (ICWR)*, pages 145–149. IEEE.

Oh, M., Lee, S., and Ok, J. (2025). Comparison-based active preference learning for multi-dimensional personalization. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 33145–33166.

Padmakumar, V., Wang, Z., Arbour, D., and Healey, J. (2025). Principled content selection to generate diverse and personalized multi-document summaries. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 29884–29899.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., et al. (2021). The prisma 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, 372.

Prahlad, D., Lee, C., Kim, D., and Kim, H. (2025). Personalizing large language models using retrieval augmented generation and knowledge graph. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 1259–1263.

Qiu, Y., Zhao, X., Zhang, Y., Bai, Y., Wang, W., Cheng, H., Feng, F., and Chua, T.-S. (2025). Measuring what makes you unique: Difference-aware user modeling for enhancing llm personalization. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 21258–21277.

Ryan, M. J., Shaikh, O., Bhagirath, A., Frees, D., Held, W. B., and Yang, D. (2025). Synthesizeme! inducing persona-guided prompts for personalized reward models in llms. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8045–8078.

Salemi, A., Kallumadi, S., and Zamani, H. (2024a). Optimization methods for personalizing large language models through retrieval augmentation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 752–762.

Salemi, A., Mysore, S., Bendersky, M., and Zamani, H. (2024b). Lamp: When large language models meet personalization. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7370–7392.

Shafi, F. R., Hossain, M. A., and Choudhury, S. (2025). Personalized mental health assistance with large language models. In *2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 815–823. IEEE.

Srivastava, H., Sunil, S., and Kumari, K. S. (2022). Neural text style transfer with custom language styles for personalized communication systems. In *2022 International Conference on Knowledge Engineering and Communication Systems (ICKES)*, pages 1–6. IEEE.

Tan, Z., Liu, Z., and Jiang, M. (2024a). Personalized pieces: Efficient personalized large language models through collaborative efforts. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6459–6475.

Tan, Z., Zeng, Q., Tian, Y., Liu, Z., Yin, B., and Jiang, M. (2024b). Democratizing large language models via personalized parameter-efficient fine-tuning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6476–6491.

Tang, Y., Wang, B., Zhao, D., Jinxiaojia, J., Zhangjijun, Z., He, R., and Hou, Y. (2024). Morpheus: Modeling role from personalized dialogue history by exploring and utilizing latent space. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 7664–7676.

Wang, Y. (2024). Iterative feedback-enhanced prompting: A green algorithm for reducing household food waste. In *2024 International Conference on Machine Learning and Applications (ICMLA)*, pages 1–4. IEEE.

Wang, Y. (2025). Personalized llm response generation system for reducing household food waste. In *2025 IEEE Conference on Technologies for Sustainability (SusTech)*, pages 1–7. IEEE.

Wang, Z., Li, Z., Jiang, Z., Tu, D., and Shi, W. (2024). Crafting personalized agents through retrieval-augmented generation on editable memory graphs. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 4891–4906.

Wasilewski, A. (2025). Harnessing generative ai for personalized e-commerce product descriptions: A framework and practical insights. *Computer Standards & Interfaces*, 94:104012.

Xie, H., Chen, Y., Xing, X., Lin, J., and Xu, X. (2025). Psydt: Using llms to construct the digital twin of psychological counselor with personalized counseling style for psychological counseling. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1081–1115.

Xu, H., Liu, H., Lv, Z., Yang, Q., and Wang, W. (2023). Pre-trained personalized review summarization with effective salience estimation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 10743–10754.

Zhan, H., Lin, X., Cui, S., Zhao, Z., Zhou, W., and Chen, H. (2023). Towards zero-shot personalized table-to-text generation with contrastive persona distillation. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

Zhang, J. (2024). Guided profile generation improves personalization with large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 4005–4016.

Zhang, J., Liu, Y., Wang, W., Liu, Q., Wu, S., Wang, L., and Chua, T.-S. (2025a). Personalized text generation with contrastive activation steering. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7128–7141.

Zhang, L., Wu, J., Zhou, D., and He, Y. (2025b). Proper: A progressive learning framework for personalized large language models with group-level adaptation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16399–16411.

Zhang, Y., He, Y., and Zhou, D. (2025c). Rehearse with user: Personalized opinion summarization via role-playing based on large language models. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 15194–15211.

Zhang, Y., Yu, Z., Jiang, W., Shen, Y., and Li, J. (2023). Long-term memory for large language models through topic-based vector database. In *2023 International Conference on Asian Language Processing (IALP)*, pages 258–264. IEEE.