

Personalized Summarization of Documents Using Role-Aware User Profiles

¹Manvi Breja, ² Dhruv Ghildiyal

1, 2 The NorthCap University, Gurugram, Haryana, India,

*Corresponding Author: Manvi Breja (manvibreja@ncuindia.edu)

Information of Article

Article history:

Received: 15 May 2026

Revised: 22 July 2026

Accepted: 29 July 2026

Online: 20 October 2026

Keywords:

Personalized document
summarization

role-aware summarization

user profiling

natural language processing

large language models

ABSTRACT

The rapid spread of digital documents across multiple online sources has exaggerated the need to implement effective document summarizers. Traditional document summarization techniques generate normal summaries that may or may not fulfill users' exact needs. Users with different roles and specializations seek various types of focused information from the documents. This paper analyzes different summarization techniques with different application areas. Further, the paper proposes a framework for personalized document summarization, which is based on role-based user profiles. The summaries of documents are created according to the user's needs. The need is captured from the users' role, their interests, and their interaction within their personal and professional networks. The proposed framework integrates the techniques of NLP, user profiling, and feature modelling to generate relevant personalized document summaries.

1. Introduction

With the widespread amount of digital information across diverse resources like research papers, academic repositories, policy reports, online articles, blogs, and other knowledge bases, users are overwhelmed by the volume of information. There is a need for automatic text summarization from one or multiple documents to reduce information overload. Nowadays, readers don't have enough time to read the whole document and extract valuable information out of it; instead, they require a system that can generate a summary valuable to the user. Most existing summarization techniques are user-agnostic, producing the same summary regardless of individual user needs, which highlights the varying ways users prefer to consume content. Users seek different information from documents based on their desired roles, such as researchers, educators, and students in the research domain, or managers, policy makers, and analysts in the policy and legal domain, depending on various real-world scenarios and application contexts. Each user with a different role seeks a different perspective on the same document. This motivates the need to implement a personalized summarization module that can summarize the documents based on the users' roles.

Traditional summarization systems, including both extractive and abstractive, summarize the text irrespective of the users' role. Li and Chaturvedi [1] highlighted the importance of reader's personalization for multi-document summarization. For example, a medical report can be summarized differently for a patient and doctor, where the summarized medical report comprises medical terms for the doctor, whereas a simple summary is for a patient. The shortcoming in all existing summarization models is that they consider the same priority of the reader, i.e., one summary can fulfill every user's needs. This provides the basis to work on role-aware personalization.

Role-aware summarization summarizes the document according to the user's background, expertise, and intent. The system utilizes a user log and creates a user profile that records their occupation, goals, intention, and expertise. When a summary is needed, the system generates a document summary tailored to the user. Venington et al. [2] highlighted personalized multi-document summarization, where the authors integrated the preferences of users with their reading logs, which are utilized by the AI model to rank the content of each user. Fendji et al. [3] integrated the concepts of both extractive and abstractive methods to produce user-oriented summaries.

With the widespread increase in online data available, it is very difficult to filter out the relevant and easy-to-understand information. Zhang et al. [4] succinctly summarized their review on text summarization using LLMs, finding that even a high-quality summarization AI does not effectively address personalization. The research produced cohesive summaries of the content, but it did not consider the needs of individual readers. The resulted document is either too technical, too vapid, or simply not relevant.

The “how” is achieved through the integration of natural language processing (NLP), deep learning, and user modeling. In practice, the system first creates a structured user model: for instance, a researcher may have “technical depth: high” and “goal: data insights,” while a policymaker may be characterized by “technical depth: medium” and “goal: understanding implications.” When a document is to be summarized, the model (which is often based on transformer-based architectures, as in Roy and Mercer [5] analyses in part, document summarization in a scientific context and uses both content and user models to produce a relevant summary. The EROS framework by Singh et al. [6] provides another aspect: it illustrates how the control of entity coverage (such as “Data,” “Usage,” and “Target”) in policy document summaries ensures that essential information is available for particular purposes, e.g., privacy officers/end-users.

The motivation to do this study arises from a gap in existing summarization types, where most systems focus only on analyzing content and do not take reader comprehension into account. Understanding is always contextual in the real world. For example, Zaman et al. [7] present simplification-aware summarization to make scientific texts more accessible for non-experts, which shows the importance of contextualization and, thus, the adaptation to the audience's capacity. Yet, these attempts often stop with highly readable texts and do not explore the avenue of personalization. This study aims to develop a system that closes that gap by combining role-aware user profiling with modern summarization models.

2. Literature Review

This section discusses key research papers focused on document summarization, and Table 1 highlights the methodologies used for summarization techniques, along with their key findings and tools.

Table 1: Significant research findings on document summarization

Authors	Goal / Target	Methodology	Key Findings	Shortcomings	Tools / Models / Datasets Used
Zhang et al. [8]	Create a user-subjective dataset for personalized summarization.	Annotated personalized summaries reflect diverse reader needs and explore user signals.	LLMs without personalization struggle; personal signals (entities, topics) are essential.	Focuses on the dataset, not a personalization model.	PersonalSum dataset, LLM few-shot evaluation.
Dasgupta et al. [9]	Improve evaluation metrics for personalized summarization.	The designed PerSEval metric correlated with human judgments.	PerSEval better reflects personalization than accuracy metrics like ROUGE.	Only evaluation; no new summarizer model.	PerSEval metric, PENS benchmark.
Vansh et al. [10]	Propose new personalization evaluation metrics (EGISES).	Developed the EGISES metric and benchmarking.	Standard metrics miss personalization; EGISES better captures user differences.	Early approach; not widely adopted yet.	PENS dataset, summarization models for testing.

Authors	Goal / Target	Methodology	Key Findings	Shortcomings	Tools / Models / Datasets Used
Khurana et al. [11]	Personalized extractive summaries for scientific texts.	wNMF includes positive/negative user knowledge signals.	Tailors summaries to user knowledge preferences.	Extractive only—lacks abstraction.	The study focuses on wNMF, semantic evaluation, and four scientific corpora.
Ghodratnama & Zakershahra [12]	Generate summaries by learning from user feedback.	The process involves two stages: first, extracting preferences, and second, generating results through a feedback loop.	Personalized summaries aligned well with feedback.	Requires interactive feedback – scalability challenge.	User preference extractor, personalized summarizer.
Ghodratnama et al. [13]	Interactive personalization via concept preferences.	Users accept/reject concepts; iterative summary refinement.	Better inclusion of desired content.	Interactive loop—hard to scale automatically.	Concept selection, human feedback.
Nam et al. [14]	Condition RLHF on user summary signals for personalization.	Learn user info summaries to condition the reward models.	Personalized outputs in broader settings improved.	Not specific to document summarization alone.	RLHF, user preference summarizer.
Li & Chaturvedi [1]	Personalize multi-document summaries via preference comparisons.	Comparative personalization among user profiles.	Enhanced relevance & coherence for individuals.	Requires detailed preference data per user.	LLMs with comparative profiling and evaluation frameworks.
Chatterjee et al. [15]	Improve dataset diversity for personalized summarization.	PerAugy augmentation (trajectory shuffling + perturbation).	Better personalization learning with augmented data.	Synthetic diversity may not match real user patterns.	Data augmentation, AUC & PSE-SU4 metrics.
Ning et al. [16]	Summarize educational texts based on explicit user intent.	Encoder for user intent + pointer-generator transformer.	Better ROUGE, BERTScore, and factual consistency.	Domain-specific (education); generalization unknown.	EduLect-Sum & EduCN-Sum datasets; LLM backbone.
Avinesh et al. [17]	Interactive multi-doc summarization with user concepts.	ILP optimization + user concept selection.	More relevant summaries with direct preference input.	Requires manual interaction.	Optimization solver, concept input.
Díaz & Gervás [18]	Early user-model-based personalization.	User profiles with long/short-term interests.	Personalized summaries are more relevant than generic.	Classic extractive—outdated by current neural methods.	User profile model, extractive summarizer.

Authors	Goal / Target	Methodology	Key Findings	Shortcomings	Tools / Models / Datasets Used
Kolagar & Zarcone [19]	Build a practical personalized summarization tool for education.	LLM + user survey for personalization in the UI.	Improved relevance for specific users.	More of a system/demo, not core algorithmic innovation.	GPT-4 via LangChain, Streamlit UI.

3. Research Gap with Aligned Research Objectives

3.1 Research Gap

From the above-discussed literature review, this section discusses the systematic research gaps that open new avenues for producing the systematic research objectives:

Gap 1: Limited Use of Role-Aware User Modelling: Most personalization approaches utilize topic/entity preferences, feedback loops, and intent-aware summarization. With these methods, the user roles are either assumed or ignored. Currently, role-playing LLM approaches exist, but they do not maintain role persistence. Thus, there is no unified framework that summarizes the documents by integrating user role, preferences, and intent together.

Gap 2: Weak Personalization in LLM-Based Summarization: LLMs help generate generalized summaries of documents very fluently but could not give role-specific summaries. One must engage in prompt engineering to achieve the desired results. The gap is the absence of a structured, role-aware mechanism without prompting.

Gap 3: Lack of End-to-End Personalized Summarization Frameworks: Since existing work focuses on components such as creating datasets, evaluation metrics, and approaches for extracting preferences. There is no unified end-to-end pipeline that implements a framework integrating context-aware roles, preferences extraction, and evaluating personalized summaries.

Gap 4: Evaluation Metrics Do Not Fully Capture Role Alignment: In existing research, PerSEval [20] improves the performance of personalization but does not measure the role of relevance. Along with this, ROUGE/BERTScore also fails to assess whether the summary suits the user's role. Thus, there is a lack of role-oriented evaluation metrics.

Gap 5: Limited Role Generalization Across Domains: Nowadays most of the studies are domain-specific, such as education, science, and policy, with role behavior not being tested across domains. Thus, there is no such study focused on cross-domain role generalization.

3.2 Proposed Research Objectives

Based on the above-identified research gaps, the following are the research objectives proposed:

RO1: To design and validate a role-aware personalized document summarization framework that integrates explicit user roles, preferences, and intent to generate summaries aligned with individual user needs.

RO2: To model different user roles (e.g., student, researcher, policymaker, general reader) and analyze their influence on summary content selection, granularity, and tone.

RO3: To propose an end-to-end architecture for personalized summarization that combines role-aware user profiling, preference extraction, and LLM-based summary generation

RO4: To develop a role-conditioned summarization mechanism that ensures role alignment during abstractive summary generation beyond prompt-based methods.

RO5: To evaluate the proposed framework using personalization-aware and role-specific evaluation metrics.

RO6: To compare role-aware personalized summaries against generic summaries, preference-only personalized summaries, and prompt-based role-playing summaries

RO7: To assess the robustness and transferability of role-aware personalization across multiple document domains (e.g., scientific, educational, and policy).

4. Proposed Work

Existing document summarization techniques are afflicted with limitations such as (i) (i) personalization according to keywords or prompts, (ii) summaries not updated according to users' needs, and (iii) lack of user role modelling. These limitations reduce the efficacy of summaries depending on stakeholders with diverse needs and expertise. To address these limitations, this section proposes a role-enhanced personalized document summarization framework that generates summaries according to the users' needs, preferences, demands, and expertise. The proposed summarization framework incorporates user roles as a core component, which takes single and multiple documents as an input and generates a personalized role-enhanced summary.

The framework integrates user profiles with transformer-based summarization modules, which dynamically update the users' interests and expertise and generate distinct summaries for users. For example, a research paper plays a distinct and significant role for a researcher, undergraduate, and teacher. Similarly, a medical report plays a different role for the doctor, researcher, and patient. Each user profile is represented by the following components:

- **User role**, which describes the intent of each user depending on his/her role.
- **Domain Expertise Level**, which depicts the users' expertise in particular domains.
- **Preferences in summarized content** comprise the length of the summary generated with topic focus and the level of abstraction required in generating the summaries.
- **Personalized logs**, which help in adaptive personalization.

These components are included as attributes in the user profile vector, which helps in generating the correct summary content from the documents. This integration of user roles into the summarization framework helps generate relevant summaries from the documents adapted to the users' needs.

5. Architecture of the Proposed Work

This section discusses the architectural flow of the proposed framework and the workings of different modules in the framework. The framework comprises and integrates document preprocessing, semantic content representation, role-aware user profiling, role-based content scoring, and personalized summary generation, followed by feedback-driven adaptation. Figure 1 depicts the architecture diagram of the framework.

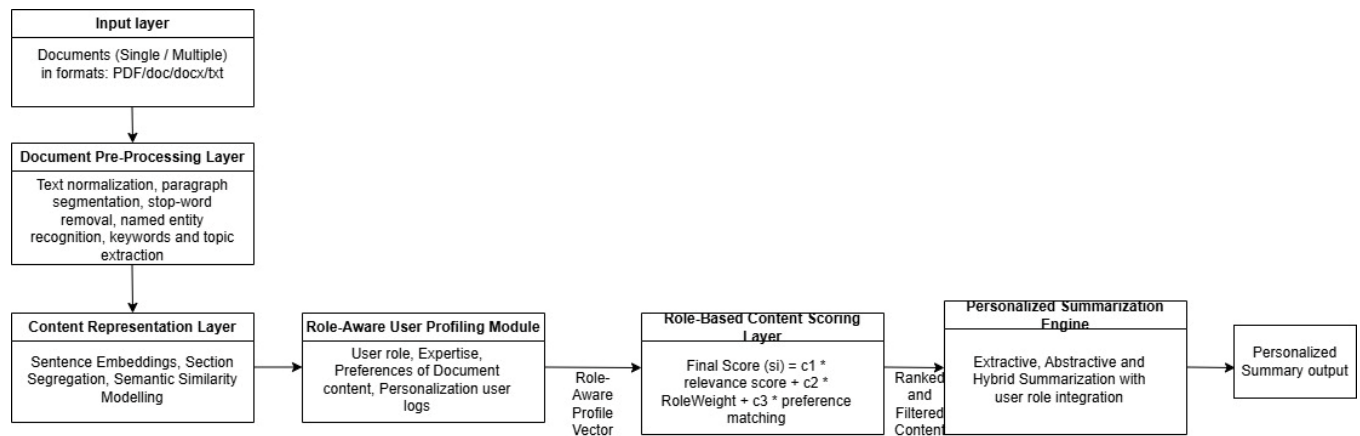


Fig. 1: Architectural Flow Diagram of Proposed Summarization Framework

Module-Wise Description of the Proposed Architecture

5.1 Input Layer:

The input layer is capable of gathering the documents from different heterogeneous sources. The framework is adapted to support both single-document and multi-document. These documents can be in different formats, such as PDF, doc, docx, and txt. This framework can be adapted across diverse applications such as academics, policy analysis, enterprises, medical ports, and so on.

5.2 Document Pre-processing Layer:

The document preprocessing layer processes the raw documents and transforms them into a clean and structured representation. This comprises several NLP tasks such as text cleaning, normalization, paragraph segmentation, stop-word removal, Named Entity Recognition (NER), and keyphrase and topic extraction. These NLP tasks help annotate the documents, which further enables content- and role-based representation.

5.3 Content Representation Layer:

This layer converts the pre-processed documents into semantic vector representations. This layer utilizes transformer-based embedding models to convert into sentence-level embeddings. The structured semantic representation identifies the topic relevance of documents and helps to identify the personalization and section focus.

5.4 Role-Aware User Profiling Module:

This module of the framework models the individual characteristics of users to ingest personalized document summarization. The role-aware user profile comprises (1) users' role with capturing their intent and need, (2) domain expertise, which captures the users' familiarity and expertise in the particular domain, (3) preferences of the content of the document, capturing the length of the summary and their content focus, and (4) personalized logs to ensure the adaptive user's personalization. These features are integrated into the user profile to produce a role-aware profile vector, which integrates the scoring of document content with personalized summary generation.

5.5 Role-Based Content Scoring Layer:

This layer helps to evaluate, score, and rank the segmented documents based on their relevance to the users' needs. A composite score of the summarized document is calculated by integrating semantic document relevancy, users' role alignment, and matching users' preferences. A weighted score depicts the dynamically adjusted summary according to the user's needs and preferences, which results in adaptive, fine-grained, personalized document summaries.

$$\text{Score}(s) = \alpha \cdot \text{Relevance}(s) + \beta \cdot \text{RoleAlignment}(s) + \gamma \cdot \text{PreferenceMatch}(s) \quad (1)$$

5.6 Personalized Summarization Engine:

This engine helps generate the final personalized summary by utilizing extractive, abstractive, and hybrid summarization techniques [21]. Further, the engine integrates the users' role and their preferences to generate a focused summary with granularity of the user's needs.

5.7 Personalized Summary Output Layer:

This module helps in producing the role-enhanced document summary, which ensures that the content of the document is correctly aligned with the users' objectives and preferences and adheres to the desired segment length with different abstraction levels. The output is shown through dashboards and decision-support systems.

6. Algorithm for Role-Aware Personalized Document Summarization

This section discusses the detailed step-by-step algorithm as in Algorithm 1, used for implementing the role-enhanced personalized document summarization framework.

Algorithm 1: Role-Aware Personalized Document Summarization

Input:

$D = \{d1, d2, ..., dn\}$ // Input documents

$U = \{\text{role, expertise, preferences, history}\}$ // User profile

L = desired summary length

Output:

S = Personalized summary

Steps:

1. Preprocess documents D

- Clean text
- Segment sentences
- Extract entities and keyphrases

2. Generate semantic embeddings for each sentence $s_i \in D$.

3. Construct Role-Aware Profile Vector (RAPV) from U

4. For each sentence, s_i :

- Compute relevance score $R(s_i)$
- Compute role alignment score $RA(s_i, \text{RAPV})$
- Compute preference match score $PM(s_i, \text{RAPV})$
- $\text{FinalScore}(s_i) = \alpha R(s_i) + \beta RA(s_i) + \gamma PM(s_i)$

5. Rank sentences based on final score.

6. Select top- k sentences according to summary length L .

7. Generate summary S using:

- Extractive / Abstractive / Hybrid model
- Inject role-aware control signals

8. Collect user feedback and update RAPV.

9. Return personalized summary S

7. Application-Oriented Mapping of Role-Aware Personalized Document Summarization Framework

This section discusses in detail different applications and use cases where this framework can be introduced to enhance the efficiency of the summarized documents. Table 2 describes different application domains.

Table 2: Application-oriented mapping of personalized document summarization

Application / Use Case	User Roles	Recommended Datasets	Recommended Tools / Models	Summarization Strategy
Academic Research Documents	Undergraduate, Postgraduate, Researcher, Reviewer	PubMed, CL-SciSumm, SciSummNet [22]	SciBERT, SBERT, PEGASUS, BART-Large, SciSpaCy.	Hybrid (extractive + abstractive), section-aware summarization
Education & E-Learning	Student, Instructor, Curriculum Designer	CNN/Daily Mail, XSum, S2ORC [23]	T5 (Base/Large), KeyBERT, Readability metrics (FKGL)	Abstractive with simplification-aware summarization
Healthcare & Clinical Documents	Clinician, Researcher, Patient, Administrator	PubMed, MEDIQA-Sum [24], MIMIC-III	BioBERT, ClinicalBERT, PEGASUS (medical), SciSpaCy	Extractive for clinicians; abstractive for patients
Policy, Legal & Government Documents	Policymaker, Legal Expert, Analyst, Citizen	GovReport, BillSum [25], EUR-LEX	Long-T5, Longformer-Encoder-Decoder, NER pipelines	Entity-driven, controlled extractive + abstractive
Corporate Knowledge Management	Executive, Manager, Engineer	AMI Meeting Corpus, Multi-News	BART, T5, BERTopic	Topic-focused hybrid summarization
News & Media	Journalist, Analyst, General Reader	CNN/Daily Mail, Multi-News, Newsroom	BART-CNN, event extraction, sentiment analysis	Fact-preserving abstractive summarization
Social Media & Online Communities	Moderator, Community Member, Researcher	Reddit TIFU [26], Multi-Dialog, Twitter/X datasets	Dialogue-T5, PEGASUS-Dialog, and stance detection tools	Opinion-aware and dialogue summarization

8. Analysis of Existing Document Summarization Systems vs. Proposed Framework

This section compares the proposed framework with the existing document summarizers. Table 3 compares the personalized document summarizer with the existing summarizer.

Table 3: Comparative Analysis of Proposed Document Summarizer with existing summarizations

Comparison Dimension	Existing Document Summarization Systems	Proposed Role-Aware Personalized Summarization Framework
User Modeling	No explicit user modeling or shallow preference-based profiles	Explicit role-aware user profiles incorporating role, expertise level, and intent

Comparison Dimension	Existing Document Summarization Systems	Proposed Role-Aware Personalized Summarization Framework
Personalization Depth	Limited to keyword overlap or reading history	Fine-grained personalization using role-based content weighting and filtering
Summary Style	Uniform summary for all users	Adaptive summary length, abstraction level, and terminology per role
Summarization Technique	Primarily extractive or abstractive	Hybrid (extractive + abstractive + controlled) strategy selected per role and domain
Domain Adaptability	Domain-specific models; limited cross-domain use	Domain-agnostic core with domain adapters (academic, healthcare, legal, policy, etc.)
Application Coverage	Typically single-domain (news, scientific articles)	Multi-domain and multi-role applicability
User Satisfaction	Implicitly assumed	Explicitly optimized through role relevance and cognitive load reduction
Real-World Suitability	Limited in high-stakes domains	Suitable for high-stakes domains (healthcare, legal, policy)

9. Conclusion and Future Scope

This paper proposed a role-enhanced personalized document summarization framework to address the challenges of generic summarizations. After reviewing the literature, it was analyzed that recent summarization approaches, despite being implemented by deep learning and large language models, ignore the real context of the user, which results in generic and insufficient information according to the user's role. To address such limitations, the proposed pipeline integrates role modelling, content filtering, and dynamic strategies.

Incorporating preferences of users, their roles, expertise, and needs enables generating document summaries that are contextually relevant and efficiently enhanced according to users' needs. The proposed architecture integrated the extractive, abstractive, and hybrid summarization techniques to make the summarization system adaptable to diverse application areas such as academics, research, education, healthcare, law records, and policy analysis, and so on.

In conclusion, the framework significantly improves the explainability and transparency of summarization for deployment in high-end domains. The proposed personalized framework with human-centric assessment largely targets fulfilling the need of the user and thus shifting the paradigm from content-centric to user-centric summary generation.

References

- [1] H. Li and S. Chaturvedi, "Comparative Personalization for Multi-document Summarization," arXiv preprint <https://doi.org/10.48550/arXiv.2509.21562>.
- [2] K. Veningston, P. V. Rao, and M. Ronalda, "Personalized multi-document text summarization using deep learning techniques," *Procedia Computer Science*, vol. 218, pp. 1220–1228, 2023 <https://doi.org/10.1016/j.procs.2023.01.100>.
- [3] J. L. K. Fendji, D. Donatien, and M. Atemkeng, "Hybrid profile-based multi-document text summarization," *Procedia Computer Science*, vol. 252, pp. 862–872, 2025.
- [4] H. Zhang, X. Liu, and J. Zhang, "Extractive summarization via ChatGPT for faithful summary generation," arXiv preprint arXiv:2304.04193, 2023 <https://doi.org/10.1016/j.procs.2025.01.047>

- [5] S. S. Roy and R. E. Mercer, “Enhancing scientific document summarization with research community perspective and background knowledge,” in Proc. Joint Int. Conf. Computational Linguistics and Language Resources and Evaluation (LREC-COLING), 2024, pp. 6048–6058 <https://doi.org/10.18653/v1/2023.findings-emnlp.214>
- [6] J. Singh, S. Fazili, R. Jain, and M. S. Akhtar, “EROS: Entity-driven controlled policy document summarization,” arXiv preprint arXiv:2403.00141, 2024 <https://doi.org/10.63317/4y96285ovxbu>
- [7] F. Zaman, F. Kamiran, M. Shardlow, S. U. Hassan, A. Karim, and N. R. Aljohani, “SATS: Simplification-aware text summarization of scientific documents,” *Frontiers in Artificial Intelligence*, vol. 7, Art. no. 1375419, 2024 <https://doi.org/10.63317/3rwmson3bq4h>
- [8] L. Zhang, P. Liu, M. Henriksboe, E. Lauvraak, J. A. Gulla, and H. Ramampiaro, “PersonalSum: A user-subjective guided personalized summarization dataset for large language models,” *Advances in Neural Information Processing Systems*, vol. 37, pp. 99333–99346, 2024 <https://doi.org/10.3389/frai.2024.1375419>
- [9] S. Dasgupta, A. Chander, P. Borad, I. Motiyani, and T. Chakraborty, “PerSEval: Assessing personalization in text summarizers,” arXiv preprint arXiv:2407.00453, 2024 <https://doi.org/10.52202/079017-3151>
- [10] R. Vansh, D. Rank, S. Dasgupta, and T. Chakraborty, “Accuracy is not enough: Evaluating personalization in summarizers,” in *Findings of the Association for Computational Linguistics: EMNLP*, 2023, pp. 2582–2595 <https://doi.org/10.48550/arXiv.2407.00453v2>
- [11] A. Khurana, V. Bhatnagar, and V. Kumar, “Personalized summarization of scientific scholarly texts,” *Expert Systems with Applications*, vol. 277, Art. no. 127177, 2025 <https://doi.org/10.18653/v1/2023.findings-emnlp.169>
- [12] S. Ghodratnama and M. Zakershahraak, “SumRecom: A personalized summarization approach by learning from users’ feedback,” arXiv preprint arXiv:2408.07294, 2024 <https://doi.org/10.1016/j.eswa.2025.127177>
- [13] S. Ghodratnama, M. Zakershahraak, and F. Sobhanmanesh, “Adaptive summaries: A personalized concept-based summarization approach by learning from users’ feedback,” in *Proc. Int. Conf. Service-Oriented Computing*, Cham, Switzerland: Springer, 2020, pp. 281–293 https://doi.org/10.1007/978-3-030-76352-7_29
- [14] H. Nam, Y. Wan, M. Liu, J. Lian, P. Ahnn, and N. Jaques, “Learning to summarize user information for personalized reinforcement learning from human feedback,” arXiv preprint arXiv:2507.13579, 2025 https://doi.org/10.1007/978-3-030-76352-7_29
- [15] P. Chatterjee, S. Sonawane, A. Hengle, A. Tanna, S. Dasgupta, and T. Chakraborty, “Diversity augmentation of dynamic user preference data for boosting personalized text summarizers,” arXiv preprint arXiv:2510.10082, 2025 <https://doi.org/10.48550/arXiv.2507.13579v4>
- [16] W. Ning, S. Shi, Y. Yin, Z. Guo, and J. Xie, “Intent-aware personalized summarization for educational texts with large language models,” *Journal of King Saud University – Computer and Information Sciences*, vol. 37, no. 9, p. 308, 2025 <https://doi.org/10.48550/arXiv.2510.10082v1>
- [17] P. V. S. Avinesh, C. Binnig, B. Hättasch, C. M. Meyer, and O. Özyurt, “Sherlock: A system for interactive summarization of large text collections,” *Proc. VLDB Endowment*, vol. 11, no. 12, pp. 1902–1905, 2018 <https://doi.org/10.1007/s44443-025-00333-7>
- [18] A. Díaz and P. Gervás, “User-model based personalized summarization,” *Information Processing & Management*, vol. 43, no. 6, pp. 1715–1734, 2007 <https://doi.org/10.1145/3328519.3329129>
- [19] Z. Kolagar and A. Zarcone, “HumSum: A personalized lecture summarization tool for humanities students using LLMs,” in *Proc. 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE)*, 2024, pp. 36–70 <https://doi.org/10.1016/j.ipm.2007.01.009>
- [20] R. Ahmed and D. J. Hemanth, “Hybrid text summarization: Integrating extractive and abstractive models for enhanced cross-domain summarization,” *Intelligent Decision Technologies*, 2025 <https://doi.org/10.18653/v1/2024.personalize-1.4>
- [21] L. Li, Y. Xie, W. Liu, Y. Liu, Y. Jiang, S. Qi, and X. Li, “CIST@CL-SciSumm 2020, LongSumm 2020: Automatic scientific document summarization,” in *Proc. First Workshop on Scholarly Document Processing*, 2020, pp. 225–234 <https://doi.org/10.1177/18724981251322745>
- [22] K. Lo, L. L. Wang, M. Neumann, R. Kinney, and D. S. Weld, “S2ORC: The Semantic Scholar Open Research Corpus,” in *Proc. 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020, pp. 4969–4983 <https://doi.org/10.18653/v1/2020.sdp-1.25>

- [23] W. W. Yim, A. B. Abacha, G. Adams, N. Snider, and M. Yetisgen, “Overview of the MEDIQA-Sum task at ImageCLEF 2023: Summarization and classification of doctor-patient conversations,” in CLEF Working Notes, 2023, pp. 1347–1360 <https://doi.org/10.18653/v1/2020.acl-main.447>
- [24] A. Kornilova and V. Eidelman, “BillSum: A corpus for automatic summarization of U.S. legislation,” in Proc. 2nd Workshop on New Frontiers in Summarization, 2019, pp. 48–56 <https://doi.org/10.18653/v1/2023.clinicalnlp-1.52>
- [25] Reddit TIFU Dataset, “Reddit TIFU dataset,” available at https://www.tensorflow.org/datasets/catalog/reddit_tifu, accessed Jan. 5, 2026 <https://doi.org/10.18653/v1/d19-5406>
- [26] E. Daraghmi, L. Atwe, and A. Jaber, “A comparative study of PEGASUS, BART, and T5 for text summarization across diverse datasets,” *Future Internet*, vol. 17, no. 9, Art. no. 389, 2025.