

Article

Not peer-reviewed version

SHARE: Synthesizing Heterogeneous Autism-Support Records into Evidence-Based Recommendations

[Saumya Chauhan](#)^{*} and Mila Hong

Posted Date: 28 November 2025

doi: 10.20944/preprints202511.2299.v1

Keywords: Keywords: large language models; human-in-the-loop AI; recommender systems; explainable AI



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

SHARE: Synthesizing Heterogeneous Autism-Support Records into Evidence-Based Recommendations[†]

Saumya Chauhan * and Mila Hong *

California Institute of Technology; 1200 E. California Blvd., Pasadena, CA 91125; saumya.s.chau@gmail.com, milazoehwork@gmail.com

[†] **Note.** This is the authors' preprint version of the paper to be presented at the 40th AAAI Conference on Artificial Intelligence (AAAI-26). The final version will appear in the AAAI-26 proceedings.

Abstract

Supporting children with Autism Spectrum Disorder (ASD) requires highly individualized knowledge, but crucial information is often scattered across documents such as Individualized Education Plans (IEPs), diagnostic assessments, and caregiver notes. Thus, we propose SHARE (Synthesis of Heterogeneous Autism-support Records into Evidence-based Recommendations), a framework that transforms diverse autism-related documents into a concise, actionable set of recommendations directed towards caregivers of children with autism. Recommendations are generated with OpenAI's large language model API, grounded in user-provided evidence with optional web-based augmentation for missing details, and each recommendation is citation-linked to ensure traceability. When caregivers rate attempted recommendations, SHARE applies a lightweight Bayesian bandit with Upper Confidence Bound (UCB) re-ranking to refine and personalize future outputs. This adaptive feedback loop sets SHARE apart from prior systems, which have focused on static goal drafting or summaries, by combining LLM-based generation, caregiver input, and interpretable ranking in a pipeline that evolves over time.

Keywords: large language models; human-in-the-loop AI; recommender systems; explainable AI

1. Introduction

ASD (Autism Spectrum Disorder) is characterized by challenges in social, nonverbal, repetitive communication and behavior [1] that is estimated to affect 1 in 127 children globally [2], and 1 in 31 children in the US [1]. Over the past 20 years, we've seen a almost a 300% increase in ASD diagnoses largely cited to (1) an expansion in the definition of ASD and (2) an increase in ASD screening for children between 18-24 months as promoted by successful public health programs [3]. With this increase in ASD diagnosis, further intervention and early childhood support is needed.

However, scalable and homogeneous care for children with ASD becomes challenging, as ASD is characterized by clinical heterogeneity across affected genes, behavioral attributes, and developmental projections [4]. Therefore, ASD support necessitates thorough individualized understanding of a child's needs.

However, vital information such sensory triggers, accommodations, and progress milestones are often trapped in various documents, notes, or staff memories. This leads to time-consuming onboarding, missed insights, and low continuity of care, especially as staff turn over [5].

1.1. Related Works

Thus, ASD research largely focuses on developing diagnostic tools, such as at-home multimodal evaluations [6]. Beyond diagnostics, both academia and industry have concentrated on Individualized Education Plans (IEPs): structured documents outlining a child's goals, accommodations, and services [7]. Many efforts explore how LLMs can generate IEP goals [8,9] or summarize IEPs [10]. In parallel,

bandit algorithms, particularly UCB-style methods, are widely used in large-scale recommender systems to balance exploration and exploitation [11]. However, this mathematically-grounded approach has not been applied to autism-related document synthesis. SHARE builds on the aforementioned group of prior work by synthesizing insights across diverse documents and delivering them to audiences beyond educators, such as caregivers, who may not have the time to read entire IEPs on the job.

1.2. Our Approach

We present SHARE, a system that organizes autism-related records from across a child's life into clear, evidence-linked recommendations for caregivers. The system works in three steps: (1) it accepts common document formats or zip files, including IEPs, diagnostic reports, teacher notes, and caregiver feedback; (2) it cleans and splits the text into smaller chunks with stable IDs so each recommendation can be traced back; and (3) it generates recommendations using OpenAI API's large language model with enforced citations. Missing details are flagged, and outputs are further refined through caregiver feedback in a simple Gradio-based UI.

Previous tools in this space have focused mainly on generating goals or fixed summaries [8–10,12]. SHARE goes further by combining multi-document, citation-grounded recommendations with an adaptive ranking process. At the core of this process is a lightweight bandit strategy, adapting UCB-style selection from large-scale recommendation systems [11], where caregiver ratings update probabilities and guide the balance between exploring new suggestions and reinforcing proven ones. This yields actionable recommendations with an interpretable reason for their ranking, helping caregivers trust and act on the system's outputs.

2. System Architecture

All stages of SHARE are accessible through a Gradio-based interface, launched from a Google Colab notebook (see Supplementary Materials).

2.1. Data Source Ingestion

SHARE begins by ingesting user-uploaded documents, including diagnostic assessments, teacher observations, caregiver feedback, or IEPs. Supplying more documents gives a fuller picture of a child's needs, triggers, and behaviors. Files can be uploaded in DOCX, TXT, or PDF format, as well as in a .zip folder containing documents of the same type. Text is then parsed with the pypdf library for page-level extraction and python-docx for educator notes. Finally, SHARE standardizes whitespace, removes personal identifiers (when chosen by the user), and normalizes formatting to prepare text for downstream processing.

2.2. Chunking and Stable IDs

Next, the preprocessed text is tokenized using OpenAI's tiktoken library (cl100k_base), segmented into windows of about 750 tokens, and assigned a stable identifier at a segment-by-segment level. These IDs provide stable citations and identifiers such that each insight can be traced back to its source.

2.3. Recommendation Generation

Through the UI, caregivers trigger generation using the OpenAI API with gpt-4.1-mini. Outputs must contain (1) a brief summary, (2) 5–10 recommendations with at least one citation in the form (source: file#chunkNN) and (3) a “missing information” section. No recommendation is provided without cited evidence. If enabled, a second pass with web_search fills gaps with web citations of the form (web: Title, URL).

2.4. Post-Processing and UI

After generation, SHARE saves outputs in Markdown and JSON. A parser extracts citation-bearing lines into a schema with IDs, text, and version history. In the Gradio interface, caregivers can edit and rate recommendations (1–10). These ratings drive refinement of the top- k recommendations selected by the ranking algorithm, producing updated outputs. Results can be exported as a .zip bundle or used to restart the cycle of re-rating and re-ranking on the refined set.

2.5. Ranking Algorithm

We use a ranking algorithm to convert the user-generated ratings into Bayesian posteriors. Each recommendation starts with a uniform prior as modeled by a Beta distribution:

$$\text{prior} \sim \text{Beta}(\alpha=1, \beta=1).$$

We convert each rating $r \in \{1, \dots, 10\}$ into a fractional score $p = r/10$ and update the posterior parameters as

$$\alpha \leftarrow \alpha + p, \quad \beta \leftarrow \beta + (1 - p).$$

And define μ as an estimation of the usefulness or effectiveness of a recommendation to the user.

$$\mu = \frac{\alpha}{\alpha + \beta}$$

Thus, as high ratings increase α , the recommendation's current μ or usefulness increases. Likewise, as low ratings increase β , the recommendation's current μ or usefulness decreases. Since users do not have to rate every item, but we want to encourage the exploration of un-tried or less-rated recommendations, we also include an Upper Confidence Bound (UCB):

$$\text{score} = \mu + \sqrt{\frac{2 \ln T}{\alpha + \beta}},$$

where T is the total number of ratings across all items. As the user rates more and more recommendations, the second term decreases, giving new/rarely rated items a higher score.

2.6. Conclusion and Future Work

We introduce SHARE, a framework that organizes diverse autism-related records into concise recommendations, each traceable to its original document and open to caregiver refinement. For future work, we plan to extend beyond text input by incorporating videos of the child across different settings, enabling recommendations to be supported with specific video segments alongside text citations.

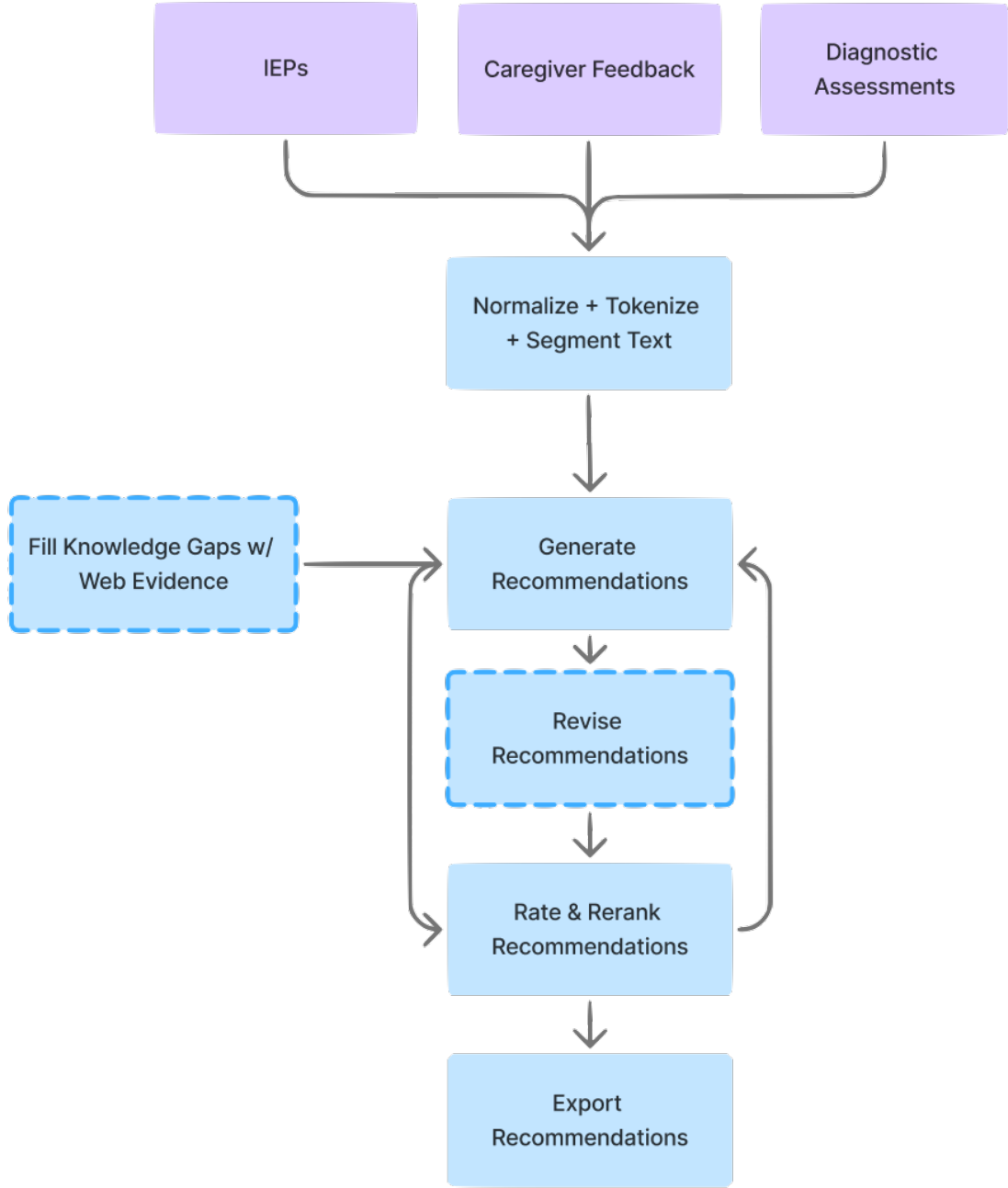


Figure 1. System pipeline. Document ingestion, chunking, generation with citations, re-ranking, rewriting, and export.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

References

1. Autism Speaks. What is Autism. In Proceedings of the Autism Speaks Information Pages. Autism Speaks, 2023.
2. World Health Organization. Autism Spectrum Disorders. In Proceedings of the WHO Fact Sheets. World Health Organization, 2023.
3. Johns Hopkins Bloomberg School of Public Health. Is There an Autism Epidemic? In Proceedings of the Public Health On Call / Johns Hopkins Magazine. Johns Hopkins University, 2025.

4. Siller, M. Editorial: Individualizing Interventions for Young Children With Autism: Embracing the Next Generation of Intervention Research. In *Proceedings of the Journal of the American Academy of Child & Adolescent Psychiatry*. Elsevier, 2021, pp. 680–682. <https://doi.org/10.1016/j.jaac.2020.11.003>.
5. Sulek, R.; Trembath, D.; Paynter, J.; Keen, D.; Simpson, K. Inconsistent staffing and its impact on service delivery in ASD early-intervention. In *Proceedings of the Research in Developmental Disabilities*. Elsevier, 2017, pp. 18–27. <https://doi.org/10.1016/j.ridd.2017.02.007>.
6. Abbas, H.; Garberson, F.; Liu-Mayo, S.; Glover, E.; Wall, D.P. Multi-modular AI Approach to Streamline Autism Diagnosis in Young Children. *Scientific Reports* **2020**, *10*. <https://doi.org/10.1038/s41598-020-61213-w>.
7. University of Washington. What is an Individualized Education Plan?, 2025.
8. Waterfield, D.A. Examining IEP Goals Written with and Without ChatGPT. *Journal of Special Education Technology* **2025**, *40*, 3–13. <https://doi.org/10.1177/01626434251324592>.
9. Monsha. How to Generate IEP Goals Using AI, 2025.
10. Botón, S.B.G.; InnovateUS. AI for Family Advocacy and Learning: Making Individualized Education Plans Accessible, 2025. Workshop, InnovateUS; showcasing Burnes Center’s IEP summarization and translation tool.
11. Yi, X.; Wang, S.C.; He, R.; Chandrasekaran, H.; Wu, C.; Heldt, L.; Hong, L.; Chen, M.; Chi, E.H. Online Matching: A Real-time Bandit System for Large-scale Recommendations. In *Proceedings of the Proceedings of the Seventeenth ACM Conference on Recommender Systems (RecSys '23)*, New York, NY, USA, 2023; pp. 1–10. <https://doi.org/10.1145/3604915.3608792>.
12. Chu, L.; Wu, H.; Pan, Y. ChatASD: A Dialogue Framework for LLMs Enhanced by Autism Knowledge Graph Retrieval. In *Proceedings of the Proceedings of the 15th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '24)*, 2024.

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.