

Using AI to Enhance Healthcare Resource Management and Allocation: A Focus on the Autism Community in Alabama

--Manuscript Draft--

Manuscript Number:	PONE-D-25-02370R3
Article Type:	Research Article
Full Title:	Using AI to Enhance Healthcare Resource Management and Allocation: A Focus on the Autism Community in Alabama
Short Title:	Using AI to Enhance Healthcare Resource Management and Allocation: A Focus on the Autism Community in Alabama
Corresponding Author:	Daniel Adamek, Ph.D. The University of Alabama in Huntsville Huntsville, Alabama UNITED STATES OF AMERICA
Keywords:	natural language processing (NLP); large language models (LLMs); retrieval-augmented generation (RAG); Autism Resource Management; AI in Social Work; Healthcare Delivery Efficiency
Abstract:	<p>This study investigates the potential of artificial intelligence, particularly Natural Language Processing and large-scale language models, to improve resource management and service access for individuals with autism in Alabama. The research aims to address existing challenges in navigating complex datasets and enhancing social work practices through the development and evaluation of AI-driven tools. We designed and tested AI systems, including general language models, domain-specific chatbots powered by advanced language models, and a Retrieval-Augmented Generation framework. A standardized set of queries was used to simulate real-world scenarios encountered by social workers and families seeking autism-related resources. System performance was evaluated based on precision, recall, and response accuracy.</p> <p>Results demonstrated that the Retrieval-Augmented Generation framework achieved superior performance compared to traditional methods and general models. The integration of domain-specific information significantly enhanced the accuracy, contextual relevance, and usability of the generated responses. This improvement was particularly evident for queries requiring geographic specificity, multiple criteria, or complex contextual understanding.</p> <p>The findings highlight the transformative potential of AI-driven tools in improving social work efficiency and enhancing healthcare equity. By streamlining care coordination and delivering accurate, contextually relevant information, these systems offer scalable solutions to improve access to autism-related services. Future research should focus on addressing data quality, minimizing biases, and ensuring ethical deployment to build trust and support widespread adoption of these tools.</p>
Order of Authors:	Armin Ahmadi, Ph.D., MBA Jerome Baudry, Ph.D. Nathan Tenhundfeld, Ph.D. Kelly Goff, MS, BCBA, LBA Daniel Adamek, Ph.D.
Opposed Reviewers:	
Response to Reviewers:	Thank you so much for all your comments. We have answered them in Athena-RespondTo_additional_comments_Aug2025.docx file which is uploaded along with the manuscript with Track Change.
Additional Information:	
Question	Response
Financial Disclosure	Yes

Enter a financial disclosure statement that describes the sources of funding for the work included in this submission. Review the [submission guidelines](#) for detailed requirements. View published research articles from [PLOS ONE](#) for specific examples.

This statement is required for submission and **will appear in the published article** if the submission is accepted. Please make sure it is accurate.

Funded studies

Enter a statement with the following details:

- Initials of the authors who received each award
- Grant numbers awarded to each author
- The full name of each funder
- URL of each funder website
- Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript?

Did you receive funding for this work?

Please add funding details.
as follow-up to "**Financial Disclosure**

Enter a financial disclosure statement that describes the sources of funding for the work included in this submission. Review the [submission guidelines](#) for detailed requirements. View published research articles from [PLOS ONE](#) for specific examples.

This statement is required for submission and **will appear in the published article** if the submission is accepted. Please make sure it is accurate.

This research was financially supported in part by Little Orange Fish (grant 25159B 2D2AC), and in part by by the Office of the Provost, The University of Alabama in Huntsville. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

<p>Funded studies</p> <p>Enter a statement with the following details:</p> <ul style="list-style-type: none"> • Initials of the authors who received each award • Grant numbers awarded to each author • The full name of each funder • URL of each funder website • Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript? 	
<p>Did you receive funding for this work?"</p>	
<p>Please select the country of your main research funder (please select carefully as in some cases this is used in fee calculation).</p> <p>as follow-up to "Financial Disclosure</p> <p>Enter a financial disclosure statement that describes the sources of funding for the work included in this submission. Review the submission guidelines for detailed requirements. View published research articles from PLOS ONE for specific examples.</p> <p>This statement is required for submission and will appear in the published article if the submission is accepted. Please make sure it is accurate.</p> <p>Funded studies</p> <p>Enter a statement with the following details:</p> <ul style="list-style-type: none"> • Initials of the authors who received each award • Grant numbers awarded to each author • The full name of each funder • URL of each funder website • Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript? <p>Did you receive funding for this work?"</p>	<p>UNITED STATES - US</p>
<p>Competing Interests</p>	<p>The authors declare that they have no competing interests, financial or non-financial,</p>

<p>Use the instructions below to enter a competing interest statement for this submission. On behalf of all authors, disclose any competing interests that could be perceived to bias this work—acknowledging all financial support and any other relevant financial or non-financial competing interests.</p> <p>This statement is required for submission and will appear in the published article if the submission is accepted. Please make sure it is accurate and that any funding sources listed in your Funding Information later in the submission form are also declared in your Financial Disclosure statement.</p> <p>View published research articles from PLOS ONE for specific examples.</p> <p>NO authors have competing interests</p> <p>Enter: <i>The authors have declared that no competing interests exist.</i></p> <p>Authors with competing interests</p> <p>Enter competing interest details beginning with this statement:</p> <p><i>I have read the journal's policy and the authors of this manuscript have the following competing interests: [insert competing interests here]</i></p> <p>* typeset</p>	<p>directly or indirectly related to the work submitted for publication.</p>
<p>Ethics Statement</p> <p>Enter an ethics statement for this submission. This statement is required if the study involved:</p> <ul style="list-style-type: none"> • Human participants • Human specimens or tissue • Vertebrate animals or cephalopods • Vertebrate embryos or tissues 	<p>N/A</p>

- Field research

Write "N/A" if the submission does not require an ethics statement.

General guidance is provided below.
Consult the [submission guidelines](#) for detailed instructions. **Make sure that all information entered here is included in the Methods section of the manuscript.**

Format for specific study types

Human Subject Research (involving human participants and/or tissue)

- Give the name of the institutional review board or ethics committee that approved the study
- Include the approval number and/or a statement indicating approval of this research
- Indicate the form of consent obtained (written/oral) or the reason that consent was not obtained (e.g. the data were analyzed anonymously)

Animal Research (involving vertebrate animals, embryos or tissues)

- Provide the name of the Institutional Animal Care and Use Committee (IACUC) or other relevant ethics board that reviewed the study protocol, and indicate whether they approved this research or granted a formal waiver of ethical approval
- Include an approval number if one was obtained
- If the study involved *non-human primates*, add *additional details* about animal welfare and steps taken to ameliorate suffering
- If anesthesia, euthanasia, or any kind of animal sacrifice is part of the study, include briefly which substances and/or methods were applied

Field Research

Include the following details if this study involves the collection of plant, animal, or other materials from a natural setting:

- Field permit number
- Name of the institution or relevant body that granted permission

Data Availability

Authors are required to make all data underlying the findings described fully available, without restriction, and from the time of publication. PLOS allows rare exceptions to address legal and ethical concerns. See the [PLOS Data Policy](#) and [FAQ](#) for detailed information.

Yes - all data are fully available without restriction

A Data Availability Statement describing where the data can be found is required at submission. Your answers to this question constitute the Data Availability Statement and **will be published in the article**, if accepted.

Important: Stating 'data available on request from the author' is not sufficient. If your data are only available upon request, select 'No' for the first question and explain your exceptional situation in the text box.

Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?

Describe where the data may be found in full sentences. If you are copying our sample text, replace any instances of XXX with the appropriate details.

- If the data are **held or will be held in a public repository**, include URLs, accession numbers or DOIs. If this information will only be available after acceptance, indicate this by ticking the box below. For example: *All XXX files are available from the XXX database (accession number(s) XXX, XXX).*
- If the data are all contained **within the manuscript and/or Supporting Information files**, enter the following:
All relevant data are within the manuscript and its Supporting Information files.
- If neither of these applies but you are able to provide **details of access elsewhere**, with or without limitations, please do so. For example:

Data cannot be shared publicly because of [XXX]. Data are available from the XXX Institutional Data Access / Ethics Committee (contact via XXX) for researchers who meet the criteria for access to confidential data.

The data underlying the results presented in the study are available from (include the name of the third party

All relevant data are within the manuscript and its Supporting Information files.

<p><i>and contact information or URL).</i></p> <ul style="list-style-type: none"> • This text is appropriate if the data are owned by a third party and authors do not have permission to share the data. <p>* typeset</p>	
Additional data availability information:	



Department of Biological Sciences,
Shelby Center for Science and Technology
301 Sparkman Drive
Huntsville, AL 35899

Jerome Baudry, Ph.D.
Interim Chair, Department of Biological Sciences
Ms. Pei-Ling Chan Endowed Chair and Professor
jerome.baudry@uah.edu

Huntsville, December 17th 2024

Dear Editor,

Please find attached a manuscript submitted for publication to *PLOS ONE journal*.

The title of the manuscript is: **"Using AI to Enhance Healthcare Resource Management and Allocation: A Focus on the Autism Community in Alabama."** and the authors are *Armin Ahmadi, Jerome Baudry (myself), Nathan Tenhundfeld, Kelly Goff, and Daniel Adamek*

This manuscript describes our original work and is not submitted anywhere else.

The manuscript is, to our knowledge, the first description of an Artificial Intelligence approach based on of Natural Language Processing (NLP) and Large Language Models (LLMs) to enhance resource management and access for individuals with autism. We are using cases and data related to Alabama, but the approach is valid regardless of the area. The research aims to address challenges in navigating complex datasets and improving the efficiency of social work practices by leveraging LLM-based chatbots and Retrieval-Augmented Generation (RAG) frameworks.

We hope that our manuscript will be found suitable for publication, and we thank you for your consideration.

Very sincerely yours,



Jerome Baudry, Ph.D.
Mrs. Pei-Ling Chan Chair and Professor.
Interim Chair, Department of Biological Sciences.

Using AI to Enhance Healthcare Resource Management and Allocation: A Focus on the Autism Community in Alabama

Armin Ahmadi*, Jerome Baudry*, Nathan Tenhundfeld\$, Kelly Goff^, Daniel Adamek&,#

*: Department of Biological Sciences, The University of Alabama in Huntsville, Huntsville, AL, 35899

\$: Department of Psychology, The University of Alabama in Huntsville, Huntsville, AL, 35800

Current address: Heterodox Academy, New York, NY 10038

^: Autism Services, Alabama Department of Mental Health, Montgomery, AL, 36130

&: Little Orange Fish, Huntsville, AL, 35806

#: Corresponding author: Daniel@littleorangefish.org

Abstract

This study investigates the potential of artificial intelligence, particularly Natural Language Processing and large-scale language models, to improve resource management and service access for individuals with autism in Alabama. The research aims to address existing challenges in navigating complex datasets and enhancing social work practices through the development and evaluation of AI-driven tools.

We designed and tested AI systems, including general language models, domain-specific chatbots powered by advanced language models, and a Retrieval-Augmented Generation framework. A standardized set of queries was used to simulate real-world scenarios encountered by social workers and families seeking autism-related resources. System performance was evaluated based on precision, recall, and response accuracy.

Results demonstrated that the Retrieval-Augmented Generation framework achieved superior performance compared to traditional methods and general models. The integration of domain-

specific information significantly enhanced the accuracy, contextual relevance, and usability of the generated responses. This improvement was particularly evident for queries requiring geographic specificity, multiple criteria, or complex contextual understanding.

The findings highlight the transformative potential of AI-driven tools in improving social work efficiency and enhancing healthcare equity. By streamlining care coordination and delivering accurate, contextually relevant information, these systems offer scalable solutions to improve access to autism-related services. Future research should focus on addressing data quality, minimizing biases, and ensuring ethical deployment to build trust and support widespread adoption of these tools.

Keywords

Natural Language Processing (NLP), Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Autism Resource Management, AI in Social Work, Healthcare Delivery Efficiency.

Introduction

Navigating fragmented healthcare systems presents significant challenges for patients, caregivers, and associated services, particularly for those managing chronic or complex conditions like autism. These inefficiencies often result in unmet medical needs and reduced patient satisfaction (Griese et al., 2020). Integrated care, which emphasizes coordination and communication among healthcare providers, has consistently demonstrated improvements in patient outcomes and satisfaction by streamlining services and reducing redundancies (Renghea et al., 2022; Kurpas et al., 2021). Holistic care platforms further enhance healthcare efficiency and sustainability, though implementing such systems is not without obstacles (Eriskin, Karatas, and Zheng, 2024).

Human social workers play a pivotal role in care coordination but often face overwhelming workloads due to high caseloads and undefined roles, which can limit their effectiveness (Yeo et al., 2022). Furthermore, only 30% of healthcare systems report robust collaborations with community-based service providers, exacerbating inefficiencies in care delivery (Kreuter et al., 2021). Issues such as human error and limited information processing further hinder case management, making it difficult to monitor outcomes and ensure accountability (Lee, 2001). While addressing social needs within clinical care demonstrates a strong return on investment, fragmented systems continue to drive high financial and human costs (Karlin, 2022).

Artificial Intelligence (AI) emerges as a transformative tool for addressing these challenges. By processing large datasets, identifying patterns, and providing evidence-based insights in real time, AI improves resource management and care coordination (Gandhi et al., 2023; Alowais et al., 2023). Automating administrative tasks like scheduling and record-keeping reduces inefficiencies and human error, with McKinsey estimating up to a 30% reduction in healthcare costs through AI-

driven operational improvements (Friesdorf, 2019). Beyond administrative efficiencies, AI also enables personalized care through data analysis, delivering tailored interventions that enhance treatment outcomes (Schork, 2019). For instance, during the COVID-19 crisis, AI-powered tools effectively guided patients to appropriate care, demonstrating the potential for streamlined coordination and improved outcomes (McRae et al., 2020). Global applications further illustrate how AI can optimize resource allocation, reduce delays, and minimize redundant treatments (Chen et al., 2019; Kitsios et al., 2023). Social workers, in particular, benefit from AI's ability to track interventions, forecast challenges, and allow for a focus on higher-level care needs (Bin et al., 2022).

However, Autism Spectrum Disorder (ASD) exemplifies a domain where these systemic challenges are particularly acute and where AI-driven solutions could provide critical support. ASD management requires a coordinated, multidisciplinary approach due to its complexity and the necessity for collaboration among diverse stakeholders (CDC, 2023). Effective care transitions, particularly from childhood to adult services, remain critical but are often hampered by personnel shortages and systemic inefficiencies (Fulceri et al., 2023). Socioeconomic and cultural factors further compound these challenges, necessitating adaptive and equitable interventions (Koum et al., 2023). The prevalence of ASD has significantly increased, with the CDC reporting that one in 36 children in the U.S. is diagnosed with ASD, a trend mirrored in Alabama with a 22% rise in prevalence since 2021 (Maenner, 2023; Shaw, 2023). Navigating autism-related services presents additional challenges due to complex eligibility requirements, insurance constraints, and a fragmented landscape of healthcare, education, and community resources. Families and social workers often face difficulties coordinating services across these sectors, which leads to delays or gaps in accessing vital services. These challenges are especially pronounced in rural and

underserved communities, where resource scarcity intersects with geographic and economic barriers (Magaña et al., 2016; Zuckerman et al., 2015). Families of autistic individuals frequently report high caregiver burden due to the constant navigation of disconnected systems, including healthcare providers, educational institutions, therapeutic services, and legal advocates (Warfield et al., 2015). These barriers result in inequities in care and missed opportunities for timely intervention.

Social workers play a critical role in supporting individuals with ASD and their families, coordinating care across healthcare, education, and community settings. AI-driven tools enhance their effectiveness by streamlining data access, enabling better intervention tracking, and providing actionable insights for individualized care (Ramadoss, 2024). By integrating diverse datasets, AI offers a holistic view of each client's needs, leading to tailored care plans and improved collaboration among service providers. Studies show a generally positive reception toward AI tools in healthcare, with about 67% of internet users expressing acceptance of AI-led services (Åhs, Mozelius, & Espvall, 2022). However, successful integration of AI requires targeted training to navigate challenges in prediction, adaptivity, and decision-making (Åhs et al., 2022).

By focusing on the autism community in Alabama, this study illustrates how AI can transform healthcare resource management. AI technologies empower social workers by streamlining care coordination, enabling more personalized and effective support, and addressing critical gaps in service delivery. Here, we focused specifically on social workers as the primary end users because the tool was developed as a proof-of-concept social worker assistant. This work leverages such use cases, workflows, and data sources, and provide a basis for the extension of this work to other stakeholders such as caregivers, physicians, and educators This research provides a scalable model for leveraging AI to improve care for underserved populations, highlighting its potential to

enhance resource allocation, reduce systemic inefficiencies, and deliver better outcomes across complex healthcare systems.

Background and Technical Foundations

Leveraging AI in Healthcare Resource Management

Artificial Intelligence (AI), particularly Natural Language Processing (NLP) and Large Language Models (LLMs), has emerged as a transformative force in healthcare resource management. These technologies efficiently process large volumes of unstructured data, such as clinical notes and patient records, which are often underutilized in traditional systems (Mah, 2022). For example, NLP applications in radiology have demonstrated 95-96% consistency in decision-making, highlighting their potential to improve outcomes (Morozov et al., 2018). NLP and LLMs extract critical insights from unstructured data, enhancing treatment planning, clinical trial recruitment, and patient engagement (Hossain et al., 2023). LLMs, particularly when integrated into Clinical Decision Support Systems (CDSS), further revolutionize care delivery by leveraging IoT devices for real-time insights (De Vito, 2024).

Traditional NLP in Healthcare: Strengths and Limitations

Traditional NLP systems rely on rule-based algorithms and machine learning models to perform tasks like medical coding and information extraction, often achieving high accuracy in specific applications. For instance, an XGBoost classifier achieved an F1 score of 0.8881 in diagnosing psychosis from psychiatric notes, and other models have reached 99% accuracy in classifying PHI-related data (Hua et al., 2024; Zhang & Jiang, 2023). However, these systems struggle with

scalability, contextual understanding, and adaptability to new or complex medical terminology. Scalability challenges, particularly in cloud-based environments, can increase latency by up to 50% (Kumar, Dubey, & Pandey, 2021). Additionally, traditional models experience performance declines of 15-20% when faced with new ICD codes without retraining (Jha, Almagro, & Tissot, 2023). Manual feature engineering is another limitation, requiring intensive human effort that can be reduced by up to 70% with deep learning models (Gupta, 2019). These systems also fail to interpret nuanced queries effectively, underperforming by 20-30% compared to transformer-based models like BERT in tasks involving complex language patterns (Dong et al., 2023; D, 2024).

Advancements with LLMs: Benefits and Challenges

LLMs, including GPT models, represent significant advancements in language processing due to their transformer-based architectures and ability to handle complex, context-rich queries with minimal retraining (García-Méndez & de Arriba-Pérez, 2024). BERT-based models, for instance, improved patient record summarization accuracy by 20-30% over traditional approaches (D, 2024). These models excel in few-shot or zero-shot learning scenarios, reducing the need for extensive fine-tuning, with GPT-3 demonstrating a 30% improvement in task accuracy using natural-language prompts (Gao, Fisch, & Chen, 2021). Furthermore, studies have shown that LLM-generated summaries are preferred over human-generated ones by up to 40% for their completeness and correctness (Van Veen et al., 2024).

Despite these advantages, LLMs require significant resources for fine-tuning and deployment. A fine-tuned GPT-3.5 model, for example, achieved an F1 score of 0.80 compared to 0.46 for an out-of-the-box version, but such training can cost millions of dollars (Stewart, Hodkiewicz, & Li, 2023; Chen et al., 2024). Biases in training data and the “black box” nature of LLMs also pose

challenges for trust and interpretability, with studies revealing significant biases in applications like physician use cases (Ayoub et al., 2024; Tao et al., 2024). Additionally, prompt engineering plays a critical role in optimizing LLM performance, particularly in specialized domains like healthcare, where well-designed prompts can achieve nearly 90% precision and recall (Polak & Morgan, 2024).

Retrieval-Augmented Generation (RAG): Enhancing LLMs

RAG combines the generative capabilities of LLMs with dynamic information retrieval, improving relevance and accuracy in healthcare applications. By incorporating real-time domain-specific data, RAG systems address the limitations of static LLMs, such as knowledge gaps and biases. For instance, BiomedRAG achieved micro-F1 scores of 81.42 and 88.83 in healthcare tasks, surpassing traditional NLP systems (Li et al., 2024). This approach also enhances transparency in decision-making, as demonstrated in its success managing chronic Hepatitis C Virus infections and reducing biases in handling sensitive data (Krešević et al., 2024; Pouplin et al., 2024).

RAG systems are highly scalable and efficient for time-sensitive healthcare scenarios. In one preoperative medicine case study, an LLM-RAG model processed 1,260 responses using 35 guidelines within 15-20 seconds, compared to the 10 minutes required by human experts. This model improved GPT-4's accuracy from 80.1% to 91.4%, exceeding human performance (Ke et al., 2024). By dynamically integrating retrieved documents into LLM outputs, RAG enhances open-domain question answering and supports better patient outcomes (Siriwardhana et al., 2023).

Methodology

This study develops an AI-driven model to improve resource management and support social workers in autism care in Alabama. By integrating diverse primary (healthcare providers), secondary (social services), and tertiary (community organizations) stakeholders, the model tries to address inefficiencies, enhances care coordination, and improves access to services. Key stakeholders were identified and categorized (Table S1), and their roles and interactions mapped to reveal gaps in existing systems to feed the model with a broad understanding of healthcare systems and public services, incorporating typical roles and responsibilities from healthcare, social services, education, legal services, and other relevant areas.

Data Collection and Integration:

To construct a comprehensive database of autism care service providers in Alabama, manual and automated web scrape identified 513 providers, and collaboration with the Alabama Department of Mental Health added another 1,102 stakeholders entities from over 350 textual documents including therapy providers, community centers, schools, nonprofit advocacy groups, and public agencies relevant to autism care in Alabama. This resulted in a final dataset of 1,615 entries. The database supported both chatbot systems (Athena and Minerva) used in this study. In this context, “providers” refers to both individual practitioners and organizations offering autism-related services. Each provider’s entry represents a unique service unit with a verifiable address and operational role. Key details such as provider names, addresses, contact information, and service descriptions were digitized using Optical Character Recognition (OCR) software (Adobe, 2024). All digitized records were then manually reviewed to validate key fields for accuracy. This included human verification of names, service descriptions, and addresses to ensure data integrity. Data originally in unstructured formats, including PDFs and handwritten notes, were converted into structured CSV databases using GPT-4 (OpenAI, 2024) for standardization and consistency.

Service Provider Tagging

A tagging system was developed to facilitate querying and data retrieval. Each service provider was assigned binary tags ('1' for offered services, '0' for non-offered) across 300 categories (example in Table 1), covering diverse needs of the autism community, such as insurance, healthcare services, therapeutic supports, and education (Hervieux and Wheatley, 2022). The binary tagging system was selected to enable scalable and interpretable matching between user queries and provider offerings. “Offered services” were defined as those that providers publicly listed or were documented as delivering in source materials. When services were ambiguous or conditional (e.g., only offered to certain age groups), we opted to create more specific tags (e.g., “early intervention” or “adult services”) rather than apply weighted labels. Some proprietary tags (e.g., “Cigna,” “United Healthcare”) were included to reflect insurance acceptance, which is a critical filtering dimension for many families. These were treated as separate but parallel attributes to broader categories such as “financial support.” Expert social workers reviewed the database to validate the accuracy and relevance of the tags.

Table 1 Sample categorization of Service Providers for Autism Care

Category	Tags
Insurance & Financial	Medicaid, AllKids, BCBS, Cigna, United, Tricare, Aetna, Medicare, Financial Supports
Healthcare Services	Pediatricians, Neurologists, Psychiatrists, Dentists, Nutritionists, Gastroenterologists, Pulmonologists, Endocrinologists, Dermatologists, Optometrists, Pediatric Neurologists
Therapeutic Services	Applied Behavior Analysis (ABA), Occupational Therapy, Physical Therapy, Speech Language Therapy, Behavioral Therapy, Art/Dance/Music Therapists, Aquatic Therapists, Equine Programs
Mental & Emotional Support	Psychologists & Counselors, Social Workers, Crisis Services, Support Groups, Substance Use Treatments, Counseling for Children, Teens, and Adults
Education & Development	Early Intervention Providers, Specialized Schools, Post-Secondary Education, Autism Evaluation, Autism Friendly Services, Educational Supports

Legal & Advocacy	Attorneys, Legal Services, Advocacy & Support, Autism Associations
Community & Social Support	Community Activities, After-school Programs, Camps, Social Skills Programs, Recreational & Community Activities, Housing & Community Living, Faith-Based Organizations
Employment & Vocational	Employment Supports, Vocational Rehabilitation, Job Training Programs
Special Needs Services	Assistive Technology, Genetic Testing, Safety/Protection, Transportation, Feeding Therapies/Interventions, Sleep Supports
Crisis & Immediate Care	Emergency Services, First Responder Resources, Inpatient Treatment Care Centers, Crisis Intervention Services
Residential & Home Care	Residential Programs, Home School, Day Programs, Housing Assistance

Socioeconomic Mapping of Resources

To identify any possible disparities in autism-related service availability, a spatial analysis was conducted by mapping service provider locations against socioeconomic factors obtained from the U.S. Census Bureau (US Census, 2024). Key factors included population density, ethnic composition, median household income, and poverty rates. Custom Python scripts utilized the Census API for data retrieval, Pandas (McKinney, 2010) for data organization, and Folium (Folium 2024) for interactive mapping. Service provider addresses were geocoded using the Nominatim geocoder (OpenStreetMap, 2024) via the GeoPy Python library and presented in a map of resource accessibility and underserved areas.

Development of the NLP-Driven Chatbot

An NLP-driven chatbot was developed to provide users with a tool to locate autism service providers based on specific needs and locations. The service provider database, initially in CSV format, was serialized into JSON files to optimize data handling and interoperability with NLP tools. (Bird et al., 2009). In house Python script were used to create comprehensive provider

profiles, encapsulating details such as name, address, services offered, and additional notes. The chatbot was built using Python 3.10 and the SpaCy v3.7 NLP library (Honnibal & Montani, 2017), utilizing the standard English NLP model (en_core_web_sm) for tasks such as tokenization, lemmatization, and named entity recognition (NER). A custom-built service_synonyms dictionary was developed to normalize user queries across heterogeneous provider terminology. For example, terms such as “speech therapy,” “speech pathologist,” and “SLP” were grouped under a unified category (“Speech Language Therapists”). The dictionary was manually curated and informed by best practices in semantic alignment and concept disambiguation (Navigli, 2009), but did not directly use external lexical resources such as WordNet or BabelNet. When exact matches were unavailable, fuzzy string matching with RapidFuzz (Gohlke, 2024) with a score cutoff of 80% was employed. Geographic information was managed by extracting locations through SpaCy’s NER capabilities and geocoding user-provided locations with Python’s geopy library (v2.4.1), specifically leveraging Nominatim geocoder (OpenStreetMap) for location identification. The geodesic distance between users and identified service providers was calculated using geopy, prioritizing providers within a 50 km radius. The system incorporated error-handling mechanisms that ensured incomplete or failed geocoding queries did not eliminate providers from recommendations but rather marked distances as "Unknown" to maintain inclusiveness. The chatbot interacts through a Flask interface (Grinberg, 2014) hosted Linux server with Ubuntu 22.04 LTS (Figure 1), prompting users to enter queries in natural language. It provides structured responses that include provider information, services offered, and contact details, enhancing accessibility and usability.

Figure 1: NLP-driven Chatbot response example, returning a filtered set of autism service providers based on user query criteria. Information shown was drawn from publicly available sources (e.g., provider websites) and manually blurred to remove identifying details. No personal health information or private individual data is presented..

Advancements with LLM-Powered Chatbots

To explore the applicability of Large Language Models (LLMs) in enhancing autism care support, we first tested widely available generative AI models (out-of-the-box LLMs), including ChatGPT (OpenAI), Google Gemini, Microsoft Copilot, and Claude (Anthropic) which are recognized for their advanced natural language understanding, ability to handle complex and nuanced queries, and general versatility across a range of domains (Minaee et al. 2024). These models were assessed using a standardized set of queries designed to simulate scenarios encountered by social workers and families seeking autism-related services in Alabama. The queries aimed to evaluate their ability to handle complex, multi-criteria requests, such as identifying local service providers based on specific needs and insurance requirements.

Based on observed strengths and limitations, we developed a domain-specific chatbot named "Minerva," built specifically upon OpenAI's GPT-4 model (GPT-4 API, version gpt-4-1106-preview). The GPT-4 model was employed with carefully engineered prompt templates (Table 2), emphasizing concise, clear instructions explicitly linked to autism service criteria and geographic specificity within Alabama. Prompt engineering utilized simple few-shot prompting techniques, explicitly listing query-response examples to guide model behavior without further fine-tuning. The structured autism provider database described earlier, serialized as JSON, provided the domain-specific context for generating responses. Iterative refinement cycles integrated structured user feedback obtained through a binary ("Yes/No") feedback loop built into the user interface to progressively improve response relevance and accuracy. This approach was designed to address

limitations observed in out-of-the-box LLMs, specifically their inability to reliably produce actionable local recommendations without supplementary domain context."

Figure 2: LLM driven Chatbot response example. Provider details are sourced from public records and have been blurred for privacy. No sensitive or user-specific information is shown.

Table 2 Criteria for GPT-4 Prompt Design

Prompt criteria	Detailed instructions
Clarity and Brevity	providing clear and concise responses, avoiding overwhelming users with excessive details, and adhering to brevity unless additional information is necessary
Data-Driven Responses	uses a structured database of autism care providers in Alabama, including service types, locations, and contact information, to accurately match user queries with service recommendations
Feedback Integration	After each interaction, Athena prompts users for feedback with a simple "Yes/No" question. This feedback is vital for continuous learning, refining responses based on user input to enhance effectiveness
Privacy / Customization	respects user privacy, handling interactions without storing personal data unless explicitly authorized, and adheres to stringent privacy standards to ensure secure communications
Guided Interactions	asks direct, relevant questions for clarification, maintaining a supportive tone to precisely identify user needs without causing frustration

Implementation of Retrieval-Augmented Generation (RAG)

To further enhance the precision and transparency of the chatbot, we implemented a Retrieval-Augmented Generation (RAG) framework consisting of two main stages: document retrieval and response generation. For the retrieval stage, we employed the OpenAI text-embedding-ada-002 model to generate dense vector embeddings for provider profiles, capturing semantic relationships beyond keyword matching (Neelakantan et al., 2022). Each profile contained comprehensive details, including provider name, address, services offered, and contact information. These embeddings were stored in a vector database, enabling efficient similarity searches. When a user query was received, it was converted into an embedding, and cosine similarity scores were

calculated between the query embedding and stored provider embeddings. Provider profiles with similarity scores exceeding a predefined threshold (0.5) were deemed relevant.

In the subsequent generation stage, relevant documents retrieved from the first stage were incorporated into prompts provided to the OpenAI GPT-3.5-turbo model. Generation parameters were explicitly set, including temperature (set to 0) and maximum tokens (set to 500), to optimize precision. The GPT-3.5-turbo assistant was explicitly instructed to generate outputs solely based on the retrieved information, enhancing transparency and minimizing hallucinations.

To further refine recommendations, geographical filtering was applied. User-provided locations were geocoded using the Nominatim geocoder from the geopy library (Gerber, 2016). Provider locations were similarly geocoded, and the geodesic distance between user and provider locations was calculated. Providers located within a specified radius (50 km) were included in the results. Missing or incomplete geocoding data were handled by marking providers' distances as "Unknown," ensuring inclusivity in recommendations. Finally, a Flask-based web application provided a user-friendly interface, allowing users to input natural language queries and receive structured responses detailing provider names, distances, services offered, and contact information.

Evaluation Framework

The performance of the developed systems, was then evaluated aimed to assess their ability to provide accurate, contextually relevant, and actionable responses for autism-related queries. We developed a standardized set of test queries reflecting real-world scenarios encountered by families and social workers seeking autism services in Alabama. These queries were categorized into six criteria:

- Simple Queries: Single-service requests with minimal criteria.
- Complex Queries: Multi-criteria requests involving multiple services or conditions.
- Geographically Specific Queries: Requests targeting specific locations.
- Synonyms and Varied Phrasing: Queries using alternative terminology or colloquial expressions.
- Ambiguous Queries: Broad or unclear requests requiring interpretation.
- Less Common Services: Queries for specialized or rare services like equine therapy.

Each category included ten representative queries (e.g., “Find an Occupational Therapist in Huntsville, AL, who accepts Medicaid and works with young adults”). These queries were developed in collaboration with the Alabama Department of Mental Health to ensure relevance and practical alignment with user needs. The output query results were categorized as Correct Matches only if they fully satisfied all constraints in the query, including service type, geographic location, insurance acceptance, and age group where applicable. Partially correct matches (e.g., matching location and service but not insurance) were not counted as correct. These results were analyzed using precision (how many of the returned results were correct), recall (how many of the correct results were returned compared to the total possible correct results), and F1 score (the harmonic mean of precision and recall, balancing both metrics) (Powers 2020). In the result section we use the term “accuracy” as a shorthand label to represent the F1 score, which provides a balanced summary of system performance across multiple query constraints. This evaluation aimed to provide a descriptive, albeit semi-quantitative, performance comparison across the tested models (Traditional NLP, Out-of-the-box LLMs, Domain-specific LLM, and RAG-based system). Formal statistical significance tests (e.g., McNemar’s test, t-tests) were not conducted. Performance was evaluated using precision, recall, and F1 scores metrics across representative

query types. Future evaluations will benefit from incorporating inferential statistical analyses to validate model performance more quantitatively.

$$Precision = \frac{Correct\ Matches}{Total\ Matches\ Returned} \quad Recall = \frac{Correct\ Matches}{Total\ Matches\ Available} \quad F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Results

Socioeconomic Mapping of Autism Services

The resulting spatial representation (Figure 3) of socioeconomic indicators allows for clear understanding of areas underserved by autism-related services. While this study's primary aim was not to explore or explain these disparities, this analysis revealed disparities in the availability of autism-related services across Alabama. Regions with poverty rates exceeding 19% and minority populations above 20% demonstrated significantly fewer service providers per capita compared to the state average. Some counties lacked autism-specific providers altogether (Figure 3). These findings underscore inequities in resource distribution, suggesting a need for targeted interventions to address underserved areas. Although a range of factors may contribute to these service gaps (Aylward, Gal-Szabo, and Taraman 2021), the maps suggest a potential need for a data-driven approach to improve resource allocation.

Figure 3: Geographic distribution of autism-related service providers in Alabama, overlaid on key socioeconomic indicators. Heatmap represents provider density. Insets show demographic context, including poverty rate, population distribution, and median income. The figure highlights service disparities that emerged during database construction and motivate the need for equitable access tools like Athena and Minerva.

Evaluation of the NLP-Driven Chatbot

The initial NLP-driven chatbot was evaluated to assess its capability to process autism-related service queries (Table 3). Traditional NLP methods implemented in our system primarily relied on syntactic processing and rule-based matching. These approaches required predefined tags and depended on exact matches or cosine similarity to link user queries to services.

Table 3 Performance Metrics of the NLP-Driven Chatbot

Sample Query	accuracy rate
Find an Occupational Therapist (OT) and Speech Therapist in Huntsville, AL who accepts Medicaid and works with 20-year-olds.	63.2 %.
Locate a Speech Therapist in Birmingham, AL specializing in early intervention and accepting private insurance	74.5 %.
Search for a Pediatric Neurologist in Montgomery, AL who offers telehealth services and accepts Medicare.	58.9 %.

This method, while effective in certain scenarios, exhibited limitations in handling the nuances of natural language, particularly when users employed varied phrasings, idiomatic expressions and synonyms, or complex query intents which are reported in research articles (Olujimi and Ade-Ibijola 2023) where the system struggled to accurately match these to the corresponding services. As a result in our case, the chatbot occasionally produced incomplete or less accurate responses, particularly in cases where the user's language deviated from the predefined tags. The chatbot's performance across previously randomly generated test queries of Table 3 shows where non-standard language was used, the accuracy of the responses dropped (Table 4).

Table 4 Challenges Faced by the NLP Chatbot in Query Processing

Sample Query	Challenges	accuracy rate
Therapist in Huntsville, AL who does OT and speech therapy for a young adult covered by Medicaid.	Uses "recommend a therapist" instead of directly specifying "Occupational Therapist (OT) and Speech Therapist," and "young adult" instead of specifying age.	57.1 %.

Where in Birmingham, AL can I find a specialist for early language intervention who works with private insurers?	Uses "specialist for early language intervention" instead of "speech therapist specializing in early intervention," and "private insurers" instead of "private insurance"	48.6 %.
Are there pediatric neurologists in Montgomery, AL who offer virtual consultations and accept Medicare?	Uses "virtual consultations" instead of "telehealth services" and "pediatric neurologists" is rephrased slightly as "Are there pediatric neurologists" instead of "Search for a pediatric neurologist,"	53.7 %.

Table 5 presents the different aspects of the NLP system, including its handling of simple, complex, and geographically specific queries, as well as its ability to process synonyms, varied phrasing, ambiguous language, and requests for less common services highlighting where the system performed well and where it struggled.

Table 5 Scenario-Based Performance Evaluation of the NLP Chatbot

Criteria	Query	Expected Service	Matched Service	Reason for Mismatch
Simple, straightforward queries	"Find a Speech Therapist in Birmingham, AL who accepts Medicaid."	Speech Therapist in Birmingham accepting Medicaid	Speech Therapist in Birmingham accepting Medicaid	Exact Match
Complex queries with multiple criteria	"I need an Occupational Therapist and a Speech Therapist in Huntsville, AL who accept Medicaid and specialize in working with young adults aged 20."	OT and Speech Therapist in Huntsville accepting Medicaid	Only Speech Therapist in Huntsville accepting Medicaid	Couldn't match both OT and Speech Therapist together
Queries with synonyms or varied phrasing	"Looking for a language specialist in Birmingham, AL who handles early intervention and takes private health insurance."	Speech Therapist specializing in early intervention	General Speech Therapist	Synonym "language specialist" not correctly linked to "Speech Therapist"
Geographically specific queries	"Are there any pediatric neurologists near downtown Montgomery, AL who offer telehealth services and accept Medicare?"	Pediatric Neurologist in Montgomery offering telehealth	Pediatric Neurologist in Montgomery	Location specificity not considered effectively by NLP system

Ambiguous or incomplete queries	"Can I find someone in Alabama who helps with autism?"	Autism specialist or center	Various unrelated services	Ambiguous phrasing, system unable to accurately determine intent
Queries with less common or rare services	"Where in Mobile, AL can I find an equine therapy program for children with autism that accepts Medicaid?"	Equine therapy program for children in Mobile accepting Medicaid	General therapy services in Mobile	Rare service not sufficiently represented in training data

Table 6 shows the distribution of the results of each query type (averaged on ten queries), alongside calculated precision, recall and F1 score. The NLP chatbot performed well on simple, single-service requests with direct phrasing while queries requiring semantic understanding, multi-criteria matching, or complex phrasing resulted in lower performance suggesting that the traditional NLP methods are effective for straightforward tasks but lack the semantic depth and contextual awareness needed for nuanced or highly specific queries.

Table 6 Precision, Recall, and F1 Scores of the NLP Chatbot

Criteria	Correct Matches	Partially Correct Matches	Incorrect Matches	Precision (%)	Recall (%)	F1 Score
Simple, straightforward queries	7	2	1	87.5	70.0	77.8
Complex queries with multiple criteria	5	3	2	71.4	50.0	58.8
Queries with synonyms or varied phrasing	4	3	3	57.1	40.0	47.0
Geographically specific queries	6	2	2	75.0	60.0	66.7
Ambiguous or incomplete queries	3	2	5	37.5	30.0	33.3
Queries with less common or rare services	2	1	7	22.2	20.0	21.0

Performance of Out-of-the-Box LLMs

Generative AI models were evaluated using the same query set (Table 7). While these models exhibited superior language comprehension and handled nuanced queries better than the NLP chatbot, they struggled with regional specificity and consistency.

Table 7 Comparison of Established LLMs for Autism Resource Queries

Sample Query	accuracy rate			
	ChatGPT (%)	Google Gemini (%)	Microsoft Copilot (%)	Claude (%)
Find an Occupational Therapist (OT) and Speech Therapist in Huntsville, AL who accepts Medicaid and works with 20-year-olds.	47.8	51.6	42.3	0.0
Locate a Speech Therapist in Birmingham, AL specializing in early intervention and accepting private insurance	43.2	37.5	33.7	2.1
Search for a Pediatric Neurologist in Montgomery, AL who offers telehealth services and accepts Medicare.	57.9	46.3	48.6	3.4

For instance, they were able to parse queries involving multiple provider types or specific insurance requirements without difficulty. Their broad knowledge base and advanced linguistic capabilities allowed them to interpret complex sentence structures and varied phrasing more effectively than the rule-based approach previously used. But despite their language processing abilities, the out-of-the-box LLMs struggled to provide region-specific or highly specialized information related to autism care in Alabama. For example, they often failed to distinguish between service providers who were available locally and those who were not. As an example, ChatGPT's failure to understand the local context led it to suggest service providers that were either not in Alabama or did not specialize in autism care, providing generic advice rather than specific, actionable information. Or Microsoft Copilot's lack of awareness of regional service providers resulted in recommendations that were irrelevant to the user's location, often suggesting national organizations or services unavailable in Alabama. In several cases, the models returned generic

responses rather than the specialized advice expected in the autism care context. For instance, in answer to a query on finding an OT and speech therapist in Huntsville, AL, Claude gives general suggestions like contacting Alabama Medicaid, doing an online search, or contacting local universities for recommendations (Figure 4). Also, the outputs might not be consistent, if a query is being asked multiple times, each time a different answer might show up. For instance Table 8, shows a query that was asked (verbatim) five times from Gemini LLM and each time a different response was generated by LLM.

Figure 4: Example of LLM Fails to Provide Relevant Information

Table 8 Inconsistencies in LLM Responses to Identical Queries

Query	Summary of Response
Find two Occupational Therapist (OT) and Speech Therapist in Huntsville, AL who accepts Medicaid and works with 20-year-olds.	1- General suggestions on finding therapists, including contacting Medicaid providers, local hospitals, universities, and online directories like Beaming Health.
	2- Provides names of specific therapy centers (Child'sPlay, Therapy Achievements, Pediatric Outpatient Center) that accept Medicaid and suggests further searches.
	3- Suggests using Beaming Health, contacting Medicaid providers for in-network therapists, and checking with local hospitals or rehab centers for therapists.
	4- Lists specific therapy centers (Child'sPlay, More Than Words, Pediatric Therapy Link) that accept Medicaid and work with 20-year-olds.
	5- Provides a list of therapy centers with addresses on Google Maps but recommends contacting them directly to confirm Medicaid acceptance for 20-year-olds.

Interestingly, in terms of region-specific accuracy, the domain-focused NLP chatbot (Athena) we developed using traditional NLP methods outperformed the LLMs. This indicates that while LLMs offer considerable potential for general tasks, their utility in highly specialized, context-heavy domains like autism care may be limited without additional context or training on domain-specific data. The LLMs often returned less accurate results when asked for specific services that required

deep local knowledge, especially when compared to our earlier system, which was trained on curated, location-specific datasets.

Domain-Specific LLM-based Chatbot Testing

The customized GPT-4-based chatbot, Athena, demonstrated substantial improvements over both traditional NLP and out-of-the-box LLMs (Table 9, 10).

Table 9 Performance Comparison: NLP vs. LLMs

Query	NLP System Result	LLM Result	Observation
"Find a provider for ABA therapy in Birmingham"	Partial match; misses some relevant providers	Comprehensive match with all relevant providers	NLP struggled with synonyms, LLM recognized intent
"Looking for speech therapy near Montgomery"	No relevant results	Accurate and relevant results	NLP failed to match non-exact phrasing, LLM succeeded
"Are there providers for OT in Huntsville?"	Accurate but limited	Accurate and more comprehensive	NLP provided limited results, LLM covered more options
"Who accepts Medicaid for ABA in Mobile?"	Incorrect match (focused on location only)	Correct match with relevant insurance filter	NLP failed to account for insurance filter, LLM did well

Table 10 Accuracy Metrics for Different Models

Sample Query	NLP Accuracy (%)	Out-of-the-Box LLM Accuracy (%)	Domain-Specific LLM Accuracy (%)
Find an Occupational Therapist and Speech Therapist in Huntsville, AL who accepts Medicaid.	66.7	44.9	91.2
Locate a Speech Therapist in Birmingham specializing in early intervention.	72.1	38.4	93.7

Search for a Pediatric Neurologist in Montgomery offering telehealth services.	61.3	56.5	89.1
--	------	------	------

More granular comparison of these models by comparing F1 scores across different scenarios also shows domain specific LLM-based chatbot outperforms the NLP based (Table 11).

Table 11 Scenario-Specific Performance Across NLP and Athena

Criteria	Traditional NLP	Domain-Specific LLM
Simple Queries	79.4	94.6
Complex Queries	51.2	89.5
Queries with Varied Phrasing	42.7	86.3
Geographically Specific Queries	58.9	87.4
Queries for Less Common Services	23.5	81.7

Although the numbers reflect qualitative assessments rather than precise, scientifically measured values, these results indicate that while traditional NLP systems generally struggled with varied phrasing and complex queries, out-of-the-box LLMs showed improvement but still lacked the precision and recall necessary for highly specialized autism care services. In contrast, using domain-specific tuning of GPT-4 consistently outperformed both methods which highlights a clear improvement in relevance and response accuracy, even if the evaluations are based on observed trends rather than rigid statistical metrics.

Performance of RAG-based chatbot

While the GPT-4-powered Athena chatbot demonstrated significant improvements in handling nuanced queries and domain-specific autism care services, it relied on a custom version of OpenAI's ChatGPT with limited control over model parameters such as temperature, token limits, and conversation context (OpenAI, 2023). To enhance functionality and precision, we

implemented a RAG approach to address the shortcomings of traditional NLP and out-of-the-box LLMs by integrating large language models with a dynamic retrieval mechanism from our curated dataset of autism service providers (Gao et al., 2023). This provides greater control over model parameters, enabling fine-tuning of responses to meet user expectations. By integrating real-time retrieval from our curated dataset, the chatbot provides contextually relevant and up-to-date information. Responses are based on specific retrieved documents, allowing us to trace the source of information, which is crucial in healthcare applications for building user trust. Comparative performance observation of these models (Table 12) shows RAG enhanced semantic matching, enabling precise recommendations even for complex, multi-criteria queries. In addition to these qualitative comparisons, RAG offers superior control over system performance metrics.

Table 12 Comparative Observations: RAG-Based vs. GPT-Based Athena

Sample Query	Athena GPT-4 Result	RAG-Based Athena Result	Observation
Find an Occupational Therapist and Speech Therapist in Huntsville, AL who accepts Medicaid.	Partial match	Accurate and specific match	RAG retrieved more precise data on providers and services.
Locate a Speech Therapist in Birmingham specializing in early intervention.	General match	Specific match	RAG incorporated specific local data not available to GPT-4 alone.
Search for a Pediatric Neurologist in Montgomery offering telehealth services.	Incomplete match	Full match	RAG handled multiple criteria more effectively.

While GPT-4-based system performed well in complex queries, RAG's structured retrieval mechanism resulted in even higher accuracy, precision, and recall (Table 13).

Table 13 - Precision and Recall of RAG-Based vs. GPT-Based

Criteria	GPT-4-Based Precision (%)	RAG-Based Precision (%)	GPT-4-Based Recall (%)	RAG-Based Recall (%)
Simple Queries	91.3	96.2	87.5	97.2
Complex Queries	87.1	94.3	83.3	95.5

Geographically Specific Queries	84.6	92.7	79.2	93.1
---------------------------------	------	------	------	------

In comparison to the custom GPT-4-based system, which was reliant on predefined knowledge, this RAG-based architecture provides Enhanced Control, Improved Accuracy, Transparency, and Explainability. With RAG, we can adjust critical generation parameters such as temperature (for controlling the randomness of output), token limit (for managing the length of responses), and the number of conversation turns kept in context (Surla 2023). The ability to incorporate real-time retrieval from our own curated dataset means that the RAG-based model is not limited by the pre-trained knowledge of an LLM. It can dynamically pull the most relevant information from a vast repository of autism service data, ensuring that responses are not only contextually relevant but also accurate and up-to-date. With RAG, every response is based on retrieved documents, allowing us to trace the origin of the information. This transparency is a key differentiator when compared to Athena’s GPT-4-based responses, where the sources of information were more opaque. For sensitive healthcare domains like autism services, the ability to verify and explain responses is crucial for user trust.

Conclusion and Discussion

This study systematically explored the potential of AI-driven tools for improving autism-related healthcare resource management in Alabama. By developing and evaluating several chatbot models—including traditional NLP methods, out-of-the-box large language models (LLMs), a customized domain-specific GPT-4 chatbot, and a Retrieval-Augmented Generation (RAG) chatbot—we were able to clearly demonstrate the feasibility and comparative strengths of different AI approaches in this specialized social work context.

Among the systems evaluated, traditional NLP-based methods showed effectiveness for simple, clearly structured queries, achieving up to 87.5% precision and 70% recall. However, these models were limited by their reliance on exact keyword matching and struggled to handle complex, multi-criteria queries and linguistic variations. Out-of-the-box LLMs, such as ChatGPT, Google Gemini, Microsoft Copilot, and Claude, exhibited strong capabilities in understanding nuanced, natural language queries but faced significant limitations in terms of geographic specificity, consistently returning generic and less actionable recommendations due to the absence of localized, domain-specific training. In contrast, our domain-specific GPT-4 chatbot (Minerva) substantially improved upon these approaches by effectively leveraging structured local datasets, clearly highlighting the importance of domain-specific fine-tuning. This model achieved significantly higher accuracy (approximately 85–95%) across diverse query types, demonstrating its utility in real-world scenarios faced by social workers and caregivers. Finally, the Retrieval-Augmented Generation (RAG) chatbot emerged as the strongest model, integrating real-time retrieval of relevant provider information with generative AI capabilities. It provided enhanced accuracy (approximately 90–96% precision and recall), improved transparency, and higher explainability of recommendations. These characteristics are particularly valuable for sensitive healthcare applications where trust and interpretability are critical.

The primary limitation of this study stems from its proof-of-concept nature, relying on exploratory evaluations rather than extensive computational benchmarks or statistical validation. Additionally, the practical deployment of these AI tools faces inherent challenges, including the need for continuous updating of service provider information, potential computational costs associated with maintaining vector embeddings, and ethical considerations surrounding algorithmic biases and data privacy. Our results clearly underscore the promise and practicality of tailored AI tools—

particularly domain-specific GPT-4 and RAG-based systems—for enhancing social work practices related to autism care. However, successful real-world implementation will require thoughtful integration strategies, regular data management practices, and attention to ethical and operational considerations. Observed marked disparities in the geographic and socioeconomic distribution of autism-related services across Alabama were observed (Figure 3). These disparities, while not the central focus of this study, illustrate the need for intelligent assistive tools that can help families and social workers identify appropriate resources despite systemic access gaps, when identified.

While the findings of this study demonstrate the potential of AI-driven tools in autism care, several important ethical and regulatory considerations must be addressed to ensure responsible deployment and real-world impact. One primary concern is algorithmic bias. AI models, particularly large language models (LLMs), risk reinforcing existing disparities if trained on data that inadequately represents marginalized populations. This is particularly relevant in autism services, where racial and socioeconomic disparities are well-documented (Aylward et al., 2021). LLMs have also shown susceptibility to inheriting biases present in their training data, potentially impacting equitable service recommendations (Ayoub et al., 2024). Data privacy and security are equally critical, especially when handling sensitive healthcare information. Compliance with regulations like the Health Insurance Portability and Accountability Act (HIPAA) is essential to protect personally identifiable information. Although this study utilized mostly publicly and non-sensitive available service provider data, future implementations involving patient data must incorporate privacy-preserving techniques such as de-identification and encryption (Ghadban et al., 2023). From a regulatory standpoint, ensuring model transparency, explainability, and human-in-the-loop review processes will be necessary for ethical compliance and user trust (Gallegos et

al., 2024). Moreover, continuous auditing and bias assessment should be embedded into system maintenance to mitigate risks of harm or unequal service delivery.

While as a pilot, this study demonstrates the potential framework for an AI-driven system to support autism care resource management, practical implementation would require further development to ensure adaptability to changes in available services and usability by intended stakeholders. Future iterations should incorporate mechanisms for periodic database updates to reflect changes in provider availability, service offerings, and insurance coverage. This could involve a combination of automated data collection, manual verification, and partnerships with state agencies to maintain accuracy over time (Eriskin, Karatas, and Zheng, 2024). The system is proposed to be designed to support accessibility through a web-based interface capable of processing natural language queries. Social workers, families, and caregivers represent primary user groups. Targeted training materials, including user guides and interactive onboarding, would be necessary to ensure effective engagement, particularly for non-technical users. Prior studies emphasize that such training is essential to maximize the utility of AI tools in social work and healthcare environments (Åhs, Mozelius, and Espvall, 2022). While this study emphasized natural language interfaces to assess the feasibility of LLM-powered semantic search and to be used by users with high digital literacy. However, traditional UI designs (e.g., dropdowns, filters) may offer complementary benefits in terms of simplicity or accessibility for some user groups. Future work could compare chatbot-driven interfaces with conventional UI approaches to assess their relative performance, user preferences, and usability across diverse populations. Finally, incorporating user feedback mechanisms can enable continuous refinement of recommendations and system performance. Feedback loops are increasingly recognized as a best practice in AI system design to enhance accuracy, relevance, and user trust (Pagan et al., 2023).

Together, these considerations emphasize that while AI-driven tools hold significant promise for improving autism care resource navigation, their real-world deployment will require careful attention to ethical, regulatory, and practical implementation factors. This proof-of-concept study provides a foundation for future research and development aimed at operationalizing such tools within social work and healthcare systems.

Supplementary data:

Table S1: Stakeholder Roles and Service Mapping

References

Adobe. OCR PDF | How Tos | PDF Services API | Adobe PDF Services [Internet]. 2024 [cited 2024 Jun 17]. Available from: <https://developer.adobe.com/document-services/docs/overview/pdf-services-api/howtos/ocr-pdf/>

Åhs F, Mozelius P, Espvall M. Preparing psychologists and social workers for the daily use of AI. *Int Conf AI Res*. 2022;4(1):1–5. doi:10.34190/icaire.4.1.771

Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ*. 2023;23(1):689. doi:10.1186/s12909-023-04698-z

Aylward BS, Gal-Szabo DE, Taraman S. Racial, ethnic, and sociodemographic disparities in diagnosis of children with autism spectrum disorder. *J Dev Behav Pediatr*. 2021;42(8):682–9. doi:10.1097/DBP.0000000000000996

Ayoub NF, Balakrishnan K, Ayoub MS, Barrett TF, David AP, Gray ST. Inherent bias in large language models: a random sampling analysis. *Mayo Clin Proc Digit Health*. 2024;2(2):186–91. doi:10.1016/j.mcpdig.2024.03.003

Berge GT, Granmo O-C, Tveit TO, Ruthjersen AL, Sharma J. Combining unsupervised, supervised and rule-based learning: the case of detecting patient allergies in electronic health records. *BMC Med Inform Decis Mak*. 2023;23:188. doi:10.1186/s12911-023-02271-8

- Bishop-Fitzpatrick L, Dababnah S, Baker-Ericzén MJ, Smith MJ, Magaña SM. Autism spectrum disorder and the science of social work: a grand challenge for social work research. *Soc Work Ment Health*. 2019;17(1):73–92. doi:10.1080/15332985.2018.1509411
- Shaw, K. A., Williams, S., Patrick, M. E., et al. (2025). Prevalence and Early Identification of Autism Spectrum Disorder Among Children Aged 4 and 8 Years — Autism and Developmental Disabilities Monitoring Network, 16 Sites, United States, 2022. *MMWR Surveillance Summaries*, 74(SS-2), 1–22. <https://doi.org/10.15585/mmwr.ss7402a1>
- Cheak-Zamora NC, Teti M, First J. 'Transitions are scary for our kids, and they're scary for us': Family member and youth perspectives on the challenges of transitioning to adulthood with autism. *J Appl Res Intellect Disabil*. 2015 Nov;28(6):548–60. doi:10.1111/jar.12150
- Chen, Q., Hu, Y., Peng, X. *et al*. Benchmarking large language models for biomedical natural language processing applications and recommendations. *Nat Commun* **16**, 3280 (2025). <https://doi.org/10.1038/s41467-025-56989-2>
- Chen X, Zhao L, Liang H, Lai KK. Matching patients and healthcare service providers: a novel two-stage method based on knowledge rules and OWA-NSGA-II algorithm. *J Comb Optim*. 2019;37(1):221–47. doi:10.1007/s10878-017-0221-2
- Christoph Gohlke. RapidFuzz: Fuzzy string matching in Python. [Internet]. 2024 [cited 2024 Jul 30]. Available from: <https://github.com/maxbachmann/RapidFuzz>
- Jagtap DS. Cross-media topic modeling: a comparative analysis of NMF, LDA and CoRex combined with BERT on Quora and Twitter. In: 2024 International Conference on

Computing and Data Science (ICCDS). 2024. p. 1–6.

doi:10.1109/ICCDS60734.2024.10560395

De Vito G. Assessing healthcare software built using IoT and LLM technologies. In: Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering. New York, NY: Association for Computing Machinery; 2024. p. 476–81.

doi:10.1145/3661167.3661202

Dong H, Suárez-Paniagua V, Zhang H, Wang M, Casey A, Davidson E, et al. Ontology-driven and weakly supervised rare disease identification from clinical notes. BMC Med Inform Decis Mak. 2023;23(1):86. doi:10.1186/s12911-023-02181-9

Eriskin L, Karatas M, Zheng Y-J. A robust multi-objective model for healthcare resource management and location planning during pandemics. Ann Oper Res. 2024;335(3):1471–518. doi:10.1007/s10479-022-04760-x

Friesdorf M. Automation at scale: the benefits for payers [Internet]. McKinsey. 2019 [cited 2024 Jun 17]. Available from: <https://www.mckinsey.com/industries/healthcare/our-insights/automation-at-scale-the-benefits-for-payers>

Fulceri F, Gila L, Caruso A, Micai M, Romano G, Scattoni ML. Building bricks of integrated care pathway for autism spectrum disorder: a systematic review. Int J Mol Sci. 2023;24(7):6222. doi:10.3390/ijms24076222

Folium: Python Data. Leaflet.js Maps [Internet]. 2024 [cited 2024 Jul 30]. Available from: <https://python-visualization.github.io/folium/>

Isabel O. Gallegos, Ryan A. Rossi, Joe Barrow, Md Mehrab Tanjim, Sungchul Kim, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, Nesreen K. Ahmed; Bias and Fairness in Large Language Models: A Survey. *Computational Linguistics* 2024; 50 (3): 1097–1179. doi: https://doi.org/10.1162/coli_a_00524

Gandhi TK, Classen D, Sinsky CA, Rhew DC, Vande Garde N, Roberts A, et al. How can artificial intelligence decrease cognitive and work burden for front line practitioners? *JAMIA Open*. 2023;6(3):ooad079. doi:10.1093/jamiaopen/ooad079

Gao T, Fisch A, Chen D. Making pre-trained language models better few-shot learners. In: Zong C, Xia F, Li W, Navigli R, editors. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics; 2021. p. 3816–30. doi:10.18653/v1/2021.acl-long.295

Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, et al. Retrieval-augmented generation for large language models: a survey. *arXiv [Preprint]*. 2023 Dec 18. Available from: <https://arxiv.org/abs/2312.10997v5>

García-Méndez S, de Arriba-Pérez F. Large language models and healthcare alliance: potential and challenges of two representative use cases. *Ann Biomed Eng*. 2024;52(8):1928–31. doi:10.1007/s10439-024-03454-8

Ghadban YA, Lu H, Adavi U, Sharma A, Gara S, Das N, et al. Transforming healthcare education: harnessing large language models for frontline health worker capacity building

using retrieval-augmented generation. medRxiv [Preprint]. 2023.

doi:10.1101/2023.12.15.23300009

Griese L, Berens E-M, Nowak P, Pelikan JM, Schaeffer D. Challenges in navigating the health care system: development of an instrument measuring navigation health literacy. *Int J Environ Res Public Health*. 2020;17(16):5731. doi:10.3390/ijerph17165731

Grinberg M. *Flask Web Development*. 1st ed. Sebastopol, CA: O'Reilly Media; 2014. ISBN: 978-1-449-37262-0

Gupta P. Neural information extraction from natural language text [dissertation]. Munich: Ludwig-Maximilians-Universität München; 2019. Available from: <https://edoc.ub.uni-muenchen.de/24953/>

Herd P, Moynihan D. Health care administrative burdens: centering patient experiences. *Health Serv Res*. 2021;56(5):751–4. doi:10.1111/1475-6773.13858

Hervieux S, Wheatley A. The rise of AI: implications and applications of artificial intelligence in academic libraries. *ACRL Publications in Librarianship*; No. 78. Chicago, IL: Association of College and Research Libraries; 2022.

Honnibal, Montani. spaCy 2: natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing [Internet]. Sentometrics Research; 2017 [cited 2024 Jun 17]. Available from: <https://sentometrics-research.com/publication/72/>

Hossain E, Rana R, Higgins N, Soar J, Barua PD, Pisani AR, et al. Natural language processing in electronic health records in relation to healthcare decision-making: a systematic review. *Comput Biol Med.* 2023;155:106649. doi:10.1016/j.compbimed.2023.106649

Hua Y, Blackley S, Shinn A, Skinner J, Moran L, Zhou L. Identifying psychosis episodes in psychiatric admission notes via rule-based methods, machine learning, and pre-trained language models [Preprint]. 2024. doi:10.21203/rs.3.rs-4126574/v1

Jha J, Almagro M, Tissot H. Designing NLP applications to support ICD coding: an impact analysis and guidelines to enhance baseline performance when processing patient discharge notes. *J Digit Health.* 2023;63–81. doi:10.55976/jdh.22023119463-81

Karlin J. Sam's story: the financial and human costs of disjointed logics of care [Internet]. PMC. 2022 [cited 2024 Jun 17]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8786420/>

Ke Y, Jin L, Elangovan K, Abdullah HR, Liu N, Sia ATH, et al. Development and testing of retrieval augmented generation in large language models -- a case study report. *arXiv* [Preprint]. 2024. doi:10.48550/arXiv.2402.01733

Kim E, Jeong Y, Choi MS. MediBioDeBERTa: biomedical language model with continuous learning and intermediate fine-tuning. *IEEE Access.* 2023;11:141036–44. doi:10.1109/ACCESS.2023.3341612

Kitsios F, Kamariotou M, Syngelakis AI, Talias MA. Recent advances of artificial intelligence in healthcare: a systematic literature review. *Appl Sci.* 2023;13(13):7479. doi:10.3390/app13137479

Koum DCK, Enyama D, Kojom Foko LP, Noukeu D, Nguedia B, Mbono R, et al. Diagnosis and treatment challenges of autism spectrum disorder at a reference hospital in Douala, Cameroon. *BMC Pediatr.* 2023;23(1):459. doi:10.1186/s12887-023-04242-4

Krešević S, Giuffrè M, Ajčević M, Accardo A, Crocè LS, Shung DL. Optimization of hepatological clinical guidelines interpretation by large language models: a retrieval augmented generation-based framework. *NPJ Digit Med.* 2024;7(1):1–9. doi:10.1038/s41746-024-01091-y

Kreuter MW, Thompson T, McQueen A, Garg R. Addressing social needs in health care settings: evidence, challenges, and opportunities for public health. *Annu Rev Public Health.* 2021;42:329–44. doi:10.1146/annurev-publhealth-090419-102204

Kumar M, Dubey K, Pandey R. Evolution of emerging computing paradigm cloud to fog: applications, limitations and research challenges. In: 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). 2021. p. 257–61. doi:10.1109/Confluence51648.2021.9377050

Kurpas D, Stefanicka-Wojtas D, Shpakou A, Halata D, Mohos A, Skarbaliene A, et al. The advantages and disadvantages of integrated care implementation in Central and Eastern Europe – perspective from 9 CEE countries. *Int J Integr Care.* 2021;21(4):14. doi:10.5334/ijic.5632

Lee B. Challenges facing case management in a changing social service environment. *Care Manag J.* 2001 Fall;3(1):20-4. PMID: 11974573.

Li M, Kilicoglu H, Xu H, Zhang R. BiomedRAG: a retrieval augmented large language model for biomedicine. *Journal of Biomedical Informatics*, 2025.

<https://doi.org/10.1016/j.jbi.2024.104769>

Maenner MJ. Prevalence and characteristics of autism spectrum disorder among children aged 8 years — Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. *MMWR Surveill Summ*. 2023;72. doi:10.15585/mmwr.ss7202a1

Mah PM. Analysis of artificial intelligence and natural language processing significance as expert systems support for e-health using pre-train deep learning models. 2022.

doi:10.56578/ataiml010201

Magaña S, Parish SL, Rose RA, Timberlake M, Swaine JG. Racial and ethnic disparities in quality of health care among children with autism and other developmental disabilities.

Intellect Dev Disabil. 2012 Aug;50(4):287–99. doi:10.1352/1934-9556-50.4.287McKinney

W. Data Structures for Statistical Computing in Python. In: van der Walt S, Millman J, editors. *Proceedings of the 9th Python in Science Conference*; 2010. p. 51–6.

doi:10.25080/Majora-92bf1922-00a

McRae MP, Simmons GW, Christodoulides NJ, Lu Z, Kang SK, Fenyo D, et al. Clinical decision support tool and rapid point-of-care platform for determining disease severity in patients with COVID-19. *Lab Chip*. 2020;20(12):2075–85. doi:10.1039/d0lc00373e

Minaee S, Mikolov T, Nikzad N, Chenaghlu M, Socher R, Amatriain X, Gao J. Large language models: a survey. *arXiv [Preprint]*. 2024. doi:10.48550/arXiv.2402.06196

Morozov SP, Vladzimirskiy AV, Gombolevskiy VA, Kuz'mina ES, Ledikhova NV.

Artificial intelligence: natural language processing for peer-review in radiology. *J Radiol Nucl Med.* 2018;99(5):253–8. doi:10.20862/0042-4676-2018-99-5-253-258

Olujimi PA, Ade-Ibijola A. NLP techniques for automating responses to customer queries: a systematic review. *Discov Artif Intell.* 2023;3(1):20. doi:10.1007/s44163-023-00065-5

OpenAI. Creating a GPT. OpenAI Help Center. 2024 [cited 2024 Jun 17]. Available from: <https://help.openai.com/en/articles/8554397-creating-a-gpt>

OpenAI. GPT-4 Technical Report. OpenAI Research. 2023 Mar 27 [cited 2024 Jun 17]. Available from: <https://cdn.openai.com/papers/gpt-4.pdf>

OpenStreetMap Foundation. Nominatim: Geocoding API for OpenStreetMap [Internet]. 2024 [cited 2024 Jul 30]. Available from: <https://nominatim.openstreetmap.org/>

Pagan N, Baumann J, Elokda E, De Pasquale G, Bolognani S, Hannák A. A classification of feedback loops and their relation to biases in automated decision-making systems. In: *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23)*; 2023. doi:10.1145/3617694.3623227

Polak MP, Morgan D. Extracting accurate materials data from research papers with conversational language models and prompt engineering. *Nat Commun.* 2024;15(1):1569. doi:10.1038/s41467-024-45914-8

Pouplin T, Sun H, Holt S, van der Schaar M. Retrieval augmented thought process for private data handling in healthcare. *arXiv [Preprint]*. 2024. doi:10.48550/arXiv.2402.07812

Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *International Journal of Machine Learning Technologies*, 2229-3981, 2011. doi: 10.9735/2229-3981

Radu M, Moldovan R, Băban A. Families with complex needs: an inside perspective from young people, their carers, and healthcare providers. *J Community Genet*. 2022;13(3):293–302. doi:10.1007/s12687-022-00586-z

Ramadoss G. Social determinants of health: integrating analytics and technology in healthcare systems, *IJSR*, call for papers, online journal. *Int J Sci Res (IJSR)*. 2024. Available from: <https://www.ijsr.net/>

Renghea A, Cuevas-Budhart MA, Yébenes-Revuelto H, Gómez Del Pulgar M, Iglesias-López MT. ‘Comprehensive care’ concept in nursing: systematic review. *Investig Educ Enferm*. 2022;40(3):e05. doi:10.17533/udea.iee.v40n3e05

Sadeghi-Tehran P, Angelov P, Virlet N, Hawkesford MJ. Scalable database indexing and fast image retrieval based on deep learning and hierarchically nested structure applied to remote sensing and plant biology. *J Imaging*. 2019;5(3):33. doi:10.3390/jimaging5030033

Schork NJ. Artificial intelligence and personalized medicine. *Cancer Treat Res*. 2019;178:265–83. doi:10.1007/978-3-030-16391-4_11

Shah WS, Elkhwesky Z, Jasim KM, Elkhwesky EFY, Elkhwesky FFY. Artificial intelligence in healthcare services: past, present and future research directions. *Rev Manag Sci*. 2024;18(3):941–63. doi:10.1007/s11846-023-00699-w

Shaw KA. Early identification of autism spectrum disorder among children aged 4 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. *MMWR Surveill Summ.* 2023;72. doi:10.15585/mmwr.ss7201a1

Siriwardhana S, Weerasekera R, Wen E, Kaluarachchi T, Rana R, Nanayakkara S. Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering. *Trans Assoc Comput Linguist.* 2023;11:1–17. doi:10.1162/tacl_a_00530

Stewart M, Hodkiewicz M, Li S. Large language models for failure mode classification: an investigation. *arXiv [Preprint]*. 2023. doi:10.48550/arXiv.2309.08181

Surla A. How to Get Better Outputs from Your Large Language Model. NVIDIA Developer Technical Blog. Jun 14, 2023 [cited 2024 Jun 17]. Available from: <https://developer.nvidia.com/blog/how-to-get-better-outputs-from-your-large-language-model/>

Tao S, Yao L, Ding H, Xie Y, Cao Q, Sun F, et al. When to trust LLMs: aligning confidence with response quality. *arXiv [Preprint]*. 2024. doi:10.48550/arXiv.2404.17287

US Census Bureau. Census Bureau data [Internet]. 2024 [cited 2024 Jun 17]. Available from: <https://data.census.gov/>

Uysal AK, Gunal S. The impact of preprocessing on text classification. *Inf Process Manag.* 2014;50(1):104–12. doi:10.1016/j.ipm.2013.08.006

Van Veen D, Van Uden C, Blankemeier L, Delbrouck J-B, Aali A, Bluethgen C, et al.

Adapted large language models can outperform medical experts in clinical text summarization. *Nat Med*. 2024;30(4):1134–42. doi:10.1038/s41591-024-02855-5

Warfield ME, Crossman MK, Delahaye J, Der Weerd E, Kuhlthau KA. Physician perspectives on providing primary medical care to adults with autism spectrum disorders (ASD). *J Autism Dev Disord*. 2015 Jul;45(7):2209–17. doi:10.1007/s10803-015-2386-9

Wu H, Lu X, Wang H. The application of artificial intelligence in health care resource allocation before and during the COVID-19 pandemic: scoping review. *JMIR AI*. 2023;2(1):e38397. doi:10.2196/38397

Yeo GTS, Kannan P, Lee ES, Smith HE. Community case managers' challenges collaborating with primary care when managing complex patients in the community: a qualitative study in Singapore. *Health Soc Care Community*. 2022;30(4):1568–77. doi:10.1111/hsc.13489

Zhang K, Jiang X. Sensitive data detection with high-throughput machine learning models in electrical health records. *arXiv [Preprint]*. 2023. doi:10.48550/arXiv.2305.03169

Zhang R, Xing Y, Wang J, Shang X, Zhu X. A novel multiattribute decision-making method based on point–Choquet aggregation operators and its application in supporting the hierarchical medical treatment system in China. *Int J Environ Res Public Health*. 2018;15(8):1718. doi:10.3390/ijerph15081718

Zuchowski I, McLennan S. A systematic review of social work in general practice: opportunities and challenges. *J Evid Based Soc Work*. 2023;20(5):686–726.
doi:10.1080/26408066.2023.2202665

Zuckerman KE, Lindly OJ, Sinche BK. Parental concerns, provider response, and timeliness of autism spectrum disorder diagnosis. *J Pediatr*. 2015 Jun;166(6):1431–9.e1.
doi:10.1016/j.jpeds.2015.03.007

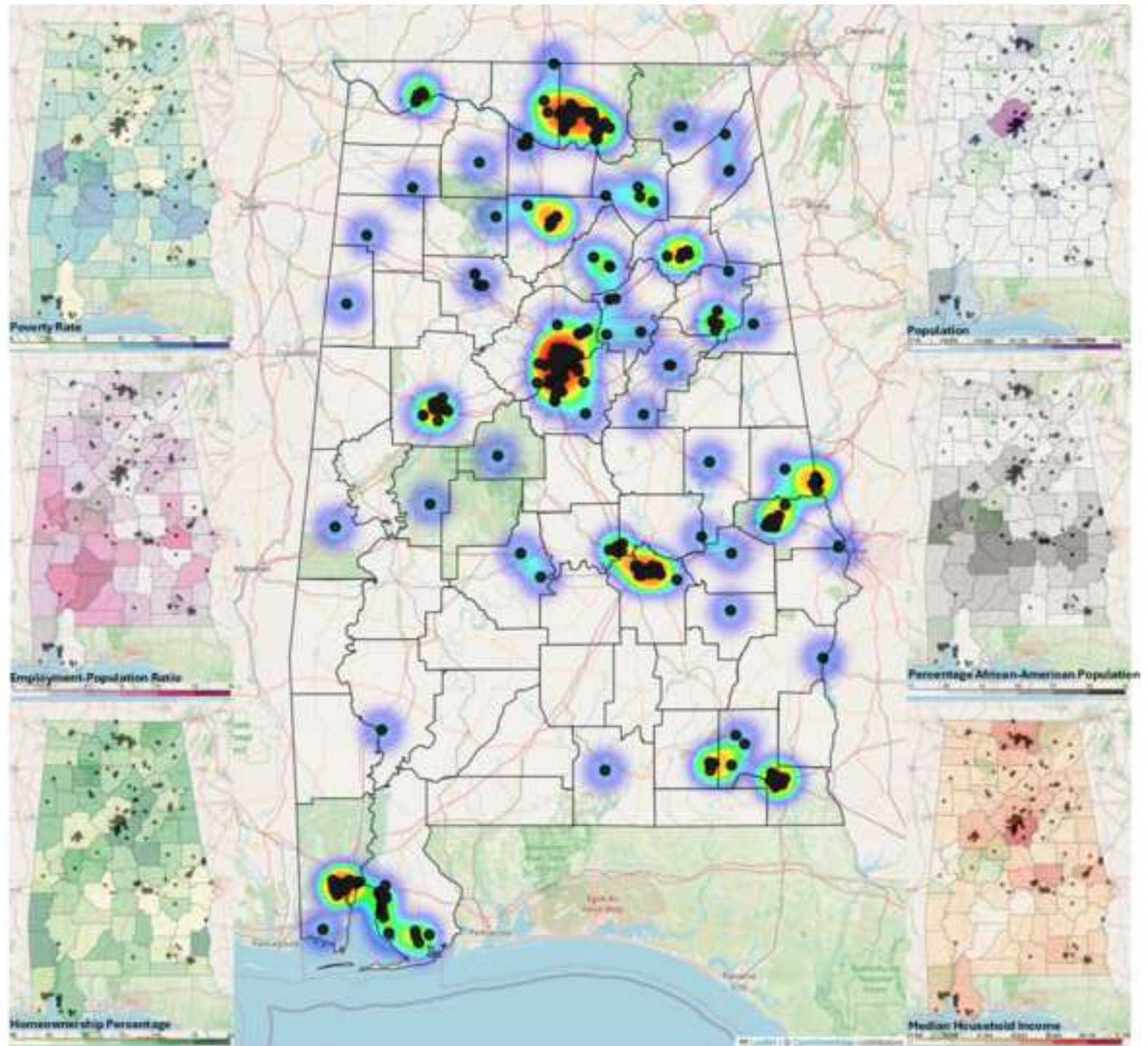
Supplementary data:

Table S1: Stakeholder Roles and Service Mapping

Organization Type	Primary Function	Primary Provider	Secondary Provider(s)	Primary Beneficiary
Art Therapist	Art-based skill development	Art therapist	Family, Various therapists	Child with autism
Autism Advocacy Organizations	Advocacy and policy making	Advocacy workers	Volunteers, Lawyers	Autism community
Autism Research Centers	Research and development	Researchers	Technicians, Volunteers	Autism community
Autism Support Group	Peer support and resources	Group coordinator	Families, Therapists	Child with autism and family
Behavioral Therapist	Behavioral assessment and intervention	Behavioral therapist	Family, Various therapists	Child with autism
Clinical Psychologist	Assessment and treatment	Clinical psychologist	Family, Therapists	Child with autism
Early Intervention Services	Early childhood intervention	Early intervention specialists	Family, Various therapists	Child with autism
Family	Support and care	Parents or siblings	None	Child with autism
Government Health Agencies	Policy making and funding	Government officials	Various health professionals	Autism community
Government Health Department	Health policy making and implementation	Government health officials	Various health professionals	Autism community
Insurance Companies	Coverage of treatments and services	Insurance agents	None	Family with autistic child

Insurance Companies	Insurance policy making and implementation	Insurance policy makers	None	Family with autistic child
Legal Services	Legal support and advice	Lawyers	None	Family with autistic child
Local Community Centers	Community support and resources	Community center staff	Volunteers, Various therapists	Autism community
Local Government	Local policy making and implementation	Local government officials	Local health and education professionals	Autism community
Music Therapist	Music-based skill development	Music therapist	Family, Various therapists	Child with autism
National Government	National policy making and implementation	National government officials	Various health and education professionals	Autism community
Neurologist	Medical assessment and treatment	Neurologist	Family, Therapists	Child with autism
Nutritionists	Diet and nutrition advice	Nutritionist	None	Child with autism
Occupational Therapist	Skill development	Occupational therapist	Family, Clinical psychologist	Child with autism
Pediatrician	General health care	Pediatrician	Family, Various therapists	Child with autism
Pharmacies	Medication supply	Pharmacists	None	Child with autism
Physical Therapist	Physical skill development	Physical therapist	Family, Various therapists	Child with autism
Public Health Departments	Public health services and policy making	Public health officials	Various health professionals	Autism community
Residential Service Providers	Residential care and support	Residential staff	Support staff, Therapists	Adult with autism
School Administration	School policy making and implementation	School administrators	Teachers, Therapists	Child with autism
Social Worker	Case management and support	Social worker	Various therapists, Family	Child with autism and family
Special Education School	Education and support	Special education teachers	Teaching assistants, Therapists	Child with autism
Speech Therapist	Language and communication skills development	Speech therapist	Family, Clinical psychologist, Occupational therapist	Child with autism
Technology Solution Providers	Technological aids provision	Tech developers	Support staff	Child with autism
Vocational Training Providers	Vocational skills training	Vocational trainers	Support staff	Adult with autism

Figure 3



OT and Speech Therapy Providers for 20-Year-Olds with Medicaid in Huntsville ▾

Do you know of providers for OT and Speech in Huntsville that accept Medicaid as insurance and serve 20-year-olds?

AA

Unfortunately I do not have specific provider recommendations for OT and speech therapy services that accept Medicaid and serve 20-year-olds in the Huntsville area. However, here are a few suggestions to help you find options:

- Contact Alabama Medicaid directly to ask for a list of Medicaid-approved OT and speech therapy providers in the Huntsville region that see adult patients. They should be able to guide you to some options.
- Search online directories like Psychology Today or online listings for terms like "Medicaid speech therapy Huntsville AL adults" or "Medicaid occupational therapy Madison County 20-year-olds." Make some calls and ask if they accept Medicaid for your age.
- Check with local clinics, rehab centers, hospitals and university therapy training clinics in the Huntsville region as some may offer Medicaid accepted speech, language and OT services for adults and young adults specifically.
- Reach out to advocacy groups like Alabama Disabilities Advocacy Program or state developmental disability agencies to get recommendations for Medicaid providers for 20-year-olds needing speech and OT services.

I hope those suggestions are a helpful starting point! Let me know if you need any other assistance in your search.

AI

Copy Retry

👍 🧡

8

You

Do you know of providers for OT and Speech in Huntsville that accept Medicaid as insurance and serve 20-year-olds?

Minerva

In Huntsville, Alabama, there are providers offering Occupational Therapy (OT) and Speech Therapy services for 20-year-olds that accept Medicaid:







- Dailey Cove Speech Therapy**
 - **Services:** Speech Therapy
 - **Address:** 1420 Weatherly Rd SE UNIT A, Huntsville, AL 35893, United States
 - **Phone Number:** 256-555-7765
 - **Website:** www.daileycove.com/
 - **Insurance:** Accepts Medicaid, AllKids, BCBS, Cigna, United, Tricare, Aetna
 - **Age Group:** Children/Adolescents, Adults.
- More than Words Speech Therapy**
 - **Services:** Speech Therapy
 - **Address:** 1101 McMurtre Dr NW Suite G1, Huntsville, AL 35893
 - **Phone Number:** 256-213-7477
 - **Website:** mtwspeechtherapyhaval.com/
 - **Insurance:** Accepts Medicaid
 - **Age Group:** Information on specific age groups not provided.


For Occupational Therapy services accepting Medicaid for adults, the search did not specify a provider that explicitly matches all criteria, including serving adults and accepting Medicaid, within the information provided. I recommend contacting **Huntsville Hospital Therapy & Sports Rehab** for potential OT services as they offer therapy and rehabilitation services, although the specific age groups and services (OT or Speech) they cater to are not detailed in the provided information:

- **Address:** 1900 Memorial Drive SW, Huntsville, AL 35894
- **Phone Number:** 256-265-7101
- **Website:** <http://www.huntsvillehospital.com/>

Please verify with each provider regarding availability of Occupational Therapy services for adults and confirm that they accept Medicaid.

Message Minerva...

  127.0.0.1:5000    |  ...

 **ATHENA**

Do you know of providers in
Huntsville for OT and Speech that
accept Medicaid as insurance and
serve 20-year-olds

Here are the top 3 service providers
out of 11 that match your query:

Provider Name: Encore
[REDACTED]

Address: Huntsville, AL 35892-1002
Phone Number: 256-350-1704
Email: [REDACTED]
Website:
<https://www.enclerchab.com/location/>

Provider Name: Parent
Training Applied Behavior Analysis
(ABA) After-school Programs Social
Skills
Address: Huntsville, AL 35890
Phone Number: 600-352-5010
Email: [REDACTED]



[Click here to access/download](#)

Supporting Information

AthenaFinal_PlosOne_SupportingInformation.docx



Using AI to Enhance Healthcare Resource Management and Allocation: A Focus on the Autism Community in Alabama

Armin Ahmadi*, Jerome Baudry*, Nathan Tenhundfeld\$, Kelly Goff^, Daniel Adamek&.#

*: Department of Biological Sciences, The University of Alabama in Huntsville, Huntsville, AL, 35899

\$: Department of Psychology, The University of Alabama in Huntsville, Huntsville, AL, 35800

Current address: Heterodox Academy, New York, NY 10038

^: Autism Services, Alabama Department of Mental Health, Montgomery, AL, 36130

&: Little Orange Fish, Huntsville, AL, 35806

#: Corresponding author: Daniel@littleorangefish.org

Abstract

This study investigates the potential of artificial intelligence, particularly Natural Language Processing and large-scale language models, to improve resource management and service access for individuals with autism in Alabama. The research aims to address existing challenges in navigating complex datasets and enhancing social work practices through the development and evaluation of AI-driven tools.

We designed and tested AI systems, including general language models, domain-specific chatbots powered by advanced language models, and a Retrieval-Augmented Generation framework. A standardized set of queries was used to simulate real-world scenarios encountered by social workers and families seeking autism-related resources. System performance was evaluated based on precision, recall, and response accuracy.

Results demonstrated that the Retrieval-Augmented Generation framework achieved superior performance compared to traditional methods and general models. The integration of domain-

specific information significantly enhanced the accuracy, contextual relevance, and usability of the generated responses. This improvement was particularly evident for queries requiring geographic specificity, multiple criteria, or complex contextual understanding.

The findings highlight the transformative potential of AI-driven tools in improving social work efficiency and enhancing healthcare equity. By streamlining care coordination and delivering accurate, contextually relevant information, these systems offer scalable solutions to improve access to autism-related services. Future research should focus on addressing data quality, minimizing biases, and ensuring ethical deployment to build trust and support widespread adoption of these tools.

Keywords

Natural Language Processing (NLP), Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Autism Resource Management, AI in Social Work, Healthcare Delivery Efficiency.

Introduction

Navigating fragmented healthcare systems presents significant challenges for patients, caregivers, and associated services, particularly for those managing chronic or complex conditions like autism. These inefficiencies often result in unmet medical needs and reduced patient satisfaction (Griese et al., 2020). Integrated care, which emphasizes coordination and communication among healthcare providers, has consistently demonstrated improvements in patient outcomes and satisfaction by streamlining services and reducing redundancies (Renghea et al., 2022; Kurpas et al., 2021). Holistic care platforms further enhance healthcare efficiency and sustainability, though implementing such systems is not without obstacles (Eriskin, Karatas, and Zheng, 2024).

Human social workers play a pivotal role in care coordination but often face overwhelming workloads due to high caseloads and undefined roles, which can limit their effectiveness (Yeo et al., 2022). Furthermore, only 30% of healthcare systems report robust collaborations with community-based service providers, exacerbating inefficiencies in care delivery (Kreuter et al., 2021). Issues such as human error and limited information processing further hinder case management, making it difficult to monitor outcomes and ensure accountability (Lee, 2001). While addressing social needs within clinical care demonstrates a strong return on investment, fragmented systems continue to drive high financial and human costs (Karlin, 2022).

Artificial Intelligence (AI) emerges as a transformative tool for addressing these challenges. By processing large datasets, identifying patterns, and providing evidence-based insights in real time, AI improves resource management and care coordination (Gandhi et al., 2023; Alowais et al., 2023). Automating administrative tasks like scheduling and record-keeping reduces inefficiencies and human error, with McKinsey estimating up to a 30% reduction in healthcare costs through AI-

driven operational improvements (Friesdorf, 2019). Beyond administrative efficiencies, AI also enables personalized care through data analysis, delivering tailored interventions that enhance treatment outcomes (Schork, 2019). For instance, during the COVID-19 crisis, AI-powered tools effectively guided patients to appropriate care, demonstrating the potential for streamlined coordination and improved outcomes (McRae et al., 2020). Global applications further illustrate how AI can optimize resource allocation, reduce delays, and minimize redundant treatments (Chen et al., 2019; Kitsios et al., 2023). Social workers, in particular, benefit from AI's ability to track interventions, forecast challenges, and allow for a focus on higher-level care needs (Bin et al., 2022).

However, Autism Spectrum Disorder (ASD) exemplifies a domain where these systemic challenges are particularly acute and where AI-driven solutions could provide critical support. ASD management requires a coordinated, multidisciplinary approach due to its complexity and the necessity for collaboration among diverse stakeholders (CDC, 2023). Effective care transitions, particularly from childhood to adult services, remain critical but are often hampered by personnel shortages and systemic inefficiencies (Fulceri et al., 2023). Socioeconomic and cultural factors further compound these challenges, necessitating adaptive and equitable interventions (Koum et al., 2023). The prevalence of ASD has significantly increased, with the CDC reporting that one in 36 children in the U.S. is diagnosed with ASD, a trend mirrored in Alabama with a 22% rise in prevalence since 2021 (Maenner, 2023; Shaw, 2023). Navigating autism-related services presents additional challenges due to complex eligibility requirements, insurance constraints, and a fragmented landscape of healthcare, education, and community resources. Families and social workers often face difficulties coordinating services across these sectors, which leads to delays or gaps in accessing vital services. These challenges are especially pronounced in rural and

underserved communities, where resource scarcity intersects with geographic and economic barriers (Magaña et al., 2016; Zuckerman et al., 2015). Families of autistic individuals frequently report high caregiver burden due to the constant navigation of disconnected systems, including healthcare providers, educational institutions, therapeutic services, and legal advocates (Warfield et al., 2015). These barriers result in inequities in care and missed opportunities for timely intervention.

Social workers play a critical role in supporting individuals with ASD and their families, coordinating care across healthcare, education, and community settings. AI-driven tools enhance their effectiveness by streamlining data access, enabling better intervention tracking, and providing actionable insights for individualized care (Ramadoss, 2024). By integrating diverse datasets, AI offers a holistic view of each client's needs, leading to tailored care plans and improved collaboration among service providers. Studies show a generally positive reception toward AI tools in healthcare, with about 67% of internet users expressing acceptance of AI-led services (Åhs, Mozelius, & Espvall, 2022). However, successful integration of AI requires targeted training to navigate challenges in prediction, adaptivity, and decision-making (Åhs et al., 2022).

By focusing on the autism community in Alabama, this study illustrates how AI can transform healthcare resource management. AI technologies empower social workers by streamlining care coordination, enabling more personalized and effective support, and addressing critical gaps in service delivery. Here, we focused specifically on social workers as the primary end users because the tool was developed as a proof-of-concept social worker assistant. Social workers were chosen not only due to their central role in managing autism-related care but also due to our direct~~This work collaboration with the Alabama Department of Mental Health, which provided extensive~~leverages input on such use cases, workflows, and data sources, and provide a basis for

the extension of this work to - While other stakeholders such as caregivers, physicians, and educators are also important actors in the autism care ecosystem, their perspectives and use cases were outside the scope of this initial prototype. Future versions of this system may extend support to those roles as well. This research provides a scalable model for leveraging AI to improve care for underserved populations, highlighting its potential to enhance resource allocation, reduce systemic inefficiencies, and deliver better outcomes across complex healthcare systems.

Background and Technical Foundations

Leveraging AI in Healthcare Resource Management

Artificial Intelligence (AI), particularly Natural Language Processing (NLP) and Large Language Models (LLMs), has emerged as a transformative force in healthcare resource management. These technologies efficiently process large volumes of unstructured data, such as clinical notes and patient records, which are often underutilized in traditional systems (Mah, 2022). For example, NLP applications in radiology have demonstrated 95-96% consistency in decision-making, highlighting their potential to improve outcomes (Morozov et al., 2018). NLP and LLMs extract critical insights from unstructured data, enhancing treatment planning, clinical trial recruitment, and patient engagement (Hossain et al., 2023). LLMs, particularly when integrated into Clinical Decision Support Systems (CDSS), further revolutionize care delivery by leveraging IoT devices for real-time insights (De Vito, 2024).

Traditional NLP in Healthcare: Strengths and Limitations

Traditional NLP systems rely on rule-based algorithms and machine learning models to perform tasks like medical coding and information extraction, often achieving high accuracy in specific applications. For instance, an XGBoost classifier achieved an F1 score of 0.8881 in diagnosing psychosis from psychiatric notes, and other models have reached 99% accuracy in classifying PHI-related data (Hua et al., 2024; Zhang & Jiang, 2023). However, these systems struggle with scalability, contextual understanding, and adaptability to new or complex medical terminology. Scalability challenges, particularly in cloud-based environments, can increase latency by up to 50% (Kumar, Dubey, & Pandey, 2021). Additionally, traditional models experience performance declines of 15-20% when faced with new ICD codes without retraining (Jha, Almagro, & Tissot, 2023). Manual feature engineering is another limitation, requiring intensive human effort that can be reduced by up to 70% with deep learning models (Gupta, 2019). These systems also fail to interpret nuanced queries effectively, underperforming by 20-30% compared to transformer-based models like BERT in tasks involving complex language patterns (Dong et al., 2023; D, 2024).

Advancements with LLMs: Benefits and Challenges

LLMs, including GPT models, represent significant advancements in language processing due to their transformer-based architectures and ability to handle complex, context-rich queries with minimal retraining (García-Méndez & de Arriba-Pérez, 2024). BERT-based models, for instance, improved patient record summarization accuracy by 20-30% over traditional approaches (D, 2024). These models excel in few-shot or zero-shot learning scenarios, reducing the need for extensive fine-tuning, with GPT-3 demonstrating a 30% improvement in task accuracy using natural-language prompts (Gao, Fisch, & Chen, 2021). Furthermore, studies have shown that LLM-generated summaries are preferred over human-generated ones by up to 40% for their completeness and correctness (Van Veen et al., 2024).

Despite these advantages, LLMs require significant resources for fine-tuning and deployment. A fine-tuned GPT-3.5 model, for example, achieved an F1 score of 0.80 compared to 0.46 for an out-of-the-box version, but such training can cost millions of dollars (Stewart, Hodkiewicz, & Li, 2023; Chen et al., 2024). Biases in training data and the “black box” nature of LLMs also pose challenges for trust and interpretability, with studies revealing significant biases in applications like physician use cases (Ayoub et al., 2024; Tao et al., 2024). Additionally, prompt engineering plays a critical role in optimizing LLM performance, particularly in specialized domains like healthcare, where well-designed prompts can achieve nearly 90% precision and recall (Polak & Morgan, 2024).

Retrieval-Augmented Generation (RAG): Enhancing LLMs

RAG combines the generative capabilities of LLMs with dynamic information retrieval, improving relevance and accuracy in healthcare applications. By incorporating real-time domain-specific data, RAG systems address the limitations of static LLMs, such as knowledge gaps and biases. For instance, BiomedRAG achieved micro-F1 scores of 81.42 and 88.83 in healthcare tasks, surpassing traditional NLP systems (Li et al., 2024). This approach also enhances transparency in decision-making, as demonstrated in its success managing chronic Hepatitis C Virus infections and reducing biases in handling sensitive data (Krešević et al., 2024; Pouplin et al., 2024).

RAG systems are highly scalable and efficient for time-sensitive healthcare scenarios. In one preoperative medicine case study, an LLM-RAG model processed 1,260 responses using 35 guidelines within 15-20 seconds, compared to the 10 minutes required by human experts. This model improved GPT-4’s accuracy from 80.1% to 91.4%, exceeding human performance (Ke et

al., 2024). By dynamically integrating retrieved documents into LLM outputs, RAG enhances open-domain question answering and supports better patient outcomes (Siriwardhana et al., 2023).

Methodology

This study develops an AI-driven model to improve resource management and support social workers in autism care in Alabama. By integrating diverse primary (healthcare providers), secondary (social services), and tertiary (community organizations) stakeholders, the model tries to address inefficiencies, enhances care coordination, and improves access to services. Key stakeholders were identified and categorized (Table S1), and their roles and interactions mapped to reveal gaps in existing systems to feed the model with a broad understanding of healthcare systems and public services, incorporating typical roles and responsibilities from healthcare, social services, education, legal services, and other relevant areas.

Data Collection and Integration:

To construct a comprehensive database of autism care service providers in Alabama, manual and automated web scrape identified 513 providers, and collaboration with the Alabama Department of Mental Health added [another](#) 1,102 stakeholders [entities](#) from over 350 textual documents [including therapy providers, community centers, schools, nonprofit advocacy groups, and public agencies relevant to autism care in Alabama. This resulted in a final dataset of 1,615 entries. The database supported both chatbot systems \(Athena and Minerva\) used in this study. In this context, “providers” refers to both individual practitioners and organizations offering autism-related services. Each provider’s entry represents a unique service unit with a verifiable address and operational role.](#) Key details such as provider names, addresses, contact information, and service descriptions were digitized using Optical Character Recognition (OCR) software (Adobe, 2024).

[All digitized records were then manually reviewed to validate key fields for accuracy. This included human verification of names, service descriptions, and addresses to ensure data integrity.](#)

Data originally in unstructured formats, including PDFs and handwritten notes, were converted into structured CSV databases using GPT-4 (OpenAI, 2024) for standardization and consistency.

Service Provider Tagging

A tagging system was developed to facilitate querying and data retrieval. Each service provider was assigned binary tags ('1' for offered services, '0' for non-offered) across 300 categories (example in Table 1), covering diverse needs of the autism community, such as insurance, healthcare services, therapeutic supports, and education (Hervieux and Wheatley, 2022). [The binary tagging system was selected to enable scalable and interpretable matching between user queries and provider offerings. “Offered services” were defined as those that providers publicly listed or were documented as delivering in source materials. When services were ambiguous or conditional \(e.g., only offered to certain age groups\), we opted to create more specific tags \(e.g., “early intervention” or “adult services”\) rather than apply weighted labels. Some proprietary tags \(e.g., “Cigna,” “United Healthcare”\) were included to reflect insurance acceptance, which is a critical filtering dimension for many families. These were treated as separate but parallel attributes to broader categories such as “financial support.”](#) Expert social workers reviewed the database to validate the accuracy and relevance of the tags.

Table 1 [Sample C](#)ategorization of Service Providers for Autism Care. ~~This table presents a small representative subset of the over 300 service tags used in the full dataset.~~

Category	Tags
Insurance & Financial	Medicaid, AllKids, BCBS, Cigna, United, Tricare, Aetna, Medicare, Financial Supports
Healthcare Services	Pediatricians, Neurologists, Psychiatrists, Dentists, Nutritionists, Gastroenterologists, Pulmonologists, Endocrinologists, Dermatologists, Optometrists, Pediatric Neurologists
Therapeutic Services	Applied Behavior Analysis (ABA), Occupational Therapy, Physical Therapy, Speech Language Therapy, Behavioral Therapy, Art/Dance/Music Therapists, Aquatic Therapists, Equine Programs
Mental & Emotional Support	Psychologists & Counselors, Social Workers, Crisis Services, Support Groups, Substance Use Treatments, Counseling for Children, Teens, and Adults
Education & Development	Early Intervention Providers, Specialized Schools, Post-Secondary Education, Autism Evaluation, Autism Friendly Services, Educational Supports
Legal & Advocacy	Attorneys, Legal Services, Advocacy & Support, Autism Associations
Community & Social Support	Community Activities, After-school Programs, Camps, Social Skills Programs, Recreational & Community Activities, Housing & Community Living, Faith-Based Organizations
Employment & Vocational	Employment Supports, Vocational Rehabilitation, Job Training Programs
Special Needs Services	Assistive Technology, Genetic Testing, Safety/Protection, Transportation, Feeding Therapies/Interventions, Sleep Supports
Crisis & Immediate Care	Emergency Services, First Responder Resources, Inpatient Treatment Care Centers, Crisis Intervention Services
Residential & Home Care	Residential Programs, Home School, Day Programs, Housing Assistance

Socioeconomic Mapping of Resources

To identify any possible disparities in autism-related service availability, a spatial analysis was conducted by mapping service provider locations against socioeconomic factors obtained from the U.S. Census Bureau (US Census, 2024). Key factors included population density, ethnic composition, median household income, and poverty rates. Custom Python scripts utilized the Census API for data retrieval, Pandas ([McKinney, 2010](#)) for data organization, and Folium ([Folium 2024](#)) for interactive mapping. Service provider addresses were geocoded using the

Nominatim geocoder (OpenStreetMap, 2024) via the GeoPy Python library and presented in a map of resource accessibility and underserved areas.

Development of the NLP-Driven Chatbot

An NLP-driven chatbot was developed to provide users with a tool to locate autism service providers based on specific needs and locations. The service provider database, initially in CSV format, was serialized into JSON files to optimize data handling and interoperability with NLP tools. (Bird et al., 2009). In house Python script were used to create comprehensive provider profiles, encapsulating details such as name, address, services offered, and additional notes. The chatbot was built using Python 3.10 and the SpaCy v3.7 NLP library (Honnibal & Montani, 2017), utilizing the standard English NLP model (en_core_web_sm) for tasks such as tokenization, lemmatization, and named entity recognition (NER). [A custom-built service_synonyms dictionary was developed to normalize user queries across heterogeneous provider terminology.](#) For example, terms such as “speech therapy,” “speech pathologist,” and “SLP” were grouped under a unified category (“Speech Language Therapists”). [The dictionary was manually curated with input from expert social workers and informed by best practices in semantic alignment and concept disambiguation.](#) ~~A service_synonyms dictionary mapped user terminology (e.g., terms like "speech therapy" and "speech pathologist" map to "Speech Language Therapists.") to standardized service categories, addressing linguistic variability~~ (Navigli, 2009), ~~but did not directly use external lexical resources such as WordNet or BabelNet.~~ When exact matches were unavailable, fuzzy string matching with RapidFuzz (Gohlke, 2024) with a score cutoff of 80% was employed. Geographic information was managed by extracting locations through SpaCy’s NER capabilities and geocoding user-provided locations with Python’s geopy library (v2.4.1), specifically leveraging Nominatim geocoder

(OpenStreetMap) for location identification. The geodesic distance between users and identified service providers was calculated using geopy, prioritizing providers within a 50 km radius. The system incorporated error-handling mechanisms that ensured incomplete or failed geocoding queries did not eliminate providers from recommendations but rather marked distances as "Unknown" to maintain inclusiveness. The chatbot interacts through a Flask interface ([Grinberg, 2014](#)) hosted Linux server with Ubuntu 22.04 LTS (Figure 1), prompting users to enter queries in natural language. It provides structured responses that include provider information, services offered, and contact details, enhancing accessibility and usability.

Figure 1: NLP-driven Chatbot [response example, returning a filtered set of autism service providers based on user query criteria. Information shown was drawn from publicly available sources \(e.g., provider websites\) and manually blurred to remove identifying details. No personal health information or private individual data is presented..](#)

Advancements with LLM-Powered Chatbots

To explore the applicability of Large Language Models (LLMs) in enhancing autism care support, we first tested widely available generative AI models (out-of-the-box LLMs), including ChatGPT (OpenAI), Google Gemini, Microsoft Copilot, and Claude (Anthropic) which are recognized for their advanced natural language understanding, ability to handle complex and nuanced queries, and general versatility across a range of domains (Minaee et al. 2024). These models were assessed using a standardized set of queries designed to simulate scenarios encountered by social workers and families seeking autism-related services in Alabama. The queries aimed to evaluate their ability to handle complex, multi-criteria requests, such as identifying local service providers based on specific needs and insurance requirements.

Based on observed strengths and limitations, we developed a domain-specific chatbot named "Minerva," built specifically upon OpenAI's GPT-4 model (GPT-4 API, version gpt-4-1106-preview). The GPT-4 model was employed with carefully engineered prompt templates (Table 2), emphasizing concise, clear instructions explicitly linked to autism service criteria and geographic specificity within Alabama. Prompt engineering utilized simple few-shot prompting techniques, explicitly listing query-response examples to guide model behavior without further fine-tuning. The structured autism provider database described earlier, serialized as JSON, provided the domain-specific context for generating responses. Iterative refinement cycles integrated structured user feedback obtained through a binary ("Yes/No") feedback loop built into the user interface to progressively improve response relevance and accuracy. This approach was designed to address limitations observed in out-of-the-box LLMs, specifically their inability to reliably produce actionable local recommendations without supplementary domain context."

Figure 2: LLM driven Chatbot [response example](#). [Provider details are sourced from public records and have been blurred for privacy. No sensitive or user-specific information is shown..shown.](#)

Table 2 Criteria for GPT-4 Prompt Design

Prompt criteria	Detailed instructions
Clarity and Brevity	providing clear and concise responses, avoiding overwhelming users with excessive details, and adhering to brevity unless additional information is necessary
Data-Driven Responses	uses a structured database of autism care providers in Alabama, including service types, locations, and contact information, to accurately match user queries with service recommendations
Feedback Integration	After each interaction, Athena prompts users for feedback with a simple "Yes/No" question. This feedback is vital for continuous learning, refining responses based on user input to enhance effectiveness
Privacy / Customization	respects user privacy, handling interactions without storing personal data unless explicitly authorized, and adheres to stringent privacy standards to ensure secure communications
Guided Interactions	asks direct, relevant questions for clarification, maintaining a supportive tone to precisely identify user needs without causing frustration

Implementation of Retrieval-Augmented Generation (RAG)

To further enhance the precision and transparency of the chatbot, we implemented a Retrieval-Augmented Generation (RAG) framework consisting of two main stages: document retrieval and response generation. For the retrieval stage, we employed the OpenAI text-embedding-ada-002 model to generate dense vector embeddings for provider profiles, capturing semantic relationships beyond keyword matching (Neelakantan et al., 2022). Each profile contained comprehensive details, including provider name, address, services offered, and contact information. These embeddings were stored in a vector database, enabling efficient similarity searches. When a user query was received, it was converted into an embedding, and cosine similarity scores were calculated between the query embedding and stored provider embeddings. Provider profiles with similarity scores exceeding a predefined threshold (0.5) were deemed relevant.

In the subsequent generation stage, relevant documents retrieved from the first stage were incorporated into prompts provided to the OpenAI GPT-3.5-turbo model. Generation parameters were explicitly set, including temperature (set to 0) and maximum tokens (set to 500), to optimize precision. The GPT-3.5-turbo assistant was explicitly instructed to generate outputs solely based on the retrieved information, enhancing transparency and minimizing hallucinations.

To further refine recommendations, geographical filtering was applied. User-provided locations were geocoded using the Nominatim geocoder from the geopy library (Gerber, 2016). Provider locations were similarly geocoded, and the geodesic distance between user and provider locations was calculated. Providers located within a specified radius (50 km) were included in the results. Missing or incomplete geocoding data were handled by marking providers' distances as "Unknown," ensuring inclusivity in recommendations. Finally, a Flask-based web application

provided a user-friendly interface, allowing users to input natural language queries and receive structured responses detailing provider names, distances, services offered, and contact information.

Evaluation Framework

The performance of the developed systems, was then evaluated aimed to assess their ability to provide accurate, contextually relevant, and actionable responses for autism-related queries. We developed a standardized set of test queries reflecting real-world scenarios encountered by families and social workers seeking autism services in Alabama. These queries were categorized into six criteria:

- Simple Queries: Single-service requests with minimal criteria.
- Complex Queries: Multi-criteria requests involving multiple services or conditions.
- Geographically Specific Queries: Requests targeting specific locations.
- Synonyms and Varied Phrasing: Queries using alternative terminology or colloquial expressions.
- Ambiguous Queries: Broad or unclear requests requiring interpretation.
- Less Common Services: Queries for specialized or rare services like equine therapy.

Each category included ten representative queries (e.g., “Find an Occupational Therapist in Huntsville, AL, who accepts Medicaid and works with young adults”). These queries were developed in collaboration with the Alabama Department of Mental Health to ensure relevance and practical alignment with user needs. [The output query results were categorized as Correct Matches only if they fully satisfied all constraints in the query, including service type, geographic location, insurance acceptance, and age group where applicable. Partially correct matches \(e.g.,](#)

matching location and service but not insurance) were not counted as correct. These results were analyzed using precision (how many of the returned results were correct), recall (how many of the correct results were returned compared to the total possible correct results), and F1 score (the harmonic mean of precision and recall, balancing both metrics). ~~The output query results were categorized as Correct Matches if fully satisfy the query criteria and incorrect matches if fail to meet the query criteria, and analyzed using metrics including precision (how many of the returned results were correct), recall (how many of the correct results were returned compared to the total possible correct results), and F1 score (harmonic mean of precision and recall, balancing both metrics) (Powers 2020).~~ In the result section we use the term “accuracy” as a shorthand label to represent the F1 score, which provides a balanced summary of system performance across multiple query constraints. This evaluation aimed to provide a descriptive, albeit semi-quantitative, performance comparison across the tested models (Traditional NLP, Out-of-the-box LLMs, Domain-specific LLM, and RAG-based system). ~~Given the exploratory nature of this study and its focus on demonstrating feasibility, formal statistical significance tests (e.g., McNemar’s test, t-tests) were not conducted. Instead, P-performance was evaluated using descriptive metrics such as precision, recall, and F1 scores metrics across representative query types. While this approach is consistent with the study’s proof of concept design, we recognize that future evaluations will need benefit from to incorporating e inferential statistical analyses to validate model performance more robustly quantitatively. This remains an important direction for ongoing and future work. Given the exploratory nature of this study and its focus on demonstrating feasibility, formal statistical significance tests (e.g., McNemar’s test, t-tests) were not conducted. The observed performance differences provide initial evidence of the system’s potential but warrant further statistical validation in future studies.~~

$$Precision = \frac{\text{Correct Matches}}{\text{Total Matches Returned}} \quad Recall = \frac{\text{Correct Matches}}{\text{Total Matches Available}} \quad F1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Results

Socioeconomic Mapping of Autism Services

The resulting spatial representation (Figure 3) of socioeconomic indicators allows for clear understanding of areas underserved by autism-related services. While this study's primary aim was not to explore or explain these disparities, this analysis revealed disparities in the availability of autism-related services across Alabama. Regions with poverty rates exceeding 19% and minority populations above 20% demonstrated significantly fewer service providers per capita compared to the state average. Some counties lacked autism-specific providers altogether (Figure 3). These findings underscore inequities in resource distribution, suggesting a need for targeted interventions to address underserved areas. Although a range of factors may contribute to these service gaps (Aylward, Gal-Szabo, and Taraman 2021), the maps suggest a potential need for a data-driven approach to improve resource allocation.

Figure 3: [Geographic distribution of autism-related service providers in Alabama, overlaid on key socioeconomic indicators. Heatmap represents provider density. Insets show demographic context, including poverty rate, population distribution, and median income. The figure highlights service disparities that emerged during database construction and motivate the need for equitable access tools like Athena and Minerva.](#)
[Disparities in Autism-Related Service Providers in Alabama](#)

Evaluation of the NLP-Driven Chatbot

The initial NLP-driven chatbot was evaluated to assess its capability to process autism-related service queries (Table 3). Traditional NLP methods implemented in our system primarily relied

on syntactic processing and rule-based matching. These approaches required predefined tags and depended on exact matches or cosine similarity to link user queries to services.

Table 3 Performance Metrics of the NLP-Driven Chatbot

Sample Query	accuracy rate
Find an Occupational Therapist (OT) and Speech Therapist in Huntsville, AL who accepts Medicaid and works with 20-year-olds.	63.2 %.
Locate a Speech Therapist in Birmingham, AL specializing in early intervention and accepting private insurance	74.5 %.
Search for a Pediatric Neurologist in Montgomery, AL who offers telehealth services and accepts Medicare.	58.9 %.

This method, while effective in certain scenarios, exhibited limitations in handling the nuances of natural language, particularly when users employed varied phrasings, idiomatic expressions and synonyms, or complex query intents which are reported in research articles (Olujimi and Ade-Ibijola 2023) where the system struggled to accurately match these to the corresponding services. As a result in our case, the chatbot occasionally produced incomplete or less accurate responses, particularly in cases where the user's language deviated from the predefined tags. The chatbot's performance across previously randomly generated test queries of Table 3 shows where non-standard language was used, the accuracy of the responses dropped (Table 4).

Table 4 Challenges Faced by the NLP Chatbot in Query Processing

Sample Query	Challenges	accuracy rate
Therapist in Huntsville, AL who does OT and speech therapy for a young adult covered by Medicaid.	Uses "recommend a therapist" instead of directly specifying "Occupational Therapist (OT) and Speech Therapist," and "young adult" instead of specifying age.	57.1 %.
Where in Birmingham, AL can I find a specialist for early language intervention who works with private insurers?	Uses "specialist for early language intervention" instead of "speech therapist specializing in early intervention," and "private insurers" instead of "private insurance"	48.6 %.

Are there pediatric neurologists in Montgomery, AL who offer virtual consultations and accept Medicare?	Uses "virtual consultations" instead of "telehealth services" and "pediatric neurologists" is rephrased slightly as "Are there pediatric neurologists" instead of "Search for a pediatric neurologist,"	53.7 %.
---	---	---------

Table 5 presents the different aspects of the NLP system, including its handling of simple, complex, and geographically specific queries, as well as its ability to process synonyms, varied phrasing, ambiguous language, and requests for less common services highlighting where the system performed well and where it struggled.

Table 5 Scenario-Based Performance Evaluation of the NLP Chatbot

Criteria	Query	Expected Service	Matched Service	Reason for Mismatch
Simple, straightforward queries	"Find a Speech Therapist in Birmingham, AL who accepts Medicaid."	Speech Therapist in Birmingham accepting Medicaid	Speech Therapist in Birmingham accepting Medicaid	Exact Match
Complex queries with multiple criteria	"I need an Occupational Therapist and a Speech Therapist in Huntsville, AL who accept Medicaid and specialize in working with young adults aged 20."	OT and Speech Therapist in Huntsville accepting Medicaid	Only Speech Therapist in Huntsville accepting Medicaid	Couldn't match both OT and Speech Therapist together
Queries with synonyms or varied phrasing	"Looking for a language specialist in Birmingham, AL who handles early intervention and takes private health insurance."	Speech Therapist specializing in early intervention	General Speech Therapist	Synonym "language specialist" not correctly linked to "Speech Therapist"
Geographically specific queries	"Are there any pediatric neurologists near downtown Montgomery, AL who offer telehealth services and accept Medicare?"	Pediatric Neurologist in Montgomery offering telehealth	Pediatric Neurologist in Montgomery	Location specificity not considered effectively by NLP system
Ambiguous or incomplete queries	"Can I find someone in Alabama who helps with autism?"	Autism specialist or center	Various unrelated services	Ambiguous phrasing, system unable to accurately determine intent

Queries with less common or rare services	"Where in Mobile, AL can I find an equine therapy program for children with autism that accepts Medicaid?"	Equine therapy program for children in Mobile accepting Medicaid	General therapy services in Mobile	Rare service not sufficiently represented in training data
---	--	--	------------------------------------	--

Table 6 shows the distribution of the results of each query type (averaged on ten queries), alongside calculated precision, recall and F1 score. The NLP chatbot performed well on simple, single-service requests with direct phrasing while queries requiring semantic understanding, multi-criteria matching, or complex phrasing resulted in lower performance suggesting that the traditional NLP methods are effective for straightforward tasks but lack the semantic depth and contextual awareness needed for nuanced or highly specific queries.

Table 6 Precision, Recall, and F1 Scores of the NLP Chatbot

Criteria	Correct Matches	Partially Correct Matches	Incorrect Matches	Precision (%)	Recall (%)	F1 Score
Simple, straightforward queries	7	2	1	87.5	70.0	77.8
Complex queries with multiple criteria	5	3	2	71.4	50.0	58.8
Queries with synonyms or varied phrasing	4	3	3	57.1	40.0	47.0
Geographically specific queries	6	2	2	75.0	60.0	66.7
Ambiguous or incomplete queries	3	2	5	37.5	30.0	33.3
Queries with less common or rare services	2	1	7	22.2	20.0	21.0

Performance of Out-of-the-Box LLMs

Generative AI models were evaluated using the same query set (Table 7). While these models exhibited superior language comprehension and handled nuanced queries better than the NLP chatbot, they struggled with regional specificity and consistency.

Table 7 Comparison of Established LLMs for Autism Resource Queries

Sample Query	accuracy rate			
	ChatGPT (%)	Google Gemini (%)	Microsoft Copilot (%)	Claude (%)
Find an Occupational Therapist (OT) and Speech Therapist in Huntsville, AL who accepts Medicaid and works with 20-year-olds.	47.8	51.6	42.3	0.0
Locate a Speech Therapist in Birmingham, AL specializing in early intervention and accepting private insurance	43.2	37.5	33.7	2.1
Search for a Pediatric Neurologist in Montgomery, AL who offers telehealth services and accepts Medicare.	57.9	46.3	48.6	3.4

For instance, they were able to parse queries involving multiple provider types or specific insurance requirements without difficulty. Their broad knowledge base and advanced linguistic capabilities allowed them to interpret complex sentence structures and varied phrasing more effectively than the rule-based approach previously used. But despite their language processing abilities, the out-of-the-box LLMs struggled to provide region-specific or highly specialized information related to autism care in Alabama. For example, they often failed to distinguish between service providers who were available locally and those who were not. As an example, ChatGPT's failure to understand the local context led it to suggest service providers that were either not in Alabama or did not specialize in autism care, providing generic advice rather than specific, actionable information. Or Microsoft Copilot's lack of awareness of regional service providers resulted in recommendations that were irrelevant to the user's location, often suggesting national organizations or services unavailable in Alabama. In several cases, the models returned generic responses rather than the specialized advice expected in the autism care context. For instance, in answer to a query on finding an OT and speech therapist in Huntsville, AL, Claude gives general suggestions like contacting Alabama Medicaid, doing an online search, or contacting local universities for recommendations (Figure 4). Also, the outputs might not be consistent, if a query

is being asked multiple times, each time a different answer might show up. For instance Table 8, shows a query that was asked (verbatim) five times from Gemini LLM and each time a different response was generated by LLM.

Figure 4: Example of LLM Fails to Provide Relevant Information

Table 8 Inconsistencies in LLM Responses to Identical Queries

Query	Summary of Response
Find two Occupational Therapist (OT) and Speech Therapist in Huntsville, AL who accepts Medicaid and works with 20-year-olds.	1- General suggestions on finding therapists, including contacting Medicaid providers, local hospitals, universities, and online directories like Beaming Health.
	2- Provides names of specific therapy centers (Child'sPlay, Therapy Achievements, Pediatric Outpatient Center) that accept Medicaid and suggests further searches.
	3- Suggests using Beaming Health, contacting Medicaid providers for in-network therapists, and checking with local hospitals or rehab centers for therapists.
	4- Lists specific therapy centers (Child'sPlay, More Than Words, Pediatric Therapy Link) that accept Medicaid and work with 20-year-olds.
	5- Provides a list of therapy centers with addresses on Google Maps but recommends contacting them directly to confirm Medicaid acceptance for 20-year-olds.

Interestingly, in terms of region-specific accuracy, the domain-focused NLP chatbot (Athena) we developed using traditional NLP methods outperformed the LLMs. This indicates that while LLMs offer considerable potential for general tasks, their utility in highly specialized, context-heavy domains like autism care may be limited without additional context or training on domain-specific data. The LLMs often returned less accurate results when asked for specific services that required deep local knowledge, especially when compared to our earlier system, which was trained on curated, location-specific datasets.

Domain-Specific LLM-based Chatbot Testing

The customized GPT-4-based chatbot, Athena, demonstrated substantial improvements over both traditional NLP and out-of-the-box LLMs (Table 9, 10).

Table 9 Performance Comparison: NLP vs. LLMs

Query	NLP System Result	LLM Result	Observation
"Find a provider for ABA therapy in Birmingham"	Partial match; misses some relevant providers	Comprehensive match with all relevant providers	NLP struggled with synonyms, LLM recognized intent
"Looking for speech therapy near Montgomery"	No relevant results	Accurate and relevant results	NLP failed to match non-exact phrasing, LLM succeeded
"Are there providers for OT in Huntsville?"	Accurate but limited	Accurate and more comprehensive	NLP provided limited results, LLM covered more options
"Who accepts Medicaid for ABA in Mobile?"	Incorrect match (focused on location only)	Correct match with relevant insurance filter	NLP failed to account for insurance filter, LLM did well

Table 10 Accuracy Metrics for Different Models

Sample Query	NLP Accuracy (%)	Out-of-the-Box LLM Accuracy (%)	Domain-Specific LLM Accuracy (%)
Find an Occupational Therapist and Speech Therapist in Huntsville, AL who accepts Medicaid.	66.7	44.9	91.2
Locate a Speech Therapist in Birmingham specializing in early intervention.	72.1	38.4	93.7
Search for a Pediatric Neurologist in Montgomery offering telehealth services.	61.3	56.5	89.1

More granular comparison of these models by comparing F1 scores across different scenarios also shows domain specific LLM-based chatbot outperforms the NLP based (Table 11).

Table 11 Scenario-Specific Performance Across NLP and Athena

Criteria	Traditional NLP	Domain-Specific LLM
Simple Queries	79.4	94.6
Complex Queries	51.2	89.5
Queries with Varied Phrasing	42.7	86.3
Geographically Specific Queries	58.9	87.4
Queries for Less Common Services	23.5	81.7

Although the numbers reflect qualitative assessments rather than precise, scientifically measured values, these results indicate that while traditional NLP systems generally struggled with varied phrasing and complex queries, out-of-the-box LLMs showed improvement but still lacked the precision and recall necessary for highly specialized autism care services. In contrast, using domain-specific tuning of GPT-4 consistently outperformed both methods which highlights a clear improvement in relevance and response accuracy, even if the evaluations are based on observed trends rather than rigid statistical metrics.

Performance of RAG-based chatbot

While the GPT-4-powered Athena chatbot demonstrated significant improvements in handling nuanced queries and domain-specific autism care services, it relied on a custom version of OpenAI's ChatGPT with limited control over model parameters such as temperature, token limits, and conversation context (OpenAI, 2023). To enhance functionality and precision, we implemented a RAG approach to address the shortcomings of traditional NLP and out-of-the-box LLMs by integrating large language models with a dynamic retrieval mechanism from our curated dataset of autism service providers (Gao et al., 2023). This provides greater control over model parameters, enabling fine-tuning of responses to meet user expectations. By integrating real-time

retrieval from our curated dataset, the chatbot provides contextually relevant and up-to-date information. Responses are based on specific retrieved documents, allowing us to trace the source of information, which is crucial in healthcare applications for building user trust. Comparative performance observation of these models (Table 12) shows RAG enhanced semantic matching, enabling precise recommendations even for complex, multi-criteria queries. In addition to these qualitative comparisons, RAG offers superior control over system performance metrics.

Table 12 Comparative Observations: RAG-Based vs. GPT-Based Athena

Sample Query	Athena GPT-4 Result	RAG-Based Athena Result	Observation
Find an Occupational Therapist and Speech Therapist in Huntsville, AL who accepts Medicaid.	Partial match	Accurate and specific match	RAG retrieved more precise data on providers and services.
Locate a Speech Therapist in Birmingham specializing in early intervention.	General match	Specific match	RAG incorporated specific local data not available to GPT-4 alone.
Search for a Pediatric Neurologist in Montgomery offering telehealth services.	Incomplete match	Full match	RAG handled multiple criteria more effectively.

While GPT-4-based system performed well in complex queries, RAG's structured retrieval mechanism resulted in even higher accuracy, precision, and recall (Table 13).

Table 13 - Precision and Recall of RAG-Based vs. GPT-Based

Criteria	GPT-4-Based Precision (%)	RAG-Based Precision (%)	GPT-4-Based Recall (%)	RAG-Based Recall (%)
Simple Queries	91.3	96.2	87.5	97.2
Complex Queries	87.1	94.3	83.3	95.5
Geographically Specific Queries	84.6	92.7	79.2	93.1

In comparison to the custom GPT-4-based system, which was reliant on predefined knowledge, this RAG-based architecture provides Enhanced Control, Improved Accuracy, Transparency, and

Explainability. With RAG, we can adjust critical generation parameters such as temperature (for controlling the randomness of output), token limit (for managing the length of responses), and the number of conversation turns kept in context (Surla 2023). The ability to incorporate real-time retrieval from our own curated dataset means that the RAG-based model is not limited by the pre-trained knowledge of an LLM. It can dynamically pull the most relevant information from a vast repository of autism service data, ensuring that responses are not only contextually relevant but also accurate and up-to-date. With RAG, every response is based on retrieved documents, allowing us to trace the origin of the information. This transparency is a key differentiator when compared to Athena's GPT-4-based responses, where the sources of information were more opaque. For sensitive healthcare domains like autism services, the ability to verify and explain responses is crucial for user trust.

Conclusion and Discussion

This study systematically explored the potential of AI-driven tools for improving autism-related healthcare resource management in Alabama. By developing and evaluating several chatbot models—including traditional NLP methods, out-of-the-box large language models (LLMs), a customized domain-specific GPT-4 chatbot, and a Retrieval-Augmented Generation (RAG) chatbot—we were able to clearly demonstrate the feasibility and comparative strengths of different AI approaches in this specialized social work context.

Among the systems evaluated, traditional NLP-based methods showed effectiveness for simple, clearly structured queries, achieving up to 87.5% precision and 70% recall. However, these models were limited by their reliance on exact keyword matching and struggled to handle complex, multi-criteria queries and linguistic variations. Out-of-the-box LLMs, such as ChatGPT, Google Gemini,

Microsoft Copilot, and Claude, exhibited strong capabilities in understanding nuanced, natural language queries but faced significant limitations in terms of geographic specificity, consistently returning generic and less actionable recommendations due to the absence of localized, domain-specific training. In contrast, our domain-specific GPT-4 chatbot (Minerva) substantially improved upon these approaches by effectively leveraging structured local datasets, clearly highlighting the importance of domain-specific fine-tuning. This model achieved significantly higher accuracy (approximately 85–95%) across diverse query types, demonstrating its utility in real-world scenarios faced by social workers and caregivers. Finally, the Retrieval-Augmented Generation (RAG) chatbot emerged as the strongest model, integrating real-time retrieval of relevant provider information with generative AI capabilities. It provided enhanced accuracy (approximately 90–96% precision and recall), improved transparency, and higher explainability of recommendations. These characteristics are particularly valuable for sensitive healthcare applications where trust and interpretability are critical.

The primary limitation of this study stems from its proof-of-concept nature, relying on exploratory evaluations rather than extensive computational benchmarks or statistical validation. Additionally, the practical deployment of these AI tools faces inherent challenges, including the need for continuous updating of service provider information, potential computational costs associated with maintaining vector embeddings, and ethical considerations surrounding algorithmic biases and data privacy. Our results clearly underscore the promise and practicality of tailored AI tools—particularly domain-specific GPT-4 and RAG-based systems—for enhancing social work practices related to autism care. However, successful real-world implementation will require thoughtful integration strategies, regular data management practices, and attention to ethical and operational considerations.

~~During the development of our provider database, we also o~~Observed marked disparities in the geographic and socioeconomic distribution of autism-related services across Alabama ~~were observed (Figure 3). These disparities, while not the central focus of this study, illustrate reinforce the urgent-need for intelligent assistive tools that can help families and social workers identify appropriate resources despite systemic access gaps, when identified. -This finding offers important motivation for continued development and future investigation.~~

While the findings of this study demonstrate the potential of AI-driven tools in autism care, several important ethical and regulatory considerations must be addressed to ensure responsible deployment and real-world impact. One primary concern is algorithmic bias. AI models, particularly large language models (LLMs), risk reinforcing existing disparities if trained on data that inadequately represents marginalized populations. This is particularly relevant in autism services, where racial and socioeconomic disparities are well-documented (Aylward et al., 2021). LLMs have also shown susceptibility to inheriting biases present in their training data, potentially impacting equitable service recommendations (Ayoub et al., 2024). Data privacy and security are equally critical, especially when handling sensitive healthcare information. Compliance with regulations like the Health Insurance Portability and Accountability Act (HIPAA) is essential to protect personally identifiable information. Although this study utilized mostly publicly and non-sensitive available service provider data, future implementations involving patient data must incorporate privacy-preserving techniques such as de-identification and encryption (Ghadban et al., 2023). From a regulatory standpoint, ensuring model transparency, explainability, and human-in-the-loop review processes will be necessary for ethical compliance and user trust (Gallegos et al., 2024). Moreover, continuous auditing and bias assessment should be embedded into system maintenance to mitigate risks of harm or unequal service delivery.

While as a pilot, this study demonstrates the potential framework for an AI-driven system to support autism care resource management, practical implementation would require further development to ensure adaptability to changes in available services and usability by intended stakeholders. Future iterations should incorporate mechanisms for periodic database updates to reflect changes in provider availability, service offerings, and insurance coverage. This could involve a combination of automated data collection, manual verification, and partnerships with state agencies to maintain accuracy over time (Eriskin, Karatas, and Zheng, 2024). The system is proposed to be designed to support accessibility through a web-based interface capable of processing natural language queries. Social workers, families, and caregivers represent primary user groups. Targeted training materials, including user guides and interactive onboarding, would be necessary to ensure effective engagement, particularly for non-technical users. Prior studies emphasize that such training is essential to maximize the utility of AI tools in social work and healthcare environments (Åhs, Mozelius, and Espvall, 2022). While this study emphasized natural language interfaces to assess the feasibility of LLM-powered semantic search and to be used by users with high digital literacy. However, ~~we recognize that~~ traditional UI designs (e.g., dropdowns, filters) may offer complementary benefits in terms of simplicity or accessibility for some user groups. Future work ~~should~~could compare chatbot-driven interfaces with conventional UI approaches to assess their relative performance, user preferences, and usability across diverse populations. Finally, incorporating user feedback mechanisms can enable continuous refinement of recommendations and system performance. Feedback loops are increasingly recognized as a best practice in AI system design to enhance accuracy, relevance, and user trust (Pagan et al., 2023).

Together, these considerations emphasize that while AI-driven tools hold significant promise for improving autism care resource navigation, their real-world deployment will require careful attention to ethical, regulatory, and practical implementation factors. This proof-of-concept study provides a foundation for future research and development aimed at operationalizing such tools within social work and healthcare systems.

Supplementary data:

Table S1: Stakeholder Roles and Service Mapping

References

Adobe. OCR PDF | How Tos | PDF Services API | Adobe PDF Services [Internet]. 2024 [cited 2024 Jun 17]. Available from: <https://developer.adobe.com/document-services/docs/overview/pdf-services-api/howtos/ocr-pdf/>

Åhs F, Mozelius P, Espvall M. Preparing psychologists and social workers for the daily use of AI. *Int Conf AI Res*. 2022;4(1):1–5. doi:10.34190/icair.4.1.771

Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ*. 2023;23(1):689. doi:10.1186/s12909-023-04698-z

Aylward BS, Gal-Szabo DE, Taraman S. Racial, ethnic, and sociodemographic disparities in diagnosis of children with autism spectrum disorder. *J Dev Behav Pediatr*. 2021;42(8):682–9. doi:10.1097/DBP.0000000000000996

Ayoub NF, Balakrishnan K, Ayoub MS, Barrett TF, David AP, Gray ST. Inherent bias in large language models: a random sampling analysis. *Mayo Clin Proc Digit Health*. 2024;2(2):186–91. doi:10.1016/j.mcpdig.2024.03.003

Berge GT, Granmo O-C, Tveit TO, Ruthjersen AL, Sharma J. Combining unsupervised, supervised and rule-based learning: the case of detecting patient allergies in electronic health records. *BMC Med Inform Decis Mak*. 2023;23:188. doi:10.1186/s12911-023-02271-8

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Formatted: Indent: Left: 0.25"

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Bishop-Fitzpatrick L, Dababnah S, Baker-Ericzén MJ, Smith MJ, Magaña SM. Autism spectrum disorder and the science of social work: a grand challenge for social work research. Soc Work Ment Health. 2019;17(1):73–92. doi:10.1080/15332985.2018.1509411

[Shaw, K. A., Williams, S., Patrick, M. E., et al. \(2025\). Prevalence and Early Identification of Autism Spectrum Disorder Among Children Aged 4 and 8 Years — Autism and Developmental Disabilities Monitoring Network, 16 Sites, United States, 2022. MMWR Surveillance Summaries, 74\(SS-2\), 1–22. <https://doi.org/10.15585/mmwr.ss7402a1> Centers for Disease Control and Prevention \(CDC\). Community report on autism, 2023: Autism and Developmental Disabilities Monitoring \(ADM\) Network \[Internet\]. 2023 \[cited 2024 Jun 17\]. Available from: <https://stacks.ede.gov/view/ede/126223>](#)

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Cheak-Zamora NC, Teti M, First J. 'Transitions are scary for our kids, and they're scary for us': Family member and youth perspectives on the challenges of transitioning to adulthood with autism. J Appl Res Intellect Disabil. 2015 Nov;28(6):548–60. doi:10.1111/jar.12150

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

[Chen, Q., Hu, Y., Peng, X. et al. Benchmarking large language models for biomedical natural language processing applications and recommendations. Nat Commun 16, 3280 \(2025\). <https://doi.org/10.1038/s41467-025-56989-2> Chen Q, Du J, Hu Y, Kuttichi Keloth V, Peng X, Raja K, et al. Large language models in biomedical natural language processing: benchmarks, baselines, and recommendations. arXiv \[Preprint\]. 2024. doi:10.48550/arXiv.2305.16326](#)

Chen X, Zhao L, Liang H, Lai KK. Matching patients and healthcare service providers: a novel two-stage method based on knowledge rules and OWA-NSGA-II algorithm. J Comb Optim. 2019;37(1):221–47. doi:10.1007/s10878-017-0221-2

Formatted: Normal, Indent: Left: 0.25", No bullets or numbering

[Christoph Gohlke. RapidFuzz: Fuzzy string matching in Python. \[Internet\]. 2024 \[cited 2024 Jul 30\]. Available from: <https://github.com/maxbachmann/RapidFuzz>](#)

Formatted: Indent: Left: 0.25"

Jagtap DS. Cross-media topic modeling: a comparative analysis of NMF, LDA and CoRex combined with BERT on Quora and Twitter. In: 2024 International Conference on Computing and Data Science (ICCDs). 2024. p. 1–6.
doi:10.1109/ICCDs60734.2024.10560395

De Vito G. Assessing healthcare software built using IoT and LLM technologies. In: Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering. New York, NY: Association for Computing Machinery; 2024. p. 476–81.
doi:10.1145/3661167.3661202

Dong H, Suárez-Paniagua V, Zhang H, Wang M, Casey A, Davidson E, et al. Ontology-driven and weakly supervised rare disease identification from clinical notes. BMC Med Inform Decis Mak. 2023;23(1):86. doi:10.1186/s12911-023-02181-9

Eriskin L, Karatas M, Zheng Y-J. A robust multi-objective model for healthcare resource management and location planning during pandemics. Ann Oper Res. 2024;335(3):1471–518. doi:10.1007/s10479-022-04760-x

Friesdorf M. Automation at scale: the benefits for payers [Internet]. McKinsey. 2019 [cited 2024 Jun 17]. Available from: <https://www.mckinsey.com/industries/healthcare/our-insights/automation-at-scale-the-benefits-for-payers>

Fulceri F, Gila L, Caruso A, Micai M, Romano G, Scattoni ML. Building bricks of integrated care pathway for autism spectrum disorder: a systematic review. *Int J Mol Sci*. 2023;24(7):6222. doi:10.3390/ijms24076222

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Formatted: Normal, Indent: Left: 0.25", No bullets or numbering

[Folium: Python Data. Leaflet.js Maps \[Internet\]. 2024 \[cited 2024 Jul 30\]. Available from: https://python-visualization.github.io/folium/](https://python-visualization.github.io/folium/)

Formatted: Indent: Left: 0.25"

[Isabel O. Gallegos, Ryan A. Rossi, Joe Barrow, Md Mehrab Tanjim, Sungchul Kim, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, Nesreen K. Ahmed; Bias and Fairness in Large Language Models: A Survey. Computational Linguistics 2024; 50 \(3\): 1097–1179. doi: https://doi.org/10.1162/coli_a_00524](https://doi.org/10.1162/coli_a_00524)
~~Gallegos IO, Rossi RA, Barrow J, Tanjim MM, Kim S, Dernoncourt F, et al. Bias and fairness in large language models: a survey. arXiv [Preprint]. 2024. doi:10.48550/arXiv.2309.00770~~

Gandhi TK, Classen D, Sinsky CA, Rhew DC, Vande Garde N, Roberts A, et al. How can artificial intelligence decrease cognitive and work burden for front line practitioners? *JAMIA Open*. 2023;6(3):ooad079. doi:10.1093/jamiaopen/ooad079

Gao T, Fisch A, Chen D. Making pre-trained language models better few-shot learners. In: Zong C, Xia F, Li W, Navigli R, editors. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics; 2021. p. 3816–30. doi:10.18653/v1/2021.acl-long.295

Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, et al. Retrieval-augmented generation for large language models: a survey. arXiv [Preprint]. 2023 Dec 18. Available from: <https://arxiv.org/abs/2312.10997v5>.

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

García-Méndez S, de Arriba-Pérez F. Large language models and healthcare alliance: potential and challenges of two representative use cases. Ann Biomed Eng. 2024;52(8):1928–31. doi:10.1007/s10439-024-03454-8

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Ghadban YA, Lu H, Adavi U, Sharma A, Gara S, Das N, et al. Transforming healthcare education: harnessing large language models for frontline health worker capacity building using retrieval-augmented generation. medRxiv [Preprint]. 2023. doi:10.1101/2023.12.15.23300009

Griese L, Berens E-M, Nowak P, Pelikan JM, Schaeffer D. Challenges in navigating the health care system: development of an instrument measuring navigation health literacy. Int J Environ Res Public Health. 2020;17(16):5731. doi:10.3390/ijerph17165731

Formatted: Normal, Indent: Left: 0.25", No bullets or numbering

[Grinberg M. Flask Web Development. 1st ed. Sebastopol, CA: O'Reilly Media; 2014. ISBN: 978-1-449-37262-0](#)

Formatted: Indent: Left: 0.25"

Gupta P. Neural information extraction from natural language text [dissertation]. Munich: Ludwig-Maximilians-Universität München; 2019. Available from: <https://edoc.ub.uni-muenchen.de/24953/>.

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Herd P, Moynihan D. Health care administrative burdens: centering patient experiences. Health Serv Res. 2021;56(5):751–4. doi:10.1111/1475-6773.13858

Hervieux S, Wheatley A. The rise of AI: implications and applications of artificial intelligence in academic libraries. ACRL Publications in Librarianship; No. 78. Chicago, IL: Association of College and Research Libraries; 2022.

Honnibal, Montani. spaCy 2: natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing [Internet]. Sentometrics Research; 2017 [cited 2024 Jun 17]. Available from: <https://sentometrics-research.com/publication/72/>

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Hossain E, Rana R, Higgins N, Soar J, Barua PD, Pisani AR, et al. Natural language processing in electronic health records in relation to healthcare decision-making: a systematic review. Comput Biol Med. 2023;155:106649. doi:10.1016/j.combiomed.2023.106649

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Hua Y, Blackley S, Shinn A, Skinner J, Moran L, Zhou L. Identifying psychosis episodes in psychiatric admission notes via rule-based methods, machine learning, and pre-trained language models [Preprint]. 2024. doi:10.21203/rs.3.rs-4126574/v1

Jha J, Almagro M, Tissot H. Designing NLP applications to support ICD coding: an impact analysis and guidelines to enhance baseline performance when processing patient discharge notes. J Digit Health. 2023;63–81. doi:10.55976/jdh.22023119463-81

Karlin J. Sam's story: the financial and human costs of disjointed logics of care [Internet]. PMC. 2022 [cited 2024 Jun 17]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8786420/>

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Ke Y, Jin L, Elangovan K, Abdullah HR, Liu N, Sia ATH, et al. Development and testing of retrieval augmented generation in large language models -- a case study report. arXiv [Preprint]. 2024. doi:10.48550/arXiv.2402.01733

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Kim E, Jeong Y, Choi MS. MediBioDeBERTa: biomedical language model with continuous learning and intermediate fine-tuning. *IEEE Access*. 2023;11:141036–44. doi:10.1109/ACCESS.2023.3341612

Kitsios F, Kamariotou M, Syngelakis AI, Talias MA. Recent advances of artificial intelligence in healthcare: a systematic literature review. *Appl Sci*. 2023;13(13):7479. doi:10.3390/app13137479

Koum DCK, Enyama D, Kojom Foko LP, Noukeu D, Nguedia B, Mbono R, et al. Diagnosis and treatment challenges of autism spectrum disorder at a reference hospital in Douala, Cameroon. *BMC Pediatr*. 2023;23(1):459. doi:10.1186/s12887-023-04242-4

Krešević S, Giuffrè M, Ajčević M, Accardo A, Crocè LS, Shung DL. Optimization of hepatological clinical guidelines interpretation by large language models: a retrieval augmented generation-based framework. *NPJ Digit Med*. 2024;7(1):1–9. doi:10.1038/s41746-024-01091-y

Kreuter MW, Thompson T, McQueen A, Garg R. Addressing social needs in health care settings: evidence, challenges, and opportunities for public health. *Annu Rev Public Health*. 2021;42:329–44. doi:10.1146/annurev-publhealth-090419-102204

Kumar M, Dubey K, Pandey R. Evolution of emerging computing paradigm cloud to fog: applications, limitations and research challenges. In: 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). 2021. p. 257–61. doi:10.1109/Confluence51648.2021.9377050

Kurpas D, Stefanicka-Wojtas D, Shpakou A, Halata D, Mohos A, Skarbaliene A, et al. The advantages and disadvantages of integrated care implementation in Central and Eastern Europe – perspective from 9 CEE countries. *Int J Integr Care*. 2021;21(4):14. doi:10.5334/ijic.5632

[Lee B. Challenges facing case management in a changing social service environment. *Care Manag J*. 2001 Fall;3\(1\):20-4. PMID: 11974573.](#)
~~Lee B. Challenges facing case management in a changing social service environment. *Care Manag J*. 2001;3(1):20-4.~~

Li M, Kilicoglu H, Xu H, Zhang R. BiomedRAG: a retrieval augmented large language model for biomedicine. [Journal of Biomedical Informatics, 2025arXiv \[Preprint\]. 2024. https://doi.org/10.1016/j.jbi.2024.104769 doi:10.48550/arXiv.2405.00465](#)

Maenner MJ. Prevalence and characteristics of autism spectrum disorder among children aged 8 years — Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. *MMWR Surveill Summ*. 2023;72. doi:10.15585/mmwr.ss7202a1

Mah PM. Analysis of artificial intelligence and natural language processing significance as expert systems support for e-health using pre-train deep learning models. 2022. doi:10.56578/ataiml010201

~~4~~ Magaña S, Parish SL, Rose RA, Timberlake M, Swaine JG. Racial and ethnic disparities in quality of health care among children with autism and other developmental disabilities. *Intellect Dev Disabil*. 2012 Aug;50(4):287–99. doi:10.1352/1934-9556-50.4.287

Formatted: Normal, Indent: Left: 0.25", No bullets or numbering

[McKinney W. Data Structures for Statistical Computing in Python. In: van der Walt S, Millman J, editors. Proceedings of the 9th Python in Science Conference; 2010. p. 51–6. doi:10.25080/Majora-92bf1922-00a](#)

Formatted: Indent: Left: 0.25"

McRae MP, Simmons GW, Christodoulides NJ, Lu Z, Kang SK, Fenyo D, et al. Clinical decision support tool and rapid point-of-care platform for determining disease severity in patients with COVID-19. *Lab Chip*. 2020;20(12):2075–85. doi:10.1039/d0lc00373e

Minaee S, Mikolov T, Nikzad N, Chenaghlu M, Socher R, Amatriain X, Gao J. Large language models: a survey. *arXiv [Preprint]*. 2024. doi:10.48550/arXiv.2402.06196

Morozov SP, Vladzimirskiy AV, Gombolevskiy VA, Kuz'mina ES, Ledikhova NV. Artificial intelligence: natural language processing for peer-review in radiology. *J Radiol Nucl Med*. 2018;99(5):253–8. doi:10.20862/0042-4676-2018-99-5-253-258

Olujimi PA, Ade-Ibijola A. NLP techniques for automating responses to customer queries: a systematic review. *Discov Artif Intell*. 2023;3(1):20. doi:10.1007/s44163-023-00065-5

[OpenAI. Creating a GPT. OpenAI Help Center. 2024 \[cited 2024 Jun 17\]. Available from: https://help.openai.com/en/articles/8554397-creating-a-gpt](#)
[OpenAI. Introducing GPTs \[Internet\]. 2023 \[cited 2024 Jun 17\]. Available from: https://openai.com/index/introducing-gpts/](#)

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

[OpenAI. GPT-4 Technical Report. OpenAI Research. 2023 Mar 27 \[cited 2024 Jun 17\]. Available from: https://cdn.openai.com/papers/gpt-4.pdf](#)
[OpenAI. GPT-4 \[Internet\]. 2024 \[cited 2024 Jun 17\]. Available from: https://openai.com/index/gpt-4/](#)

Formatted: Normal, Indent: Left: 0.25", No bullets or numbering

[OpenStreetMap Foundation. Nominatim: Geocoding API for OpenStreetMap \[Internet\]. 2024](#)

[\[cited 2024 Jul 30\]. Available from: https://nominatim.openstreetmap.org/](#)

Pagan N, Baumann J, Elokda E, De Pasquale G, Bolognani S, Hannák A. A classification of feedback loops and their relation to biases in automated decision-making systems. In: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23); 2023. doi:10.1145/3617694.3623227

Polak MP, Morgan D. Extracting accurate materials data from research papers with conversational language models and prompt engineering. Nat Commun. 2024;15(1):1569. doi:10.1038/s41467-024-45914-8

Pouplin T, Sun H, Holt S, van der Schaar M. Retrieval augmented thought process for private data handling in healthcare. arXiv [Preprint]. 2024. doi:10.48550/arXiv.2402.07812

Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. [arXiv \[Preprint\]. International Journal of Machine Learning Technologies. 2229-3981. 20202011](#). doi: [10.9735/2229-398110.48550/arXiv.2010.16061](#)

Radu M, Moldovan R, Băban A. Families with complex needs: an inside perspective from young people, their carers, and healthcare providers. J Community Genet. 2022;13(3):293–302. doi:10.1007/s12687-022-00586-z

Ramados G. Social determinants of health: integrating analytics and technology in healthcare systems, IJSR, call for papers, online journal. Int J Sci Res (IJSR). 2024. Available from: <https://www.ijsr.net/>

Formatted: Indent: Left: 0.25"

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Renghea A, Cuevas-Budhart MA, Yébenes-Revuelto H, Gómez Del Pulgar M, Iglesias-López MT. ‘Comprehensive care’ concept in nursing: systematic review. *Investig Educ Enferm*. 2022;40(3):e05. doi:10.17533/udea.ice.v40n3e05

Sadeghi-Tehran P, Angelov P, Virlet N, Hawkesford MJ. Scalable database indexing and fast image retrieval based on deep learning and hierarchically nested structure applied to remote sensing and plant biology. *J Imaging*. 2019;5(3):33. doi:10.3390/jimaging5030033

Schork NJ. Artificial intelligence and personalized medicine. *Cancer Treat Res*. 2019;178:265–83. doi:10.1007/978-3-030-16391-4_11

Shah WS, Elkhwesky Z, Jasim KM, Elkhwesky EFY, Elkhwesky FFY. Artificial intelligence in healthcare services: past, present and future research directions. *Rev Manag Sci*. 2024;18(3):941–63. doi:10.1007/s11846-023-00699-w

Shaw KA. Early identification of autism spectrum disorder among children aged 4 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. *MMWR Surveill Summ*. 2023;72. doi:10.15585/mmwr.ss7201a1

Siriwardhana S, Weerasekera R, Wen E, Kaluarachchi T, Rana R, Nanayakkara S. Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering. *Trans Assoc Comput Linguist*. 2023;11:1–17. doi:10.1162/tacl_a_00530

[SQLPad. How to convert strings to lowercase in R \[Internet\]. 2024 \[cited 2024 Jun 17\].](#)

[Available from: https://sqlpad.io/tutorial/convert-strings-lowercase/](https://sqlpad.io/tutorial/convert-strings-lowercase/)

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Stewart M, Hodkiewicz M, Li S. Large language models for failure mode classification: an investigation. arXiv [Preprint]. 2023. doi:10.48550/arXiv.2309.08181

[Surla A. How to Get Better Outputs from Your Large Language Model. NVIDIA Developer Technical Blog. Jun 14, 2023 \[cited 2024 Jun 17\]. Available from:
https://developer.nvidia.com/blog/how-to-get-better-outputs-from-your-large-language-model/Surla-How-to-get-better-outputs-from-your-large-language-model \[Internet\]. NVIDIA Technical Blog; 2023 Jun 14 \[cited 2024 Jun 17\]. Available from:
https://developer.nvidia.com/blog/how-to-get-better-outputs-from-your-large-language-model/.](#)

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Tao S, Yao L, Ding H, Xie Y, Cao Q, Sun F, et al. When to trust LLMs: aligning confidence with response quality. arXiv [Preprint]. 2024. doi:10.48550/arXiv.2404.17287

US Census Bureau. Census Bureau data [Internet]. 2024 [cited 2024 Jun 17]. Available from: <https://data.census.gov/>.

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Formatted: Font: (Default) +Headings CS (Times New Roman), 12 pt

Uysal AK, Gunal S. The impact of preprocessing on text classification. Inf Process Manag. 2014;50(1):104–12. doi:10.1016/j.ipm.2013.08.006

Van Veen D, Van Uden C, Blankemeier L, Delbrouck J-B, Aali A, Bluethgen C, et al. Adapted large language models can outperform medical experts in clinical text summarization. Nat Med. 2024;30(4):1134–42. doi:10.1038/s41591-024-02855-5

Warfield ME, Crossman MK, Delahaye J, Der Weerd E, Kuhlthau KA. Physician perspectives on providing primary medical care to adults with autism spectrum disorders (ASD). J Autism Dev Disord. 2015 Jul;45(7):2209–17. doi:10.1007/s10803-015-2386-9

Wu H, Lu X, Wang H. The application of artificial intelligence in health care resource allocation before and during the COVID-19 pandemic: scoping review. *JMIR AI*. 2023;2(1):e38397. doi:10.2196/38397

Yeo GTS, Kannan P, Lee ES, Smith HE. Community case managers' challenges collaborating with primary care when managing complex patients in the community: a qualitative study in Singapore. *Health Soc Care Community*. 2022;30(4):1568–77. doi:10.1111/hsc.13489

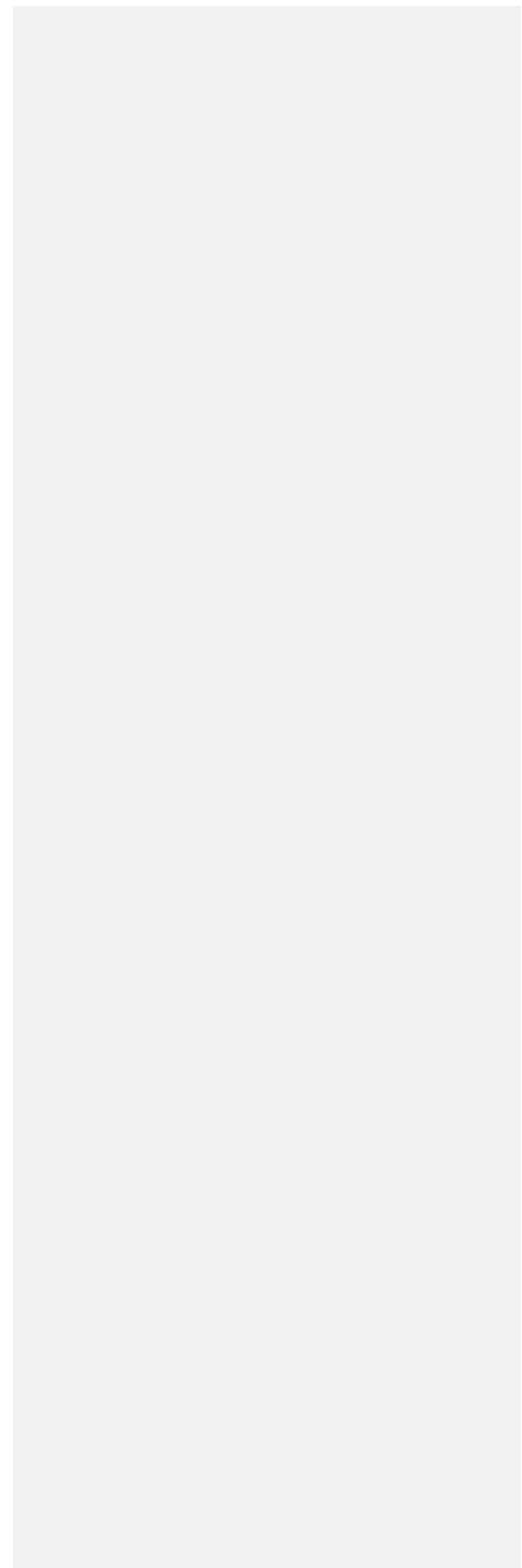
Zhang K, Jiang X. Sensitive data detection with high-throughput machine learning models in electrical health records. *arXiv [Preprint]*. 2023. doi:10.48550/arXiv.2305.03169

Zhang R, Xing Y, Wang J, Shang X, Zhu X. A novel multiattribute decision-making method based on point–Choquet aggregation operators and its application in supporting the hierarchical medical treatment system in China. *Int J Environ Res Public Health*. 2018;15(8):1718. doi:10.3390/ijerph15081718

Zuchowski I, McLennan S. A systematic review of social work in general practice: opportunities and challenges. *J Evid Based Soc Work*. 2023;20(5):686–726. doi:10.1080/26408066.2023.2202665

Zuckerman KE, Lindly OJ, Sinche BK. Parental concerns, provider response, and timeliness of autism spectrum disorder diagnosis. *J Pediatr*. 2015 Jun;166(6):1431–9.e1. doi:10.1016/j.jpeds.2015.03.007

|



Supplementary data:

Table S1: Stakeholder Roles and Service Mapping

Organization Type	Primary Function	Primary Provider	Secondary Provider(s)	Primary Beneficiary
Art Therapist	Art-based skill development	Art therapist	Family, Various therapists	Child with autism
Autism Advocacy Organizations	Advocacy and policy making	Advocacy workers	Volunteers, Lawyers	Autism community
Autism Research Centers	Research and development	Researchers	Technicians, Volunteers	Autism community
Autism Support Group	Peer support and resources	Group coordinator	Families, Therapists	Child with autism and family
Behavioral Therapist	Behavioral assessment and intervention	Behavioral therapist	Family, Various therapists	Child with autism
Clinical Psychologist	Assessment and treatment	Clinical psychologist	Family, Therapists	Child with autism
Early Intervention Services	Early childhood intervention	Early intervention specialists	Family, Various therapists	Child with autism
Family	Support and care	Parents or siblings	None	Child with autism
Government Health Agencies	Policy making and funding	Government officials	Various health professionals	Autism community
Government Health Department	Health policy making and implementation	Government health officials	Various health professionals	Autism community
Insurance Companies	Coverage of treatments and services	Insurance agents	None	Family with autistic child
Insurance Companies	Insurance policy making and implementation	Insurance policy makers	None	Family with autistic child
Legal Services	Legal support and advice	Lawyers	None	Family with autistic child
Local Community Centers	Community support and resources	Community center staff	Volunteers, Various therapists	Autism community
Local Government	Local policy making and implementation	Local government officials	Local health and education professionals	Autism community
Music Therapist	Music-based skill development	Music therapist	Family, Various therapists	Child with autism
National Government	National policy making and implementation	National government officials	Various health and education professionals	Autism community

Formatted: Justified, Space After: 8 pt, Line spacing: Multiple 1.15 li, Don't keep with next, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Neurologist	Medical assessment and treatment	Neurologist	Family, Therapists	Child with autism
Nutritionists	Diet and nutrition advice	Nutritionist	None	Child with autism
Occupational Therapist	Skill development	Occupational therapist	Family, Clinical psychologist	Child with autism
Pediatrician	General health care	Pediatrician	Family, Various therapists	Child with autism
Pharmacies	Medication supply	Pharmacists	None	Child with autism
Physical Therapist	Physical skill development	Physical therapist	Family, Various therapists	Child with autism
Public Health Departments	Public health services and policy making	Public health officials	Various health professionals	Autism community
Residential Service Providers	Residential care and support	Residential staff	Support staff, Therapists	Adult with autism
School Administration	School policy making and implementation	School administrators	Teachers, Therapists	Child with autism
Social Worker	Case management and support	Social worker	Various therapists, Family	Child with autism and family
Special Education School	Education and support	Special education teachers	Teaching assistants, Therapists	Child with autism
Speech Therapist	Language and communication skills development	Speech therapist	Family, Clinical psychologist, Occupational therapist	Child with autism
Technology Solution Providers	Technological aids provision	Tech developers	Support staff	Child with autism
Vocational Training Providers	Vocational skills training	Vocational trainers	Support staff	Adult with autism

Reviewer 3:

GENERAL

1. **The tool's focus on social workers is understandable, but why were social workers specifically chosen for input? Were other key stakeholders (e.g., such as physicians, school teachers, or caregivers) consulted or considered?**

Authors Response:

We appreciate the reviewer's attention to the intended user base of the tool. Social workers were selected as the primary end users for several reasons.

First, the core aim of this project was to develop a proof-of-concept for a social worker assistant tool, a system designed specifically to support the work of social workers in navigating autism-related resources. The responsibilities targeted by the tool, such as coordinating services, managing referrals, understanding eligibility criteria, and responding to real-world family needs, fall squarely within the scope of social work.

Second, our project was conducted in close partnership with the Alabama Department of Mental Health, where the tool was conceptualized as a practical solution to support their front-line social service staff. This collaboration directly informed the system requirements, query types, and resource datasets, ensuring that the prototype would reflect actual workflows and decision needs encountered by social workers.

While we acknowledge that caregivers, physicians, educators, and others play critical roles in autism care, they were outside the primary scope of this initial deployment. We agree that future iterations of the tool could be extended to support those broader stakeholder groups, and we have added a note in the manuscript to reflect this possible future expansion. (revision is added to page 5)

Reviewer 3:

2. **The user-friendly interface and use of AI/chatbots are timely and engaging; however, how does this approach compare to more traditional UI elements like dropdown menus for filtering? Although the manuscript acknowledges this on page 21, no quantitative performance comparisons are provided. Including such analysis would add rigor.**

Authors' Response:

We agree that comparing the chatbot interface with traditional UI mechanisms such as dropdown menus or rule-based filters could add valuable insight. However, this study was designed as a proof-of-concept to test the feasibility of using AI, particularly NLP and LLM-based approaches, for supporting natural language queries in autism care navigation. As such, our focus was on evaluating the semantic robustness and contextual responsiveness of the AI-driven system, rather than conducting a formal UI/UX comparison across interface types. That said, we acknowledge that traditional UI elements like dropdown filters may be more appropriate in certain contexts, especially for users with structured needs or lower digital literacy. We have updated the manuscript to acknowledge this explicitly and framed it as a future area of research. We appreciate the reviewer for raising this opportunity to extend the work. (revision is added to page 29)

Reviewer 3:

3. **The manuscript lacks statistical analyses and relies solely on observed trends. Incorporating relevant statistics would enhance the validity of the findings.**

Authors' Response:

We appreciate this observation. As the reviewer notes, the manuscript currently relies on descriptive performance metrics (e.g., precision, recall, F1 scores) and observed trends across different AI models. This was a deliberate design choice, as the primary aim of the study was to develop and demonstrate a proof-of-concept system for applying AI (particularly LLMs and RAG framework) to the specific challenge of autism-related service navigation. Given the exploratory nature of this project and the emphasis on technical feasibility, we did not conduct formal statistical significance testing (e.g., t-tests or McNemar's tests). However, we agree that a rigorous statistical evaluation would be in principle desirable to quantify the performance claims in a production-ready deployment. We have updated the manuscript to acknowledge this limitation more explicitly and flagged it as a critical area for future work. (revision is added to page 17)

Reviewer 3:

4. **Citations are inconsistent for programs, software, and resources. (e.g., spaCy is cited, but RapidFuzz is not). Please ensure all tools are properly referenced.**

Authors' Response:

Thank you for the helpful suggestion; the following tools and libraries are now appropriately cited in the manuscript where used: spaCy, RapidFuzz, GeoPy (with Nominatim), OpenAI GPT-3.5 / GPT-4 APIs, Flask, Pandas, Folium, Python 3.10, Adobe OCR / PDF Services, and OpenStreetMap. (revision is added to the text and reference section)

Reviewer 3:

PAGE 8

Data collection and integration:

- a. **Were any OCR or GPT-4 digitizations/conversions manually reviewed for error and reliability?**

Authors' Response:

Yes, manual review was performed. we manually verified the outputs for accuracy and reliability. In particular, provider names, service descriptions, and contact information were manually cross-checked for correctness after automated conversion. (revision is added to page 10 to clarify this note)

- b. **What is the final number of service providers included in the Minerva database?**

Authors' Response:

The final database compiled for this study included 1,615 service provider entries. This figure reflects the combination of two sources: 513 providers identified through manual and automated web scraping, and 1,102 stakeholder entities extracted from documentation shared by the Alabama Department of Mental Health. The database was used across both the Athena and Minerva chatbot systems. While some duplication may exist between the two sources, no systematic deduplication was conducted beyond basic filtering, as the primary goal of this study was to demonstrate technical feasibility and proof-of-concept rather than to create a fully production-ready dataset. (revision is added to page 9)

c. Does “providers” refer to individuals, organizations, or another entity?

Authors’ Response:

“providers” refers to both individual practitioners (e.g., pediatric neurologists, speech therapists) and organizations or institutions (e.g., clinics, therapy centers, advocacy groups). Each entry in the database represents a distinct service delivery unit with a physical address and contact information, regardless of whether the service is rendered by an individual or organization. (revision is added to page 9 to clarify this note)

d. The term “stakeholder” seems broad—could you clarify what or who the 1,102 stakeholders represent? This number appears quite large.

Authors’ Response:

We have clarified the term “stakeholder”. In this context, the stakeholder entries provided by the Alabama Department of Mental Health included individuals and organizations relevant to autism services in Alabama. These included therapy centers, nonprofit service providers, schools, advocacy groups, residential programs, and various public agencies. Each was treated as a potential service provider entry in the combined dataset used across both chatbots. (revision is added to page 9 to clarify this note)

Reviewer 3:

Table 1:

- a. **Why was a binary tagging system chosen to represent service offerings? How were nuanced or variable service offerings handled? How was “offered services” defined operationally?**

Authors’ Response:

The binary tagging system was chosen as a practical and scalable approach to enable rapid filtering and matching between user queries and provider offerings. Because service categories in our dataset were often sparse, inconsistently labeled, or unstructured, we prioritized a yes/no binary structure for each tag to reflect whether a given service type was explicitly offered by a provider.

Conditional offerings (e.g., age- or diagnosis-specific programs) were handled by adding more specific service tags, rather than introducing partial weights or probabilistic confidence. For example, instead of marking a provider as “partially offering therapy,” we might tag it under “speech therapy,” “occupational therapy,” or “ABA,” based on the available information. The term “offered services” was operationally defined as: “Any service that a provider actively advertises or is documented as delivering to individuals with autism or their families, as per publicly available data or agency-supplied documentation.”

(revision is added to page 10)

- b. **The text mentions 300 categories, but Table 1 lists only 11. Could you clarify what the remaining 200+ categories are?**

Authors’ Response:

We appreciate this opportunity to clarify. Table 1 was intended as a sample excerpt, not a full listing of the service tag set, which indeed includes over 300 tags. A categorized summary of the types of stakeholder roles and services is included in Supplementary Table S1. This table groups services by provider type and service focus, offering a clearer picture of the tag diversity. We have revised the manuscript to clarify that Table 1 is illustrative. (revision is added to page 10)

- c. **It seems inconsistent to list some specific proprietary companies (e.g., Cigna, United) alongside broad generic categories (e.g., financial supports). Are these treated differently within the tagging system? For example, since Cigna is a type of financial support, why is it separated out as its own tag rather than grouped under the broader category?**

Authors' Response:

Proprietary companies (e.g., "Cigna," "United Healthcare") were included as distinct tags only when their presence was functionally relevant to a user's ability to access services. In many cases, families searching for care require providers who accept specific insurance plans, and this information is critical to query resolution. Thus, while "financial support" is a general service type, tagging providers by accepted insurers (e.g., "accepts Cigna") was treated as an orthogonal layer of information relevant to accessibility. These tags were derived directly from provider websites or verified documents and were not mutually exclusive with broader service categories.

The edited manuscript clarifies this distinction, we have added a clarifying sentence accordingly. (revision is added to page 10)

Reviewer 3:

PAGE 10

Development of the NLP-driven chatbot:

- a. **The chatbot's use of a service synonym dictionary is a strong idea, but which specific dictionary was utilized? Reference [54] includes several possibilities.**

Authors' Response:

We appreciate the reviewer's close reading of our use of the service synonym dictionary. To clarify: although Reference [54] (Navigli, 2009) provides valuable background on semantic mapping and lexical disambiguation, we did not directly use BabelNet or WordNet in this study. Instead, the synonym dictionary was developed in-house, using a custom process:

- Terms were extracted from publicly available provider listings and documentation;
- Mapped by expert social workers into standardized service categories (e.g., "speech-language pathologist," "speech therapy," "SLP" → "Speech Language Therapists");
- Structured into a lookup table to normalize user queries.

We have revised the manuscript to clarify this. (revision is added to page 12)

- b. **Figures 1 and 2 provide helpful visualization, but the blurring of sensitive information might not be sufficient to ensure privacy.**

Authors' Response:

We would like to clarify that the content shown in Figures 1 and 2 is drawn exclusively from publicly available sources such as provider websites and public service directories. No private individual information, personal health data, or user-specific content is included. Furthermore, for the purpose of visual presentation, we applied additional manual blurring and redaction to obscure provider names, addresses, phone numbers, emails, and URLs, even though this information is publicly accessible. We believe that, in its current form, the figure ensures full compliance with ethical and privacy standards. Nonetheless, based on this comment, we made further adjustments to both figures to enhance the blurring and ensure complete clarity. (revision is added to page 13,14)

Reviewer 3:

PAGE 14

Evaluation framework:

- a. How was it determined whether chatbot responses fully satisfied query criteria? How were partially correct responses handled in multi-criteria queries?**

Authors' Response:

Our evaluation framework treated a chatbot response as “correct” only when all key query constraints were satisfied. If a returned provider entry matched only a subset of the constraints it was not counted as a true positive in precision/recall calculations. Partially correct responses were logged for manual review, but not included in the quantitative metrics. We opted for a strict correctness criterion to ensure high precision and to avoid overstating system performance. We have added a clarification in the manuscript accordingly. (revision is added to page 16)

Reviewer 3:

Figure 3:

- a. This is a very informative figure; however, I'm unsure how it relates to the manuscript.

Authors' Response:

Figure 3 reflects an important finding that emerged during the construction and validation of the autism provider database: clear geographic and socioeconomic disparities in the distribution of autism-related services across the state of Alabama. While this is not the central focus of the study, it directly illustrates the need for accessible, intelligent tools that help underserved families navigate a fragmented care landscape. We have clarified the role of Figure 3 in the revised text and added a reference in the discussion section to emphasize that this disparity warrants addressing in future studies the importance of adaptive decision-support tools such as the tools demonstrated in this study. (revision is added to page 28)

Reviewer 3:

PAGE 15

Table 3:

- a. Regarding Table 3, what does “accuracy” specifically refer to - precision, recall, F1 score, or another metric?

Authors' Response:

Thank you for the comment. In both Table 3 and Table 4, the term “Accuracy” refers to the F1 score, the harmonic mean of precision and recall, as defined in the Evaluation Framework section. This was chosen as a balanced performance summary for multi-constraint queries. To avoid confusion, we clarified this in the table notes.

Reviewer 3:

REFERENCES

Preprint reliance: The reference list is disproportionately weighted toward preprints, with 11 out of 62 citations from non-peer-reviewed sources. Consider whether some of these can be replaced with peer-reviewed equivalents, especially for key claims or data.

Authors' Response:

Thank you for this helpful comment. We agree that peer-reviewed sources are ideal when available. However, the rapid pace of AI research means that preprints (particularly from arXiv) were found to be the most current and technically detailed sources, especially in areas like LLMs and retrieval-based systems.

That said, we reviewed our citations and updated references [10], [18], [42], and [55] to their peer-reviewed published versions, thereby improving the balance of peer-reviewed to preprint sources. For the remaining preprints, no published alternatives currently exist, and they represent foundational work widely used in the field.

[8]: DOI appears to be available but is not included.

Thank you for noting that for reference [8], the CDC Community Report on Autism, lacked a DOI in the current citation. We have verified that this report is the 2023 ADDM Network community surveillance report published by CDC in Morbidity and Mortality Weekly Report (MMWR) Surveillance Summaries. It is properly cited as Shaw et al. (2025), MMWR Surveillance Summaries with DOI, and this version is now used in the revised manuscript.

[20] & [21]: Both entries share the same DOI, which points to Shaw et al.

After carefully reviewing our reference list, we confirmed that [20] and [21] are distinct publications with different authors, topics, and DOIs.

[22] & [51]: Both citations are vague and link only to general websites.

We agree that references [50] and [51], both linking to general OpenAI web pages, were too informal for academic citation. In the revised manuscript, we have replaced these with authoritative and citable sources

[34]: Incomplete citation. Only the arXiv homepage is listed. Include the full title, full author name(s).

We have carefully reviewed Reference [34] and confirm that it currently includes the full title, author names, and the appropriate arXiv DOI. It appears this issue may have stemmed from an earlier version of the manuscript, as the reference is now complete and properly formatted.

[55] & [58]: These appear to be exact duplicates of the Geopy GitHub repository. Remove one unless both are intentionally cited in different contexts (and clarify if so).

Upon review, References [55] and [58] are not duplicates and neither refers to the Geopy GitHub repository. Reference [55] cites a widely used evaluation metrics paper (Powers, 2011), and Reference [58] is a systematic review on the concept of comprehensive care in nursing.

[62]: Citation is incomplete and refers to a blog post. Consider whether this is appropriate for an academic reference.

Reference 64 is currently unused in the manuscript and likely a leftover or mistakenly added citation.