Information Retrieval Research 2 (2026) 59-112 Submitted 12/25; Revised 05/26; Published 06/26

Much Ado about Accessibility: An Exploration of Online Content Accessibility from an Autism-Informed Perspective

Hrishita Chakrabarti H.CHAKRABARTI@TUDELFT.NL
Delft University of Technology
Delft, Netherlands

Maria Soledad Pera M.S.PERA@TUDELFT.NL
Delft University of Technology
Delft, Netherlands

Editor: Haiming Liu

Abstract

Autism Spectrum Disorder is a neuro-developmental condition that may affect the way a person learns, behaves, and interacts with their environment. Autistic individuals are known to often rely on mainstream information access systems, such as search engines, as their initial point for information discovery. Despite their active online information-seeking practices, autistic individuals struggle to extract information from the resources they encounter, leading to frustration. This raises the question of whether these systems fit the needs and expectations of this population. In this work, we empirically explore how different mainstream information access systems respond to autistic users' inquiries through the lens of web content accessibility. Specifically, we focus on three text-based perspectives known to impact autistic individuals' information seeking – structure, readability, and concreteness (tangibility of terminology) of web content. For our analysis, we leverage established accessibility guidelines for autistic individuals to create accessibility profiles for Google, Bing, ChatGPT, and Gemini, providing an overview of various aspects of each system's responses to common inquiries of autistic individuals. Given the influence of query formulation on produced responses, we also investigate the impact of explicitly informing the system of the user's condition on the accessibility profile of the considered systems. Our findings reveal potential accessibility limitations of mainstream information access systems when responding to autistic users' inquiries, more so if their information need is related to the searchers' condition. Furthermore, including a diagnosis-disclosing phrase in autistic users' inquiries results in responses that are even less accessible to autistic individuals. Based on our findings, we reflect on autistic users' access to information and the role of information access technologies in supporting their information-seeking process.

Keywords: generative models, text analysis, autism, web accessibility, search engines, LLMs

1 Introduction

Autism Spectrum Disorder (**ASD**) is a neuro-developmental condition that affects how a person interacts and communicates with the people and environment around them (CDC, 2024b). The abilities of autistic individuals[1] can vary immensely, i.e., individuals' abilities

1. There is an ongoing debate on how Autism must be described. Here, we use "identity-first" (e.g., autistic individuals or autistic users) instead of "person-first" language (e.g., individual with Autism or user with Autism) based on the preferences of autistic individuals (Kenny et al., 2016; Bottema-Beutel et al., 2021).

 WWW: https://irrj.org DOI: 10.54195/irrj.25297

lie on a *spectrum* with an autistic person's needs potentially changing over time (WHO, 2023). Despite their heterogeneity, common struggles include issues with memory, attention, stimulus sensitivity, and language and visual comprehension (APA, 2022; CDC, 2024a). These conditions often hinder autistic users from socialising and carrying out their daily activities, leading to loneliness and poor quality of life (Locke et al., 2010; Lin and Huang, 2019).

Autistic individuals display a strong affinity for technology–especially that aiming to support their daily activities (Lin et al., 2013; Putnam and Chong, 2008). They also have a prominent presence on online platforms, which they use for various purposes, including establishing and maintaining relationships through social media platforms as well as searching for information related to their interests (Bosseler and Massaro, 2003; Kim et al., 2020; Gibson and Hanson-Baldauf, 2019). For the latter, they often turn to popular information access systems (**IASs**), such as search engines (**SEs**) like Google and Bing, as the portal to online information (Gibson and Hanson-Baldauf, 2019; Jang et al., 2024). However, they sometimes struggle with extracting information from the online resources they come across (e.g., Eraslan et al., 2017; Friedman and Bryen, 2008). This is known to stem from various factors; of note, autistic users find it difficult to locate relevant components of information when a website is too visually complex due to autistic individuals' attention issues and challenges with visual comprehension (White and Abou-Zahra, 2024; Eraslan et al., 2017). Additionally, the phrasing of textual content in an online source can impact an autistic person's comprehension of the content. Vague metaphors and "fancy" language, used to make content engaging, can sometimes end up confusing an autistic person (White and Abou-Zahra, 2024; Yaneva et al., 2015).

Autistic individuals are increasingly turning to generative AI-based IASs for information seeking, i.e., LLM-agents like ChatGPT and Gemini (White, 2024), (Jang et al., 2024; Spengler, 2023), which leverage Large Language Models (**LLMs**) for response generation. This is due to the generative capabilities of these agents to personalise the response based on an autistic user's specific cognitive processing styles (Jang et al., 2024; Spengler, 2023). LLM agents also remove the issue of crowded websites but respond to users' inquiries in long passages of text, which may lead to information overload, inciting challenging behaviours or meltdowns in autistic users, which hinders autistic individuals from utilising the agent's answer for their information discovery (White and Abou-Zahra, 2024). These challenges highlight the specific needs and expectations of autistic users when seeking information online. However, there is limited understanding of how well popular IASs accommodate for autistic users' distinct requirements while catering to the user group's information needs. This prompts the need to examine whether popular IASs respond to autistic individuals' inquiries in a manner suitable for autistic users.

Inspired by the Web Accessibility Initiative (**WAI**) of the World Wide Web Consortium (**W3C**), which aims to ensure that the web content users consume accommodates their distinct abilities (W3C, 2019), in this work we conduct an empirical exploration to gauge the extent to which responses by four popular commercial IASs—Google, Bing, ChatGPT, and Gemini—for autistic individuals' inquiries align with established Web Content Accessibility Guidelines (**WCAG**) for autistic individuals. Considering how an online inquiry is formulated biases the response elicited from an IAS (Alaofi et al., 2022; Murgia et al., 2023a; Schmidt et al., 2023), we further investigate the impact of explicitly stating the user's condition in autistic users' inquiry on the accessibility of the response. To guide our exploration, we pose two research questions:

- **RQ1:** Are the responses of mainstream SEs and LLM agents accessible to autistic users?
- **RQ2:** Are the responses of SEs and LLM agents more accessible when diagnosis is disclosed in a query?

For our exploration, we rely on the WCAG proposed by the Cognitive and Learning Disabilities Accessibility Task Force (**COGA**) – a task force under WAI's Accessibility Guidelines Working Group, created to understand and address the challenges faced by autistic users (amongst other user groups with cognitive and learning disabilities) when exposed to content on the web (W3C, 2021). We scope our investigations to the *textual content* of IAS responses and create *accessibility profiles*. Leveraging the outcomes of empirical studies that validate the applicability of COGA's guidelines (e.g., Yaneva et al., 2015; Eraslan et al., 2017), an accessibility profile provides an overview of various text-based aspects of an IAS response to user inquiries that are known to impact autistic individuals' information seeking. Particularly, we focus on aspects related to three broad perspectives of text accessibility:

1. *Structure*, i.e., how concise the passages are, as guidelines suggest that text should be written in short paragraphs or a list of bullet points to avoid inciting information overload in autistic users (Britto and Pizzolato, 2016; Seeman and Cooper, 2024),
2. *Readability* or in other words language complexity, as autistic individuals prefer simple and easy-to-read sentences over complex sentences with niche jargon (Yaneva et al., 2015),
3. *Concreteness*, which refers to the tangibility of the word describing a concept (Paivio, 2013), as text accessible to autistic users should explain related concepts concretely and avoid vague metaphors as much as possible (Yaneva et al., 2015; Yaneva and Evans, 2015).

To ascertain if the accessibility of IAS responses is inherent to general algorithmic behaviour intrinsic to an IAS or if it changes depending on a particular user group and their inquiries, we also create accessibility profiles of the considered IASs based on their responses for inquiries formulated by the general population, which we treat as the control group. Due to the scarcity of public logs capturing IAS responses to user queries, we generate the accessibility profiles for the considered IASs based on responses elicited using query samples that reflect the common information needs of autistic users and the general population, respectively. Query samples from a general population are commonly available for research purposes; in our case, we use a random sample from Yahoo! Search Query Tiny sample dataset (Yahoo!, 2009). Unfortunately, queries from autistic users are generally not shared or part of publicly available datasets. Consequently, we anchor our exploration on a newly generated sample of synthetic queries that represent the information needs of autistic individuals. For this, informed by similar studies conducted for other non-traditional user groups, such as those suffering from mental health disorders for whom data is also scarce (e.g., De Choudhury et al., 2016; Ríssola et al., 2022; Milton and Pera, 2023), we turn to Reddit to collect posts made by self-reported autistic users on various forums dedicated to different topics related to the target population. We extract popular noun phrases from these posts to use as *proxy queries* related to topics of interest for autistic individuals that also capture the common vocabulary and phrasing used by the target population.

Findings reveal that all the considered IASs respond with verbose, hard-to-read, and abstract texts for autistic individuals' queries compared to the responses for inquiries initiated by the general population. Moreover, when an autistic user's inquiry relates to ASD, IAS responses are much harder to read than when the inquiry pertains to other topics. Furthermore, we find that explicitly informing the IAS that the inquiry is from an autistic user results in more verbose responses, which are also more abstract in nature, compared to when the systems are unaware of the user's condition. We do, however, observe that Google responses are considerably easier to read when an autistic user's query is rephrased to explicitly state the user's condition. Furthermore, IAS responses when informed of the user's condition, although less aligned with autistic individuals' accessibility requirements on average, are more consistent in their structure, readability, and concreteness across queries.

The main contributions of this work are (i) generating accessibility profiles of IAS responses based on textual features that impact web content accessibility for autistic users during search; and (ii) providing empirical evidence that in terms of text accessibility, the considered IASs respond differently depending on the user group and their inquiries, with IAS responses being less aligned with autistic individuals' accessibility requirements when the inquiry pertains to autistic individuals' information needs compared to when the inquiry comes from the general population. This raises concerns about potential issues in autistic individuals' access to information through online mediums—a direct infringement of their fundamental human rights (United Nations, 1948)—evincing the need to adapt popular IASs to better cater to the needs and expectations of autistic people. To support research in this direction, the experimental setup we use to generate accessibility profiles can be extended to examine IAS responses from other perspectives of web content accessibility to realise other potential accessibility limitations of online IASs. For reproducibility purposes and to facilitate further analysis to understand information accessibility for not only autistic people but also other vulnerable and underserved user groups, we share our code at: `https://github.com/HrishitaC/much-ado-about-accessibility`.

2 Related Work

We discuss below related literature that offers context and informs the experimental designs of our empirical investigation into the accessibility of responses provided by popular IASs to autistic users.

Autism Spectrum Disorder. ASD is a neuro-developmental condition that causes affected individuals to communicate and behave in ways that are different from most people (CDC, 2024b). Autistic people can begin showing symptoms as early as the age of 2 years, but most individuals are diagnosed much later in their lives (Nyrenius et al., 2023; Lai and Baron-Cohen, 2015). Currently, it is estimated that 1 in every 100 children is autistic (WHO, 2023), with the condition more prevalent in boys than girls (Zeidan et al., 2022). Increased awareness and changes in diagnostic criteria have resulted in a significant rise in autism diagnoses over the past few years (Zeidan et al., 2022; Russell et al., 2022), yet ASD remains largely underdiagnosed (O'Nions et al., 2023; WHO, 2023), especially among women (Cary et al., 2023; Driver and Chester, 2021).

The abilities of autistic individuals are described to lie on a spectrum, with the symptoms and effects of the condition varying from person to person and evolving with age (CDC, 2024a; WHO, 2023). Typically, an autism diagnosis is categorised into 3 levels depending on the extent of support the affected individual requires for their daily activities, where level 1 Autism indicates low-level support, primarily with social communication, while level 3 Autism implies that the individual requires substantial support to carry out even day-to-day activities (APA, 2022). Despite the variation in their abilities, most autistic people face issues with reading comprehension, including difficulty processing long, complex sentences and understanding figurative language and abstract words (O'Connor and Klein, 2004; Frith and Snowling, 1983; Happé, 1997). Autistic individuals also depict atypical attention patterns such as their reliance on a single sensory modality when performing a task, a phenomenon called "stimulus overselectivity" (Lovaas and Schreibman, 1971). Due to this phenomenon, autistic people often tend to be more detail-oriented and fail to see the "bigger picture", i.e., they focus on local components of information and fail to consider the global context and semantic information provided in a reading material (Happé and Frith, 2006; Ploog, 2010). Challenges with reading tasks are a common reason for lower academic and professional achievements in autistic users which in turn have long-term negative impacts on autistic people's employment rates, social relationships, and mental and physical health (Howlin and Moss, 2012) resulting in lower quality of life for autistic people compared to non-autistic individuals (Lin and Huang, 2019; Mason et al., 2018).

ASD & Technology. Autistic individuals show a natural affinity to technology to aid them in their daily life (Lin et al., 2013; Putnam and Chong, 2008). Particularly, due to their struggles with social interactions, autistic individuals have a very restricted friend circle, due to which they are reported to show higher levels of loneliness compared to the general public (Shattuck et al., 2011; Locke et al., 2010). Social media platforms have proven to be an effective tool for autistic individuals to abate this problem, with studies revealing a strong preference amongst autistic individuals to maintain their social contacts and even create new ones on these platforms (Gillespie-Lynch et al., 2014; Wang et al., 2020). Online communications provide autistic individuals with a sense of control over the conversation that they find to be lacking in a face-to-face conversation (Benford and Standen, 2009; Koteyko, 2023). Apart from maintaining their social contacts, autistic individuals also often rely on computer-based assistive technology to develop or enhance essential life skills from basic motor skills (e.g., Zheng et al., 2021; Zhao et al., 2018; Cao et al., 2025) to skills required to navigate specific social scenarios (e.g., Hartholt et al., 2019; Kandalaft et al., 2013).

Researchers have also allocated efforts in improving diagnosis and interventions for autistic individuals. For instance, autistic users' posts and their interactions on platforms like Reddit or Twitter have been used to detect linguistic and behavioural signals characteristic of autistic individuals, which can be used in detection models to facilitate early diagnosis of Autism (e.g., Kim et al., 2020; Rubio-Martín et al., 2023). In a study by Dotch et al. (2023), a qualitative analysis of the posts and comments made by self-reported autistic users on two online forums resulted in rich insights into the personal experiences and coping mechanisms for their issues with noise sensitivity. The outcomes of the study have enabled the development of user-centred technological solutions to tackle noise sensitivity for autistic people (e.g., Park et al., 2024; Dotch, 2024). In our case, we take advantage of posts shared by self-reported autistic users on Reddit to overcome the lack of public data on online inquiries commonly asked by autistic individuals. Particularly, we extract common noun phrases from the Reddit posts to serve as synthetic queries that represent the typical information needs of autistic individuals.

The advent of LLMs has resulted in considerable advancements in diagnostic strategies for ASD (Haskins et al., 2024; Hu et al., 2024; Jiang et al., 2024). The generative capabilities of LLMs have introduced new opportunities for facilitating autistic individuals' social communication, such as providing personalised support in professional settings (Jang et al., 2024), dynamically generating feedback on the implicit emotional tone and intent in text messages (Haroon and Dogar, 2024), etc. Furthermore, researchers have leveraged LLMs to design conversational agents that can provide personalised clinical support for autistic individuals (Chu et al., 2024; Shubbar, 2025). For our study, considering the growing popularity of LLM-based chatbots like ChatGPT in autistic individuals' everyday lives (Choi et al., 2024), we examine the potential of popular LLM agents as accessible IAS for this target audience.

Regarding autistic individuals' information seeking, research efforts have resulted in the proposal of several accessibility guidelines to align web content and interfaces with the needs and expectations of autistic individuals (Britto and Pizzolato, 2016; Dattolo et al., 2016b; Raymaker et al., 2019). These guidelines, in turn, have been used to guide the development and evaluation of technological interventions to specifically facilitate autistic users' information seeking. Interventions in this area range from online tools that modify web content to enhance readability for autistic individuals (Pavlov, 2014; Eric, 2019) to online platforms that support autistic individuals information seeking in specific contexts such as tourism (Dattolo et al., 2016a; Rapp et al., 2017; Cena et al., 2020), e-learning (Judy et al., 2012; Rathod et al., 2024), medical question-answering (Chu et al., 2024), etc. Although individuals who are aware of their diagnosis do tend to utilise these resources to ease their search process (Gibson and Hanson-Baldauf, 2019), considering the low diagnosis rate of ASD around the world (WHO, 2023), many autistic individuals, unaware of their condition, may still rely on the resources a mainstream IAS would provide in response to an autistic user's inquiry. To investigate the degree to which popular IASs could support autistic users' in such scenarios, we examine the variance across different aspects of mainstream IAS responses based on text-based factors known to influence the target population's information-seeking behaviour.

ASD & Online Information Seeking. Autistic individuals frequently undertake self-directed online learning tasks on topics of their interest, but do not always trust the quality of the information available online and limit themselves to a few resources or tools that meet their expectations during a search task (Gibson and Hanson-Baldauf, 2019). The lack of trust stems from the frustration due to the unsuitable format of information on several web sources (Gibson and Hanson-Baldauf, 2019), which has been empirically proven to hinder autistic users from efficiently searching for the information they need (e.g., Eraslan et al., 2020; Yaneva et al., 2015; Rezae et al., 2020). Some researchers attribute the inefficiency of autistic users when looking for information on the web to the difference in the attention patterns between autistic users and non-autistic users as well as the information-processing issues common in autistic individuals (APA, 2022; Williams et al., 2015). The eye scan paths of an autistic user and a non-autistic user when searching for information on a webpage were visualised in a study by Eraslan et al. (2017) highlighting the stark differences in the attention patterns of autistic users from non-autistic users when trying to extract information from a webpage. Similar studies involving tracking users' eye movements when scanning a web page revealed that autistic users scan more items on the web page, most of which were irrelevant to the question they sought to answer (Eraslan et al., 2017, 2020). Autistic adults also prefer image-based elements and feel a strong aversion to text-heavy elements (Yaneva et al., 2015). Although the image-based elements are effective in supporting autistic users' information seeking when the image is directly related to the textual content, as the presence of abstract images and icons negatively impacts the autistic users' efficiency when searching for information on a web page (Yaneva et al., 2015; Rezae et al., 2020). The impact of a web interface on autistic individuals information access is not limited to hindering autistic users ability to extract information, but can also influence other processes related to information seeking. For instance, Alzahrani et al. (2022) observed that the presence of irrelevant animations in a web interface can hinder an autistic user from formulating effective queries during search sessions.

Existing works that investigate and address the challenges autistic individuals face when seeking information online largely involve experiments in which participants complete predefined search tasks using either curated web resources (e.g., Eraslan et al., 2020; Yaneva et al., 2015) or a specific web interface (e.g., Alzahrani et al., 2022; Rezae et al., 2020). The insights gained from such controlled experiments help identify barriers that could hinder autistic individuals from finding the information they need. However, these studies do not necessarily indicate that autistic individuals face these barriers when interacting with mainstream IASs for addressing their information needs. The work by Yechiam and Yom-Tov (2021) is one of the few empirical studies based on actual browsing records of self-reported autistic users when searching for information. However, the study focuses on discerning the characteristic search behaviours of autistic users and not the extent to which the search results are accessible to autistic users, which is what we pay attention to in our study. In our work, we focus on the system and probe mainstream IAS responding to the information needs of autistic individuals, as captured in queries related to topics commonly of interest for this user group.

3 Experimental Framework

In this section, we describe the components of the experimental framework used to conduct our investigation. This work has been approved by the Human Research Ethics Committee (HREC), which is the local institutional review board for Delft University of Technology.

3.1 Information Access Systems

Considering the popularity of SE and increasing use of LLM agents amongst autistic individuals for information seeking (Yechiam and Yom-Tov, 2021; Jang et al., 2024; Spengler, 2023), we focus our exploration on these two types of systems.

Amongst SEs, Google continues to be the most popular SE of all time, followed by Bing (Bianchi, 2024). To elicit responses from Google and Bing, we use their official APIs, i.e., Google Custom Search JSON API[2] and Bing Web Search API[3]. For the Google API, we use all default parameters except that we set the language to English using the key–value pair {`'lr': 'lang_en'`}. For the Bing API, we restrict results to the top 10 and set the language and location to English and the USA, respectively, to align with the default Google API configuration: {`'answerCount': 10, 'setLang': 'en', 'mkt': 'en-US'`}.

In the case of LLM agents, OpenAI's ChatGPT is most commonly known and used by autistic users (Jang et al., 2024), which is why we consider it in our exploration, along with Google's Gemini due to its comparable performance to ChatGPT in benchmark question-answering tasks (Gemini Team Google et al., 2023; Achiam et al., 2023). We elicit the direct answer from ChatGPT and Gemini using the official APIs of `GPT-3.5 Turbo`[4] and `Gemini Pro`[5] models, respectively which are the models driving ChatGPT and Gemini's response generation at the time of our study (OpenAI, 2022; Liedtke and O'Brien, 2023; Milmo, 2023). Like for SE responses, we limit LLM agent responses to English. Note that all API calls are made independently and anonymously to control for the influence of user profiling on the elicited responses.

3.2 Data Collection

Logs from mainstream IASs are generally scarce, especially when capturing the interactions initiated by autistic individuals specifically. Consequently, to enable the analysis of accessibility of IASs when responding to autistic individuals' inquiries, we construct *synthetic logs* that capture the content autistic individuals would be exposed to when interacting with an IAS for information search. For this, we require queries that reflect the common information needs of the target population as a starting point. Considering the different needs and varied responses to stimuli of autistic individuals (CDC, 2024a; WHO, 2023), the queries need to capture the various areas of interest for the target audience. Unfortunately, query samples from existing studies on autistic users' search behaviour are either proprietary data (Yechiam and Yom-Tov, 2021; Alzahrani et al., 2022) or cannot be shared publicly due to regulations protecting the privacy of vulnerable populations (Milton and Pera, 2023), which limits our access to queries formulated by a diverse group of real autistic users.

To overcome this challenge, motivated by similar works involving other vulnerable and under-represented populations, such as individuals afflicted with mental health issues, we adopt a synthetic approach that leverages data from public social media interactions of self-disclosed individuals belonging to a target population to capture their common needs and interests (e.g., De Choudhury et al., 2016; Paul and Dredze, 2011; De Choudhury et al., 2014; Ríssola et al., 2022). Specifically, we follow the method used in (Milton and Pera, 2023) and extract common noun phrases from social media posts created by individuals who report themselves to belong to a target population. Through this strategy, we isolate phrases containing terminology used by autistic individuals on topics of common interest for this population, which can serve as a proxy for queries that reflect the target user

2. https://developers.google.com/custom-search/v1/overview
3. https://learn.microsoft.com/en-us/bing/search-apis/bing-web-search/overview
4. https://platform.openai.com/docs/models/gpt-3-5-turbo
5. https://cloud.google.com/vertex-ai/generative-ai/docs/model-reference/gemini

group's common information needs. In our case, noting the popular use of Reddit posts to study the topics, linguistic patterns, and behavioural aspects typical for autistic users (e.g., Bahrke, 2024; Varghese et al., 2024; Thom-Jones et al., 2024; Fong et al., 2025), we turn to posts made by self-reported autistic users from 16 subreddits (Medvedev et al., 2019)[6]. In doing so, we consider data addressing different perspectives within the autistic community, enabling a wide and diverse coverage of topics and vocabulary commonly used by autistic people. See Appendix A for the description of each subreddit that was scraped for data collection.

Using the official Python Reddit API wrapper (PRAW[7]), we collected the title and text body of 11,100 Reddit posts[8]. To transform the posts into synthetic queries, we extracted the common noun phrases using `KeyBERT` (Grootendorst, 2020), a BERT embeddings-based keyphrase extraction technique, which yielded approximately 28,880 key phrases. As stated in (Van Rijsbergen, 1979), it is common for certain terminologies to be used very frequently in any given corpus to a point where the terms do not capture any distinct information and instead serve as corpus-specific stopwords. To eliminate such terms from our collection of keyphrases, we turn to Zipf's law (Zipf, 2016), which states that the frequency of any term is inversely proportional to its rank in a frequency table. Specifically, following the stopword-removal method used in (Sahu et al., 2023; Lo et al., 2005; Courseault Trumbach and Payne, 2007), we rank the keyphrases in a descending order of their frequency of occurrence (grouped by their n-gram length). We plot the keyphrases in the order of their rank and estimate a maximum threshold of occurrence for each group (listed in Table 1). Keyphrases with a frequency above the threshold are considered as stopwords and therefore excluded, leading to a collection of 7000 key phrases as candidate synthetic queries.

n-gram	**Max # of posts an included key phrase can occur in**
1	400
2	100
3	10
4	4

Table 1: Upper limit of occurrence for key phrases of different n-gram lengths to exclude in-context stopwords.

We use a sample of the most common key phrases for each n-gram length as the set of queries that capture the information needs of autistic users. Given the impact of query lengths in biasing search results (Alaofi et al., 2022; Pera et al., 2023), we pay particular attention to ensure that the set of synthetic queries follows a realistic query length distribution. However, due to the lack of information on such a distribution specifically for autistic

6. Note that no personally identifiable information was collected, and all content was de-identified before use. All collected data was discarded after the completion of our experiments.
7. `https://github.com/praw-dev/praw`
8. The data was extracted from Reddit in April 2024 and comprises posts created by self-reported autistic individuals between April 2019 and April 2024.

users at the time of the study, we turn to the Yahoo! Search Query Tiny samples dataset[9] (Yahoo!, 2009; Blanco et al., 2011), as it comprises queries formulated by real people and therefore would represent a query length distribution realistic at least for a generic user[10]. Specifically, we compute the proportion of queries for each n-gram length and mimic those proportions to generate set $\mathtt{Q_A}$ comprising 250 queries of which 78 are unigram queries, 78 bigram queries, 70 trigram queries, and the remaining 24 of n-gram length ≥ 4. The queries in $\mathtt{Q_A}$ range from phrases directly related to ASD, such as "autism", "autism diagnosis", and "autistic burnout", to other mental health related phrases like "crippling social anxiety", "sensory overload", and "executive dysfunction", as well as more generic topics like "part time job", "high school", and "facial expression".

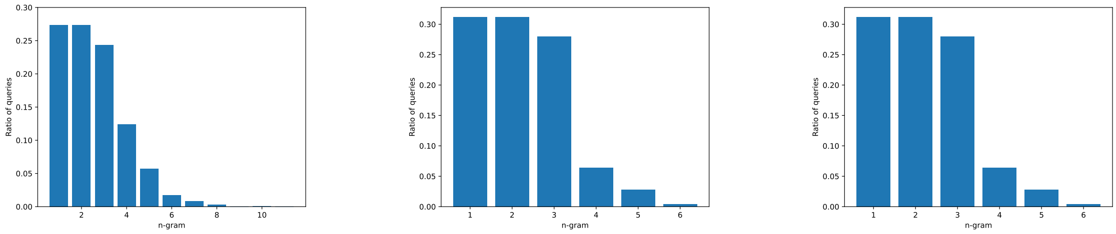
Different people express the same information need in different ways, which in turn has a strong influence on the response produced by an IAS (Alaofi et al., 2022; Pera et al., 2023). Given that diagnosed autistic users tend to report their diagnosis in their search query, we create a second *variation* of the synthetic queries in $\mathtt{Q_A}$ to reflect the information needs of autistic individuals covered in $\mathtt{Q_A}$ while also informing the system of the user's condition. To reflect autistic individuals' natural style of reporting their condition in our synthetic queries, we refer to the phrases used by autistic individuals to report their condition in their search query, as identified by Yechiam and Yom-Tov (2021) from the query log the authors used to study the online search behaviour of autistic individuals. The query log used in the study is proprietary; hence, we rely on the examples reported in the manuscript, such as *"I'm autistic"*, *"I'm on the autism"* and *"I am on the autism"*. Particularly, we create two sets $\mathtt{Q'_A}$ and $\mathtt{P'_A}$ to probe SE and LLM agent responses, respectively, with each set representing a distinct variation of queries in $\mathtt{Q_A}$. In the case of $\mathtt{Q'_A}$, we simply append the phrase "*I'm autistic*" at the end of each query in $\mathtt{Q_A}$; whereas for $\mathtt{P'_A}$, we use a longer approach to mimic the conversational nature of information-seeking in LLM agents and rephrase each query in $\mathtt{Q_A}$ as "I am autistic please explain" + query + "to me". See Table 2 for examples of queries in $\mathtt{Q_A}$, $\mathtt{Q'_A}$, and $\mathtt{P'_A}$.

$\mathtt{Q_A}$	$\mathtt{Q'_A}$	$\mathtt{P'_A}$
autism diagnosis	autism diagnosis I'm autistic	I am autistic please explain autism diagnosis to me
common autistic trait	common autistic trait I'm autistic	I am autistic please explain common autistic trait to me

Table 2: Examples of queries in $\mathtt{Q_A}$, $\mathtt{Q'_A}$, and $\mathtt{P'_A}$.

To contextualise whether the accessibility of IAS responses to autistic individuals' (synthetic) queries is inherent to a system's algorithm or shows variances when catering to the information needs of different user groups, we create another query set, $\mathtt{Q_C}$ to capture the topics of interest, vocabulary, and phrasing common amongst the general population. $\mathtt{Q_C}$ consists of 250 queries randomly sampled from the Yahoo! Search Query Tiny sample dataset (Yahoo!, 2009) with a similar n-gram length distribution as $\mathtt{Q_A}$, which we illustrate in Figure 1 along with the n-gram distribution of the original Yahoo! Search Query Tiny sample dataset.

9. The dataset used in this study is no longer publicly available; however, archived versions can be accessed at `https://web.archive.org/web/20250517143901/https://webscope.sandbox.yahoo.com/catalog.php?datatype=l`.
10. Specifically, the dataset comprises a random sample of 4496 queries posted to Yahoo's US SE in January 2009.



(a) Yahoo! Search Query Tiny sample dataset (Yahoo!, 2009).

(b) $\mathtt{Q_C}$

(c) $\mathtt{Q_A}$

Figure 1: n-gram distribution of queries in different query collections.

Using each of the aforementioned query sets, we generate synthetic logs capturing how a given IAS responds to online inquiries. Each log consists of `<query, response>` pairs, where `query` is a query from a given query set, and `response` is the response generated by an IAS for the `query`.

To probe the accessibility of SE responses, we first focus on the Search Engine Result Page (`SERP`) as it is the initial response a searcher receives from an SE for their inquiry. The `SERP` `response` comprises of the top 10 results retrieved by the SE[11]; each result represented by the associated web title and snippet[12]. Additionally, we examine SE responses focusing on individual retrieved results (`RR`) to investigate the content an autistic user would be exposed to upon clicking on a result from the SERP. The `RR` `response` consists of the entire text body of the web resources retrieved as the top 10 results by an SE for a `query`. In the case of LLM agents, `response` refers to the direct answer (`ANS`) that an LLM generates for a `query`.

In the remainder of the manuscript, we refer to a synthetic log as a possible combination of three components: **{**$\mathtt{Q_A}$**,** $\mathtt{Q_C}$**,** $\mathtt{Q'_A}$**,** $\mathtt{P'_A}$**}-{Google (**`G`**), Bing (**`B`**), ChatGPT (**`GPT`**), Gemini (**`GEM`**)}-{**`SERP`**,** `RR`**,** `ANS`**}**[13]. For instance, $\mathtt{Q_A}$-`G-SERP` is the log capturing Google's SERP snippets for each query in $\mathtt{Q_A}$.

3.3 Accessibility Indicators

To examine IAS responses through a web accessibility lens, we rely on 11 quantitative measures, which we refer to as *indicators*, based on WCAG for autistic individuals related to three text-based perspectives that may influence autistic individuals' information seeking: (1) Structure, (2) Readability, and (3) Concreteness. See Table 3 for detailed indicator definitions.

Structure. Some autistic people struggle with reading huge blocks of text, often leading to information overload in them, which could incite challenging behaviours or meltdowns

11. We only consider the top 10 results, i.e., the first SERP, as most users never go past the first result page when searching for information (Sharma et al., 2019).
12. At the time of the study, AI Overviews are not generated consistently and are therefore not included in the collected SERP responses.
13. Note that `SERP` and `RR` are only for SE responses, and `ANS` is only for LLM agent responses.

in autistic individuals, thereby hindering them from engaging with the content (White and Abou-Zahra, 2024; Happé and Frith, 2006). Accessibility guidelines, therefore, suggest presenting textual content as succinctly as possible (Britto and Pizzolato, 2016; Seeman and Cooper, 2024). Given that paragraphs usually are the longest component of a text, we consider the average number of words per paragraph in T (**Average Paragraph Length**) as a measure of *text succinctness.* Moreover, as autistic individuals are known to prefer shorter sentences over long-winded sentences (Yaneva and Evans, 2015), we also examine the text succinctness of a text sample at the sentence-level. Hence, we measure the number of sentences (**Sentence Count**) and the average number of words per sentence (**Average Sentence Length**) as two other accessibility indicators. Further, markup features like lists and headings can improve reading flow for autistic individuals (Britto and Pizzolato, 2016). Therefore, we inspect the structure of T also by the proportion of sentences in T that constitute heading titles (**Ratio of Headings**), list items (**Ratio of List Items**), or a paragraph (**Ratio of Paragraphs**) to assess the *reading flow* of T. Note that we compute the Ratio of Paragraphs for T as the proportion of sentences that are neither headings nor list items. Therefore, the three ratios computed for T add up to 1.

Readability. For an individual to extract information from a resource, they must be able to comprehend the content presented in the resource. For some autistic individuals, texts written in simple and plain language are more suitable for comprehension than texts written in complex sentences (Yaneva et al., 2015; Seeman and Cooper, 2024). Therefore, as our second perspective, we examine the language complexity of T. For this, we turn to two popular readability formulas, which look at different linguistic aspects of a text to determine the reading ease of the text. First, the Flesch Reading Ease Score (**FRES**) (Kincaid et al., 1975) due to its popularity in research, especially in studies pertaining to autistic individuals (Yaneva and Evans, 2015; Yaneva et al., 2015). The FRES score is computed based on the average words per sentence and average syllables per word in a text, and a higher score indicates higher ease of reading. The other readability formula we consider is the Coleman-Liau Readability Index (**CLI**) (Coleman and Liau, 1975) due to its suitability for digital texts (Allen et al., 2022). The index value is computed based on the average number of letters per word instead of syllables, and unlike FRES, a higher CLI index indicates that the text is *less* easy to read.

Concreteness. Autistic people prefer literal language and concrete terminology over figurative and abstract language (Kalandadze et al., 2018; Vulchanova and Vulchanov, 2022). We posit that for a text to be accessible to an autistic person it must be as concrete as possible. To assess the concreteness of T, we estimate the average concreteness of T (**Average Concreteness**) by adopting a lexical approach based on a dataset created by Brysbaert et al. (2014). The dataset consists of concreteness ratings for 40 thousand generally known English lemmas (ranging between 1: maximally abstract and 5: maximally concrete), estimated based on the average of ratings provided by over 4000 participants. Note that, as per this dataset, not all English words include a concreteness score; therefore, we compute the average concreteness of T only based on the words in T that have a concreteness rating in the dataset, i.e, words in T which exactly match those in the dataset. Since the average concreteness does not account for unrated words in T, we also measure the proportion of words in T classified as concrete (**Ratio of Concrete Words**) and those classified as ab-

Accessibility Perspective	Indicator Name	Indicator Formula
Structure	Average Paragraph Length	$ParaLen(T) = \frac{\Sigma_{i=1}^{\lvert P_T \rvert} len(p_i)}{\lvert P_T \rvert},\ p_i \in P_T$
	Sentence Count	$SC(T) = \lvert S_T \rvert$
	Average Sentence Length	$SentLen(T) = \frac{\Sigma_{i=1}^{\lvert S_T \rvert} len(s_i)}{\lvert S_T \rvert},\ s_i \in S_T$
	Ratio of Headings	$HeadingSents(T) = \frac{\lvert H_T \rvert}{\lvert S_T \rvert}$
	Ratio of List Items	$ListSents(T) = \frac{\lvert L_T \rvert}{\lvert S_T \rvert}$
	Ratio of Paragraphs	$ParaSents(T) = 1 - (\frac{\lvert H_T \rvert}{\lvert S_T \rvert} + \frac{\lvert L_T \rvert}{\lvert S_T \rvert})$
Readability	FRES	$\mathrm{FRES}(T) = 206.835 - 1.015 * \frac{\lvert W_T \rvert}{\lvert S_T \rvert} - 84.6 * \frac{syllables(T)}{\lvert W_T \rvert}$
	CLI	$\mathrm{CLI}(T) = 0.0588 * \frac{letters(T) * 100}{\lvert W_T \rvert} - 0.296 * \frac{\lvert S_T \rvert * 100}{\lvert W_T \rvert} - 15.8$
Concreteness	Average Concreteness	$Conc(T) = \frac{\Sigma_{i=1}^{\lvert W_T \rvert} conc(w_i)}{\Sigma_{i=1}^{\lvert W_T \rvert} e(w_i)},\ w_i \in W_T;\ e(w_i) = \begin{cases} 1, & \text{if } conc(w_i) > 0 \\ 0, & \text{otherwise} \end{cases}$
	Ratio of Concrete Words	$ConcreteWords(T) = \frac{\Sigma_{i=1}^{\lvert W_T \rvert} c(w_i)}{\lvert W_T \rvert},\ c(w_i) = \begin{cases} 1, & \text{if } conc(w_i) \geq 4 \\ 0, & \text{otherwise} \end{cases}$
	Ratio of Abstract Words	$AbstractWords(T) = \frac{\Sigma_{i=1}^{\lvert W_T \rvert} a(w_i)}{\lvert W_T \rvert},\ a(w_i) = \begin{cases} 1, & \text{if } 0 < conc(w_i) < 4 \\ 0, & \text{otherwise} \end{cases}$

Table 3: Definitions of indicators used in our study to assess the accessibility of a text sample T, where $len(\cdot)$ yields the number of words, $syllables(\cdot)$ the number of syllables, $letters(\cdot)$ the number of letters, $conc(\cdot)$ the concreteness rating estimated using (Brysbaert et al., 2014), and $\lvert \cdot \rvert$ the size of a set; the variable W_T is the set of words in T, P_T the set of paragraphs, S_T the set of sentences in T. H_T is the set of sentences in S_T that are headings, and L_T is the set of sentences in S_T that are list items. Note that a sentence in T is either a heading, list item, or part of a paragraph, and therefore, the ratios of headings, list items, and paragraphs add up to 1.

stract (**Ratio of Abstract Words**) to contextualise the average concreteness estimation. The classification is done based on the categorisation rule used in the original study, i.e., words rated between 1-3 are abstract and those rated between 4-5 are concrete (Brysbaert et al., 2014).

3.4 Accessibility Profile

To get an overview of a text sample (e.g. a SERP snippet, full text of a resource retrieved by an SE, and answer generated by an LLM agent) from an accessibility point of view, we turn to other works on online content analysis. Specifically, we are inspired by works which create a "profile" to capture various aspects of an item, such as the emotional undertone of IAS responses (Kazai et al., 2019; Landoni et al., 2020), implicit emotional and vocabulary-based triggers in SE responses for individuals suffering from mental health disorders (Milton and Pera, 2023), etc. A profile in these works is operationalised as a vector comprising multiple components to represent the different perspectives of analysis. In our case, to examine a sample from the perspectives of Structure, Readability, and Concreteness, we create an *accessibility vector* which accounts for all the indicators described in Section 3.3 and is illustrated in Figure 2. An accessibility vector, however, provides insights into individual samples but not an IAS in general, which is the focus of our study. Hence, we create an overall *accessibility profile* of an IAS, $\mathbf{AP}(\texttt{log})$, by taking an average of text sample vectors of all IAS responses in a synthetic `log` (§3.2). Note that an accessibility profile in our study is not meant to serve as a set of metrics that can determine whether an IAS is accessible for autistic users. Instead, the accessibility profile of an IAS provides an overview of different (text-based) aspects of the system when responding to the information needs of a user group that are known to impact autistic individuals' information seeking.

Considering an SE provides multiple results for a user's query, we create an accessibility profile for an SE `log` by first averaging text sample vectors grouped by the query, and then computing the average of the resultant query-wise vectors. For LLM agents, since a `log` contains only one response per query, we create a profile by taking an average of all the text sample vectors in the `log`. We are also interested in the impact of query length and topic of inquiry on the text accessibility of IAS responses, hence we also generate (i) *n-gram profile*, $\mathbf{AP}_{n-gram}(\texttt{log})$, by averaging the text sample vectors (query-wise vectors in case of SE-based logs) of responses to queries of $n - gram$ length; and (ii) *inquiry topic profile*, $\mathbf{AP}_{topic}(\texttt{log})$, by taking an average of the text sample vectors of responses to queries related to an inquiry topic category.

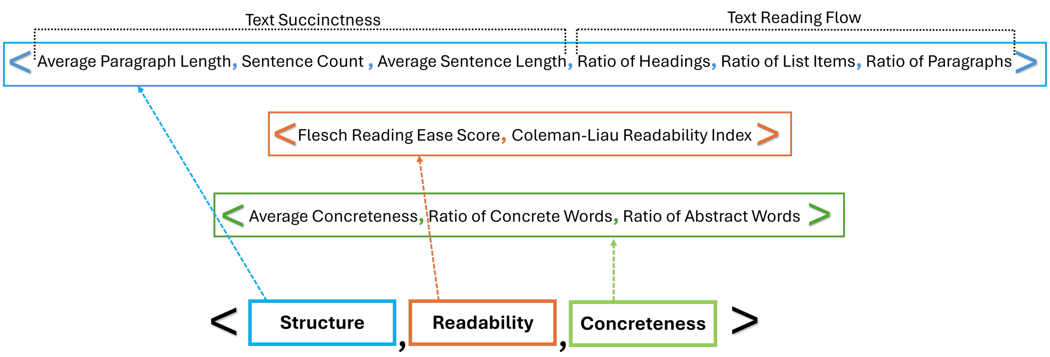


Figure 2: Structure of the accessibility vector representing text accessibility of each text sample.

Example. To illustrate the creation of a profile for a given IAS, consider a query set $\texttt{Q}_\texttt{E}$, comprising queries "banana", "white melon", and "motorcycle". We use the query set to build a synthetic SE based `log`: $\texttt{Q}_\texttt{E}$`-G-SERP`, consisting of Google's top 10 SERP results (i.e., web title and snippet of top 10 retrieved webpages) per query in $\texttt{Q}_\texttt{E}$. To build an overall accessibility profile of Google using this `log`, we first average the accessibility vectors of results per query, resulting in 3 query-wise vectors. We then take an average of the 3 vectors to obtain AP($\mathtt{Q_E}$-`G-SERP`). To create a 1-gram profile based on the same log, we compute the average of vectors corresponding to responses for queries "banana" and "motorcycle" respectively, and the 2 resultant query-wise vectors are averaged to create AP_{1-gram}($\mathtt{Q_E}$-`G-SERP`). For an inquiry topic profile pertaining to results for queries in the category "fruit", we average the vectors of results for queries "banana" and "white melon" respectively, and then the two vectors are averaged to obtain AP_{fruit}($\mathtt{Q_E}$-`G-SERP`).

4 Examining the Accessibility of IAS Responses for Autistic Users

Here, we describe the results of the experiments we conducted to answer **RQ1**.

4.1 Variation across IAS

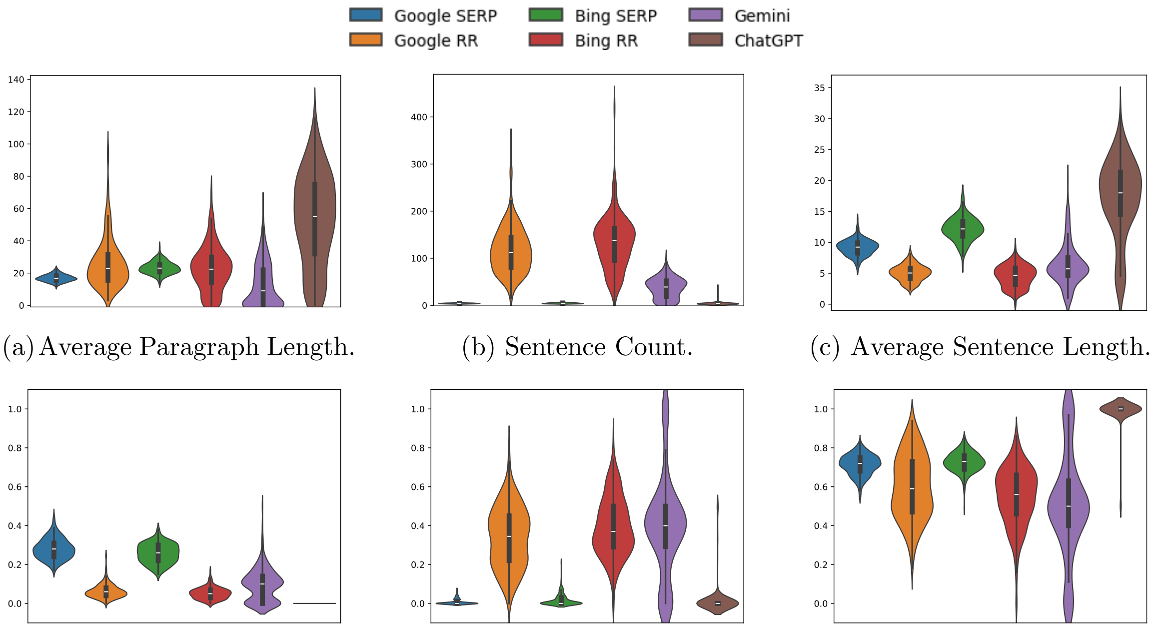
To assess the overall text accessibility of the considered SEs and LLM agents when responding to autistic individuals' inquiries, we juxtapose the accessibility profiles of all $\mathtt{Q_A}$ based logs reported in Table 4 at the component level. Note that unless stated otherwise, accessibility indicator scores are significantly different (one-way ANOVA $p < 0.05$) across the profiles.

Accessibility Perspective	Accessibility Indicator	AP($\mathtt{Q_A}$-G-SERP)	AP($\mathtt{Q_A}$-G-RR)	AP($\mathtt{Q_A}$-B-SERP)	AP($\mathtt{Q_A}$-B-RR)	AP($\mathtt{Q_A}$-GEM-ANS)	AP($\mathtt{Q_A}$-GPT-ANS)
Structure	ParaLen ↓	16.86	26.11	23.18	24.8	**12.49**	52.87
	SC ↓	**4.08**	117.61	4.43	133.16	37.01	4.17
	SentLen ↓	9.14	5.03	12.16	**4.52**	6.27	17.06
	HeadingSents ↑	**0.28**	0.06	0.26	0.05	0.09	0
	ListSents ↑	0.01	0.34	0.02	0.39	**0.4**	0.02
	ParaSents ↓	0.71	0.6	0.73	0.55	**0.51**	0.98
Readability	FRES ↑	**61.25**	58.78	57.43	59.01	42.06	47.03
	CLI ↓	**10.83**	11.89	11.74	11.79	13.83	12.13
Concreteness	Conc ↑	2.41	**2.43**	2.36	2.42	2.35	2.39
	ConcreteWords ↑	0.04	0.04	0.04	0.04	0.05	**0.06**
	AbstractWords ↓	0.43	0.39	0.54	**0.37**	0.48	0.68

Table 4: Accessibility Profiles of Synthetic Logs based on $\mathtt{Q_A}$. Each indicator score is significantly different (ANOVA one-way, $p < 0.05$) across logs. The best score of each accessibility indicator is **bolded**, where ↓ next to an indicator implies lower values to be better, and ↑ implies higher values to be better.

Structure. Comparing the components pertaining to text succinctness in the profiles from Table 4 (illustrated in Figures 3a-3c), we find that AP($\mathtt{Q_A}$-`GEM-ANS`) has the lowest average paragraph length, and AP($\mathtt{Q_A}$-`GPT-ANS`) the highest. Amongst the logs obtained from SEs, AP($\mathtt{Q_A}$-`G-SERP`) has notably shorter average paragraph length compared to AP($\mathtt{Q_A}$-`G-RR`), while the score is similar for AP($\mathtt{Q_A}$-`B-SERP`) and AP($\mathtt{Q_A}$-`B-RR`). On a sentence level, AP($\mathtt{Q_A}$-`G-RR`) and AP($\mathtt{Q_A}$-`B-RR`) have the highest sentence count but the shortest average sentence length. In fact, we observe an inverse relation between the sentence count and average sentence length across all profiles, which suggests a trade-off between the number of sentences and their length on average, perhaps to maintain a balance in overall text succinctness across all IAS responses.

In the case of components related to reading flow (Figures 3d-3f), the ratio of paragraphs is higher than the ratios of headings and list items across all profiles, particularly in AP($\texttt{Q}_\texttt{A}$-`GPT-ANS`). In AP($\texttt{Q}_\texttt{A}$-`G-SERP`) and AP($\texttt{Q}_\texttt{A}$-`B-SERP`), the sentences are mostly divided as paragraphs and then headings; in AP($\texttt{Q}_\texttt{A}$-`G-RR`), AP($\texttt{Q}_\texttt{A}$-`B-RR`), and AP($\texttt{Q}_\texttt{A}$-`GEM-ANS`), sentences mostly constitute a paragraph followed by list items. This implies that compared to other IAS, ChatGPT responses on average comprise longer blocks of text, which could impede autistic individuals' reading flow. Regarding the structure of responses produced by the other IAS, the distribution of sentences across markup features is distinct in each profile, implying that each IAS uses a different strategy to facilitate reading flow in its produced responses.

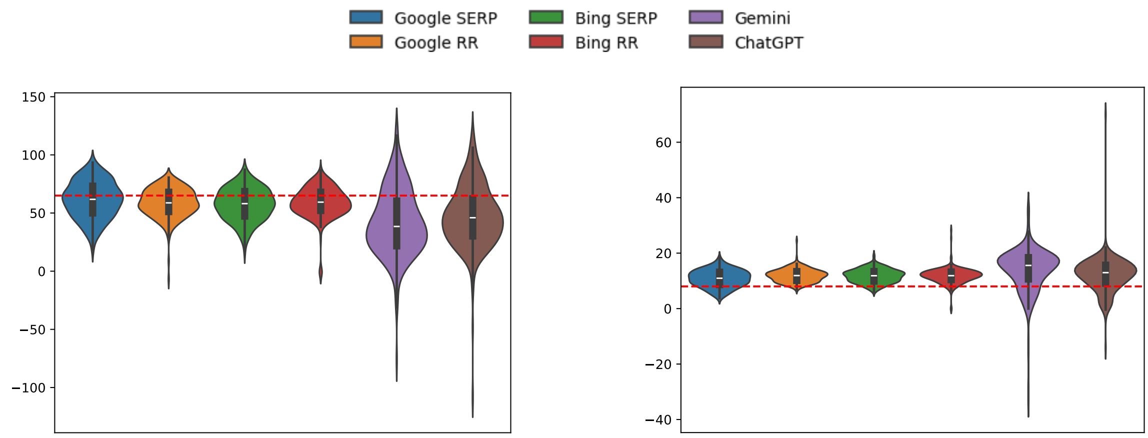


(a) Average Paragraph Length. (b) Sentence Count. (c) Average Sentence Length.

(d) Ratio of Headings. (e) Ratio of List Items. (f) Ratio of Paragraphs.

Figure 3: Text structure scores for logs generated using $\texttt{Q}_\texttt{A}$.

Readability. Yaneva et al. (2015) suggest that for a text to be readable by an autistic person, it must have a FRES score ≥ 65, which is text suitable for a student at grade 8 or 9 in the US school system (Flesch, 2016). In the case of CLI, the integer closest to the computed CLI score indicates the minimum US school grade a reader must have graduated from to be able to read the scored text, so for a text to be readable by an 8^{th} or 9^{th} grade student, the text should have a CLI score ≤ 8. When considering the FRES and CLI scores of the profiles from Table 3, none of the profiles meet either of the readability criteria to be deemed suitable for autistic individuals. From Figures 4a and 4b illustrating the distribution of readability scores of IAS responses within each log, we also note that most responses do not meet the readability criteria. We do however observe notable differences across the profiles with AP($\texttt{Q}_\texttt{A}$-`G-SERP`) having the highest FRES score and lowest CLI score, whereas AP($\mathtt{Q_A}$-GEM-ANS) has the lowest FRES score and highest CLI score. This implies that while none of the IASs produce responses that would be ideal for autistic individuals' reading ability, snippets in Google SERP, in general, are the easiest to read and Gemini answers the most challenging.

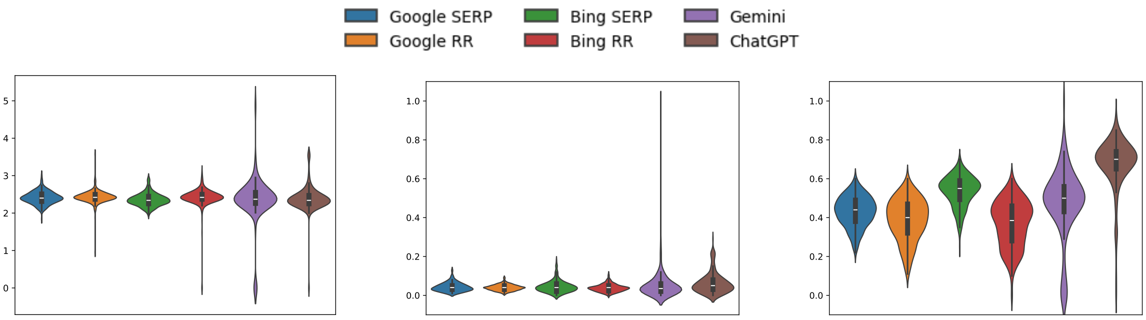


(a) FRES. (b) CLI.

Figure 4: Text readability scores for logs generated using $\mathtt{Q_A}$. The dotted red line illustrates the minimum FRES score in (a) and maximum CLI score in (b) an IAS response must have to be considered readable for an autistic person.

Concreteness. Comparing the profiles in Table 4 from the concreteness perspective does not highlight any notable differences. All profiles have an average concreteness score ranging between 2.35 for AP($\mathtt{Q_A}$-GEM-ANS) and 2.43 for AP($\mathtt{Q_A}$-G-RR) (Figure 5a). Moreover, from the scores of each profile for the ratio of concrete and abstract words respectively, we observe that all profiles have a higher ratio of abstract words than concrete words, conveying that IAS responses to autistic users' inquiries are, on average, more abstract than concrete in nature.

Examining the ratio of concrete and abstract words (Figures 5b and 5c) collectively in each profile, we find that AP($\mathtt{Q_A}$-GPT-ANS) has the highest ratio of words with a concreteness rating followed by AP($\mathtt{Q_A}$-B-SERP) and AP($\mathtt{Q_A}$-GEM-ANS); whereas AP($\mathtt{Q_A}$-G-RR) and AP($\mathtt{Q_A}$-B-RR) have the lowest ratio of words with a concreteness rating. Provided the lexicon we use to compute the Concreteness indicator scores consists of over 40000 commonly known English lemmas (Brysbaert et al., 2014), we infer that the LLM agent responses consist of more commonly known English words on average compared to SE responses, except for Bing SERP snippets. Considering that autistic individuals find it easier to comprehend a text written using vocabulary they are familiar with (Britto and Pizzolato, 2016; WebAIM, 2025), we infer that although resources retrieved by Google and Bing, in general, have a higher average concreteness than LLM agent responses, autistic individuals would find it easier to comprehend the familiar words in the answers generated by the agents compared to the unfamiliar words in the SE retrieved resources. This implies that autistic individuals' may perceive the information provided in an LLM agent to be more tangible compared to the information provided in resources retrieved by SE and the corresponding snippets presented in their SERP.



(a) Average Concreteness. (b) Ratio of Concrete Words. (c) Ratio of Abstract Words.

Figure 5: Text concreteness scores for logs generated using $\mathtt{Q_A}$.

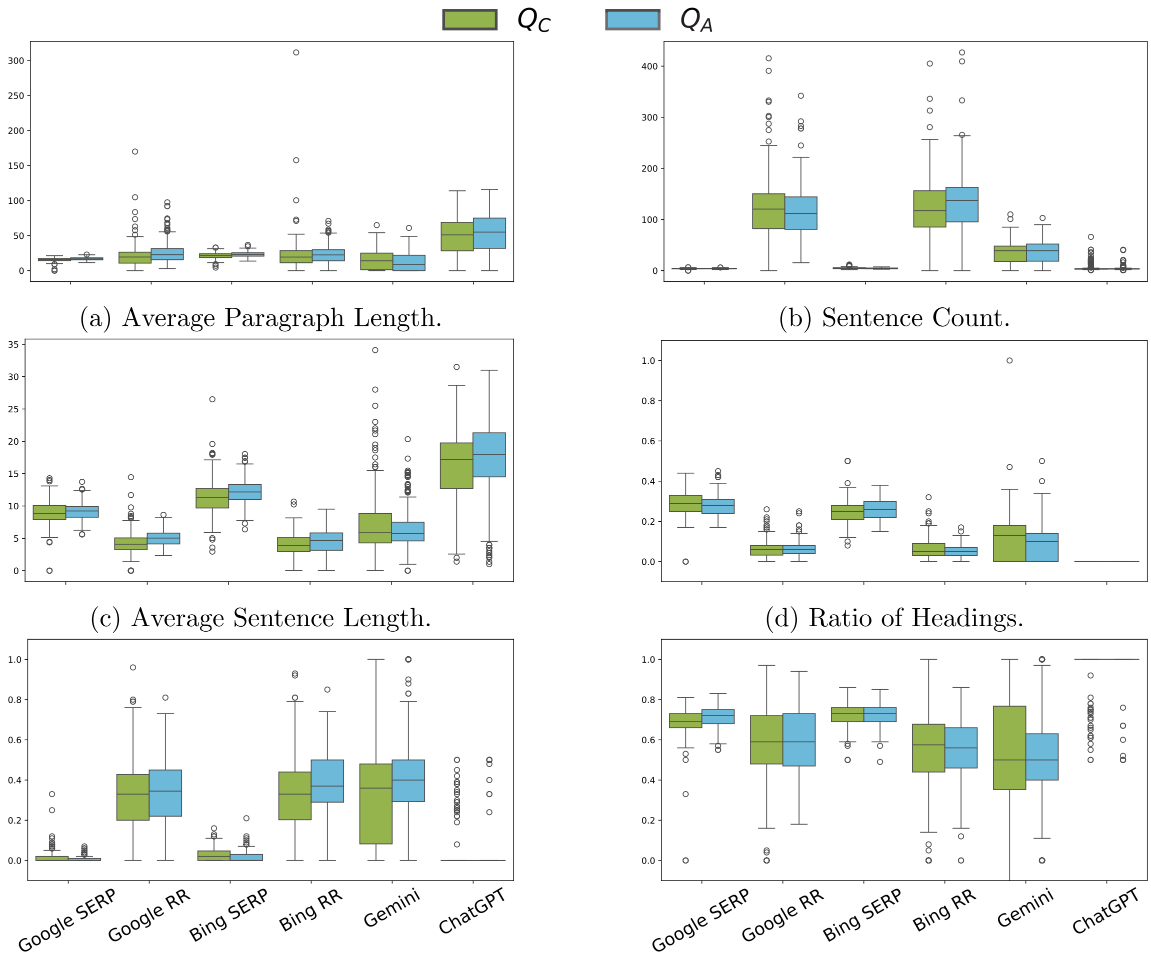
4.2 Impact across User Groups

To contextualise whether text accessibility-based aspects of IAS responses studied in this work are consistent regardless of which user group the information need originates from, we compare the accessibility profile generated for a $\mathtt{Q_A}$-based log with the profile generated for its corresponding $\mathtt{Q_C}$-based log (reported in Table 5 and illustrated in Figures 6-8), using an independent T-test ($p < 0.05$) for significance testing.

IAS	Synthetic Log	Structure						Readability		Concreteness		
		Para Len ↓	SC ↓	Sent Len ↓	Heading Sents ↑	List Sents ↑	Para Sents ↓	FRES ↑	CLI ↓	Conc ↑	Concrete Words ↑	Abstract Words ↓
Google SERP	$\mathtt{Q_A}$-G-SERP	16.86	4.08	9.14	0.28	0.01	0.71	61.25	10.83	2.41	0.04	0.43
	$\mathtt{Q_C}$-G-SERP	**15.57 ***	**4**	**8.93**	**0.29**	**0.02 ***	**0.68 ***	**63.01**	**11.03**	**2.43**	**0.05 ***	**0.31 ***
Google RR	$\mathtt{Q_A}$-G-RR	**26.11**	**117.61**	5.03	0.06	**0.34**	0.6	**58.78**	**11.89**	2.43	0.04	0.39
	$\mathtt{Q_C}$-G-RR	26.95	122.48	**4.24 ***	0.06	0.32	**0.58**	56.44	12.58	2.43	0.04	**0.28 ***
Bing SERP	$\mathtt{Q_A}$-B-SERP	23.18	**4.43**	12.16	**0.26**	0.02	0.73	57.43	**11.74**	2.36	0.04	0.54
	$\mathtt{Q_C}$-B-SERP	**21.23 ***	4.9 *	**11.38 ***	0.25	**0.03 ***	**0.72**	**61.55 ***	11.9	**2.5 ***	**0.06 ***	**0.37 ***
Bing RR	$\mathtt{Q_A}$-B-RR	24.8	133.16	4.52	0.05	**0.39**	0.55	**59.01**	11.79	**2.42**	0.04	0.37
	$\mathtt{Q_C}$-B-RR	**21.85**	**120.62 ***	**3.98 ***	**0.07 ***	0.33 *	**0.53**	56.58	**10.98 ***	2.36	0.04	**0.26 ***
Gemini	$\mathtt{Q_A}$-GEM-ANS	**12.49**	**37.01**	**6.27**	0.09	**0.4**	**0.51**	42.06	13.83	2.35	0.05	0.48
	$\mathtt{Q_C}$-GEM-ANS	15.77 *	42.32	7.21 *	**0.12 ***	0.34	0.54	**50.5 ***	**12.49**	**2.41**	0.05	**0.38 ***
ChatGPT	$\mathtt{Q_A}$-GPT-ANS	52.87	**4.17**	17.06	0	0.02	0.98	47.03	12.13	2.39	0.06	0.68
	$\mathtt{Q_C}$-GPT-ANS	**49**	5.45 *	**15.98 ***	0	**0.03**	**0.97**	**55.2 ***	**11.05 ***	**2.49 ***	**0.08 ***	**0.57 ***

Table 5: Accessibility Profiles of Synthetic Logs based on $\mathtt{Q_A}$ and $\mathtt{Q_C}$ respectively. An asterisk (*) on the $\mathtt{Q_C}$ based log denotes a significant difference (independent T-test $p < 0.05$) in indicator score compared to the corresponding $\mathtt{Q_A}$ based log. The better indicator score per compared log pair is **bolded**, where ↓ next to an indicator implies lower values to be better, and ↑ implies higher values to be better.

SE Responses. Looking at the profiles generated for Google SERP based logs produced for autistic individuals and the general population's inquiries in Table 5, we find that AP($\mathtt{Q_C}$-G-SERP) has a significantly lower average paragraph length, sentence count, and average sentence length than AP(Q_A-G-SERP). AP(Q_C-G-SERP) also has a significantly lower ratio of paragraphs and a significantly higher ratio of list items compared to AP(Q_A-G-SERP). From a Concreteness perspective, AP(Q_C-G-SERP) has a significantly higher ratio of concrete words and a significantly lower ratio of abstract words. Even the average concreteness of AP(Q_C-G-SERP) is higher than AP(Q_A-G-SERP) but the difference is not significant. From these results, we infer that the snippets in a Google SERP produced for inquiries by the general population are, on average, more succinct with an easier reading flow and also more concrete than the snippets produced for autistic individuals' inquiries. In terms of Readability, AP(Q_C-G-SERP) has a higher FRES score but also a higher CLI score than AP(Q_A-G-SERP), although the differences don't hold true in practice, from which we infer that snippets produced for autistic individuals and general population inquiries are similarly challenging to read for autistic individuals.



(a) Average Paragraph Length.

(b) Sentence Count.

(c) Average Sentence Length.

(d) Ratio of Headings.

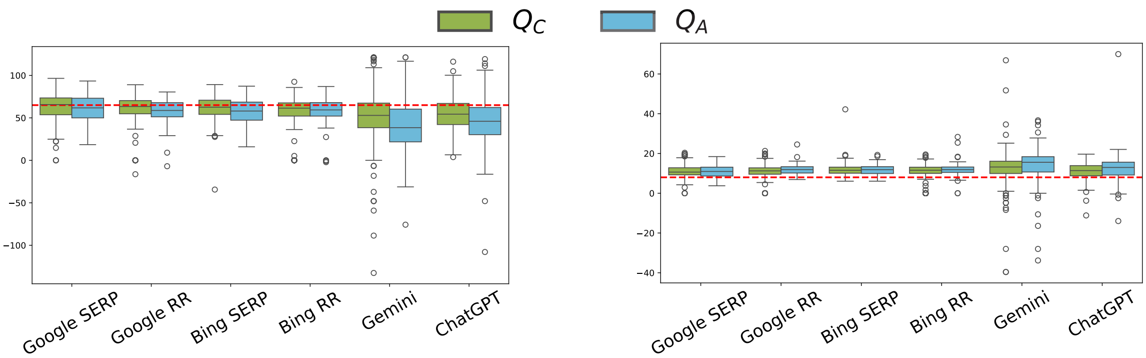
(e) Ratio of List Items.

(f) Ratio of Paragraphs.

Figure 6: Text structure scores for logs generated using Q_A and Q_C respectively.

In the case of profiles created for Google RR based logs, AP(Q_A-G-RR) has a lower average paragraph length and lower number of sentences than AP(Q_C-G-RR) as well as higher FRES score and lower CLI score than AP($\mathtt{Q_C}$-G-SERP), but none of the differences are statistically significant. Therefore, we infer that in practice, the resources retrieved by Google in response to inquiries by autistic individuals and the general population are similarly accessible for autistic individuals.

Comparing the accessibility profiles generated for Bing responses reported in Table 5, we find that AP($\mathtt{Q_C}$-B-SERP) has significantly lower average paragraph and sentence lengths, but significantly higher sentence count than AP($\mathtt{Q_A}$-B-SERP). In terms of reading flow, AP($\mathtt{Q_C}$-B-SERP) has a higher ratio of list items and a lower ratio of paragraphs, but also a lower ratio of headings than AP($\mathtt{Q_A}$-B-SERP), although the difference in the ratio of paragraphs does not hold true in practice. From the Readability perspective, AP($\mathtt{Q_C}$-B-SERP) has a significantly higher FRES score but also a higher CLI score than AP($\mathtt{Q_A}$-B-SERP), although the difference in CLI score is not significant. Based on the Concreteness perspective, AP($\mathtt{Q_C}$-B-SERP) has a higher average concreteness, a higher ratio of concrete words, and a lower ratio of abstract words than AP($\mathtt{Q_A}$-B-SERP), with differences across all three components being statistically significant. Based on these results, we infer that Bing SERP snippets are, on average, more succinct with better reading flow, easier to read, and more concrete when produced for the general population's inquiries compared to the snippets produced for autistic individuals' inquiries.



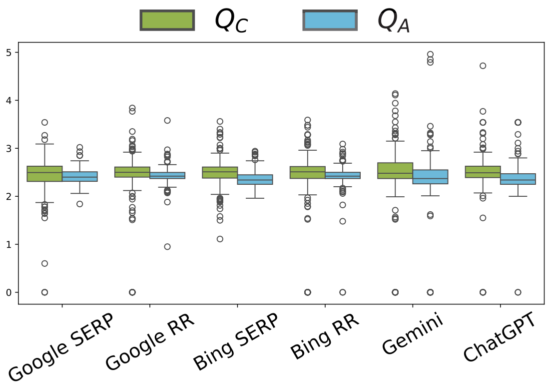
(a) FRES.

(b) CLI.

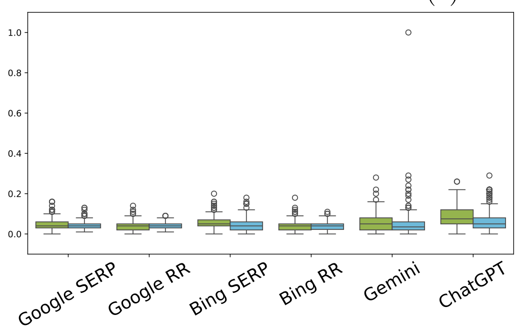
Figure 7: Text readability scores for logs generated using $\mathtt{Q_A}$ and $\mathtt{Q_C}$ respectively. The dotted red line illustrates the minimum FRES score in (a) and maximum CLI score in (b) an IAS response must have to be considered readable for an autistic person.

The difference in Bing responses becomes even more prominent upon comparing AP($\mathtt{Q_C}$-B-RR) and AP($\mathtt{Q_A}$-B-RR) reported in Table 5, due to the large margin of difference between the two profiles across most component values. AP($\mathtt{Q_C}$-B-RR) has a significantly lower sentence count and a significantly lower average sentence length than AP($\mathtt{Q_A}$-B-RR). In terms of reading flow, AP($\mathtt{Q_C}$-B-RR) has a significantly higher ratio of headings and a significantly lower ratio of paragraphs, but also a significantly lower ratio of list items than AP($\mathtt{Q_A}$-B-RR). From a Readability perspective, AP($\mathtt{Q_C}$-B-RR) has a significantly higher CLI score but a lower FRES score than AP($\mathtt{Q_A}$-B-RR), although the difference in FRES score does not hold true in practice. In terms of Concreteness, AP($\mathtt{Q_A}$-B-RR) has a higher average concreteness than AP($\mathtt{Q_C}$-B-RR) but the difference is not statistically significant; whereas AP($\mathtt{Q_C}$-B-RR) has a significantly lower ratio of abstract words than AP($\mathtt{Q_A}$-B-RR). Based on these results, we find that the resources retrieved by Bing for inquiries initiated by the general population are, in general, more succinct with easier reading flow, easier to read, as well as more concrete for autistic individuals than the resources retrieved for autistic individuals' inquiries.

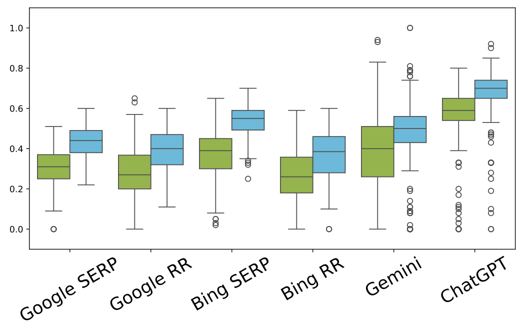
LLM Agent Responses. Comparing the profiles generated for Gemini responses to autistic individuals and the general population's inquiries, i.e., AP($\mathtt{Q_A}$-GEM-ANS) and AP($\mathtt{Q_C}$-GEM-ANS) reported in Table 5, we find that AP($\mathtt{Q_C}$-GEM-ANS) has higher FRES score and lower CLI score, as well as higher average concreteness and lower ratio of abstract words than AP($\mathtt{Q_A}$-GEM-ANS). This implies that from the Readability and Concreteness perspectives, Gemini generates, on average, more accessible responses for the general population's inquiries than for autistic individuals' inquiries. Still, the margin of difference is small and not significant except for the significant difference in the ratio of abstract words. From the Structure perspective, on the other hand, AP($\mathtt{Q_A}$-GEM-ANS) has significantly lower average paragraph and sentence lengths and significantly lower sentence count than AP($\mathtt{Q_C}$-GEM-ANS). These outcomes indicate that Gemini generates more succinct responses with easier reading flow for autistic users' inquiries compared to the general population's inquiries.



(a) Average Concreteness.



(b) Ratio of Concrete Words.



(c) Ratio of Abstract Words.

Figure 8: Text concreteness scores for logs generated using $\mathtt{Q_A}$ and $\mathtt{Q_C}$ respectively.

In the case of ChatGPT responses, AP($\mathtt{Q_C}$-GPT-ANS) has significantly lower average paragraph and sentence lengths than AP($\mathtt{Q_A}$-GPT-ANS). Even in terms of reading flow, AP($\mathtt{Q_C}$-GPT-ANS) has a significantly lower ratio of paragraphs and a significantly higher ratio of list items. These results imply that responses in $\texttt{Q}_\texttt{C}$-GPT-ANS on average are more succinct and have an easier reading flow than the responses in $\texttt{Q}_\texttt{A}$-GPT-ANS. From the Readability perspective, AP($\texttt{Q}_\texttt{C}$-GPT-ANS) has a significantly higher FRES score and significantly lower CLI score than AP($\texttt{Q}_\texttt{A}$-GPT-ANS), implying that ChatGPT generates easier to read responses for general population's inquiries compared to the responses generated for autistic individuals' inquiries. We observe a similar trend from the Concreteness perspective as well, with AP($\texttt{Q}_\texttt{C}$-GPT-ANS) having a significantly higher average concreteness and ratio of concrete words, and a significantly lower ratio of abstract words compared to AP($\texttt{Q}_\texttt{A}$-GPT-ANS). Putting all the results together, ChatGPT, like the other considered IAS, produces responses that are more aligned with text accessibility guidelines for autistic individuals when an inquiry is initiated by a user from the general population, compared to the responses generated for autistic individuals' inquiries.

4.3 Influence of Query Phrasing

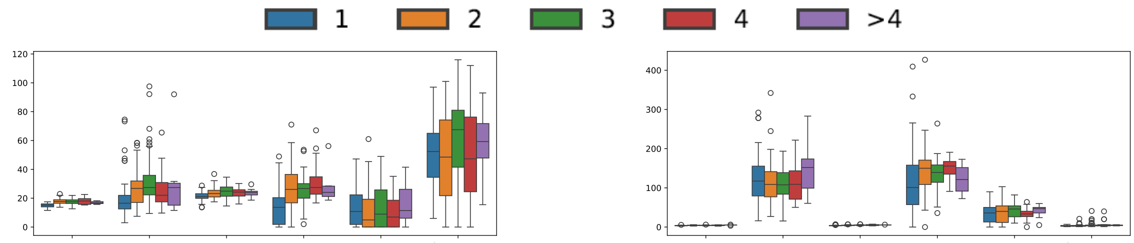
The manner in which a searcher formulates their query to articulate their information need strongly influences the responses produced by an IAS (Alaofi et al., 2022; Pera et al., 2023). Based on this, we explore the potential influence of variation in query phrasing in $\texttt{Q}_\texttt{A}$ on the text accessibility of elicited IAS responses. Particularly, we examine the impact of variation in (1) the **query length**, i.e., the number of words in the query, and (2) the **inquiry topic**, in our case a binary categorisation indicating whether the autistic users' inquiry pertains to ASD or general topics.

Query Length. We examine the variation in the text accessibility of IAS responses to autistic individuals' inquiries across five categories of query n-gram lengths: unigram, bigram, trigram, 4-gram, and queries longer than 4-gram. For this, we generate AP_{n-gram}(log) where log is an $\texttt{Q}_\texttt{A}$ based log as described in Section 3.2 and $n = 1, 2, 3, 4, > 4$. We compare the five profiles generated per log at the component level using one-way ANOVA ($p < 0.05$) to test for statistical significance.

We illustrate in Figure 9 the variation in the text structure scores of IAS responses across queries of different n-gram lengths. For a given IAS, there are statistically significant differences in Structure-related indicator scores across responses for different query lengths. However, we do not observe any prominent trends across or within a log, except for in AP_{n-gram}($\texttt{Q}_\texttt{A}$-GPT-ANS), where the ratios of headings, list items, and paragraphs remain consistent across the n-gram categories, which implies that the reading flow of ChatGPT responses is unaffected by the query length.

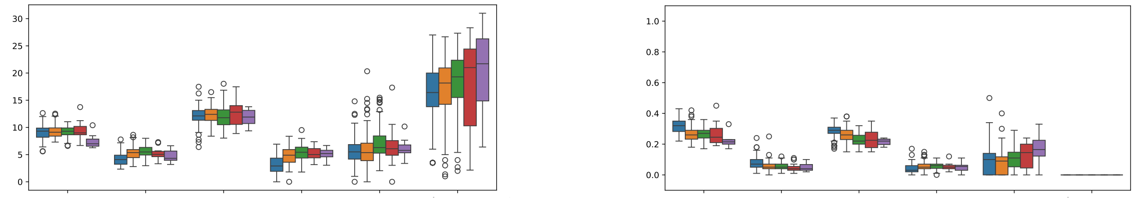
The sentence count in AP_{n-gram}($\texttt{Q}_\texttt{A}$-G-SERP), AP_{n-gram}($\texttt{Q}_\texttt{A}$-B-SERP), and AP_{n-gram}($\texttt{Q}_\texttt{A}$-GPT-ANS), respectively, also remains consistent across the n-gram categories. However, the average sentence and paragraph lengths are consistent only in AP_{n-gram}($\texttt{Q}_\texttt{A}$-G-SERP) and AP_{n-gram}($\texttt{Q}_\texttt{A}$-B-SERP) and not in AP_{n-gram}($\texttt{Q}_\texttt{A}$-GPT-ANS). This implies that the succinctness of Google and Bing SERP responses is unaffected by the query length, but in the case of ChatGPT responses, while the number of sentences remains unaffected on average, the lengths of the sentences and paragraphs vary across query lengths.

Figure 10 illustrates the impact of query length variation on the Readability indicator scores of IAS responses. From the figure, we find that the FRES score decreases significantly and the CLI score increases significantly in AP_{n-gram}($\texttt{Q}_\texttt{A}$-GPT-ANS) and AP_{n-gram}($\texttt{Q}_\texttt{A}$-GEM-



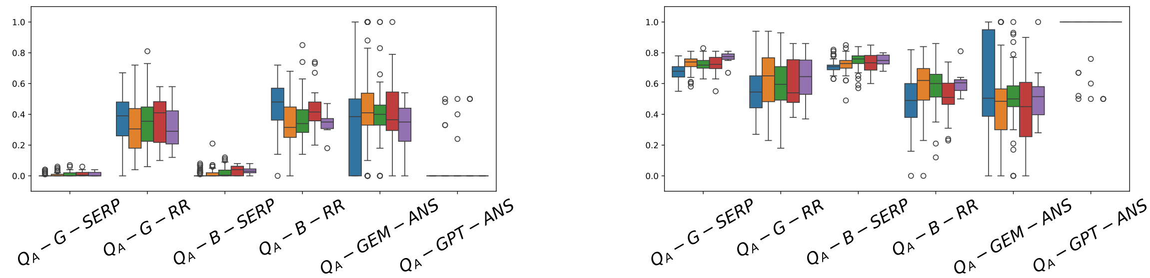
(a) Average Paragraph Length. The difference across n-gram lengths is not significant for $\texttt{Q}_\texttt{A}$-B-RR.

(b) Sentence Count. The difference across n-gram lengths is not significant for $\texttt{Q}_\texttt{A}$-G-RR and $\texttt{Q}_\texttt{A}$-GEM-ANS, respectively.



(c) Average Sentence Length. The difference across n-gram lengths is not significant for $\texttt{Q}_\texttt{A}$-B-SERP and $\texttt{Q}_\texttt{A}$-GPT-ANS, respectively.

(d) Ratio of Headings. The difference across n-gram lengths is not significant for $\texttt{Q}_\texttt{A}$-GPT-ANS.

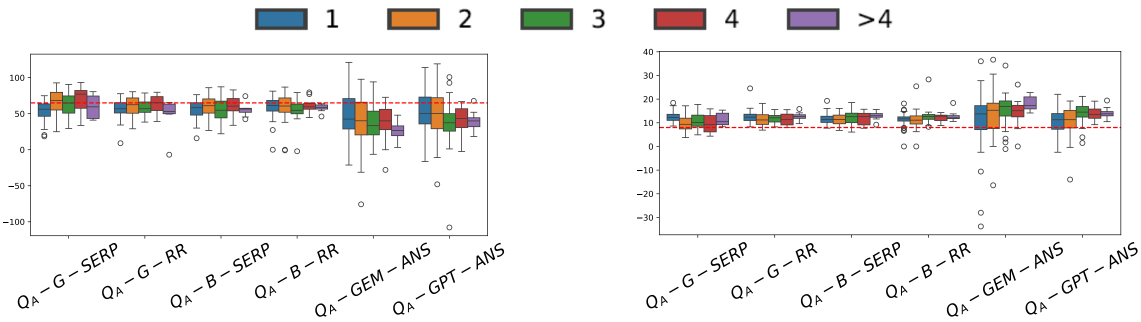


(e) Ratio of List Items. The difference across n-gram lengths is not significant for $\texttt{Q}_\texttt{A}$-G-SERP and $\texttt{Q}_\texttt{A}$-G-RR, respectively.

(f) Ratio of Paragraphs.

Figure 9: Text structure scores for $\texttt{Q}_\texttt{A}$ based logs across query n-gram length categories. The difference in indicator scores across n-gram lengths is significant (one-way ANOVA, $p < 0.05$) for a log unless stated otherwise.

ANS) respectively as the query length increases. This implies that for longer queries, both ChatGPT and Gemini generate responses that are harder to read. In case of Bing responses, FRES score in $\text{AP}_{n-gram}(\texttt{Q}_\texttt{A}\text{-B-SERP})$ and $\text{AP}_{n-gram}(\texttt{Q}_\texttt{A}\text{-B-RR})$ is notably lower and CLI score notably higher when $n > 4$, although only the CLI score difference in $\text{AP}_{n-gram}(\texttt{Q}_\texttt{A}\text{-B-RR})$ holds true in practice. For Google responses, the FRES score across $\text{AP}_{n-gram}(\texttt{Q}_\texttt{A}\text{-G-SERP})$ and $\text{AP}_{n-gram}(\texttt{Q}_\texttt{A}\text{-G-RR})$ is significantly lower when $n = 1$ and $n > 4$. The CLI score in $\text{AP}_{n-gram}(\texttt{Q}_\texttt{A}\text{-G-SERP})$ and $\text{AP}_{n-gram}(\texttt{Q}_\texttt{A}\text{-G-RR})$ is also prominently higher when $n = 1$ and $n > 4$, however the difference is significant only in $\text{AP}_{n-gram}(\texttt{Q}_\texttt{A}\text{-G-SERP})$. These trends suggest that in Bing, longer queries lead to harder-to-read responses, while in Google, extremely short or long queries similarly result in harder-to-read responses.



(a) FRES. The difference across n-gram lengths is not significant for Q$_A$-B-SERP and Q$_A$-B-RR.

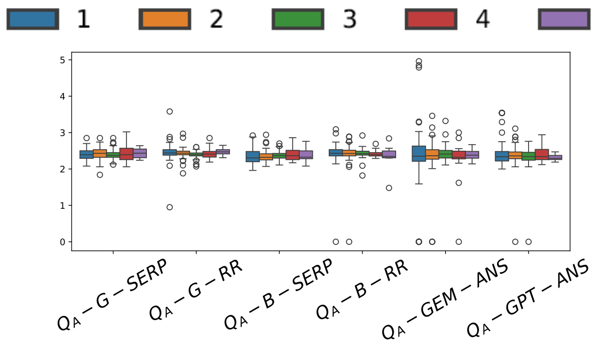
(b) CLI. The difference across n-gram lengths is not significant for Q$_A$-G-RR and Q$_A$-B-SERP.

Figure 10: Text readability for Q$_A$ based logs across query n-gram length categories. The difference in indicator scores across n-gram lengths is significant (one way ANOVA, $p < 0.05$) for a log unless stated otherwise. The dotted red line illustrates the minimum FRES score in (a) and maximum CLI score in (b) an IAS response must have to be considered readable for an autistic person.

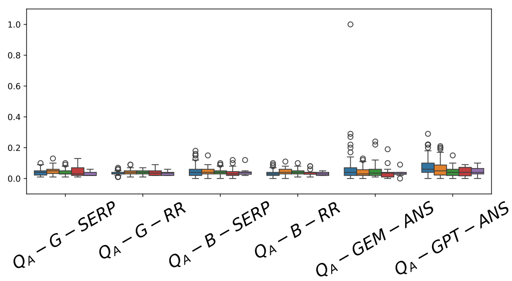
We depict the variation in the concreteness of IAS responses across varying query lengths in Figure 11 and observe notable differences only in the Ratio of Abstract words although the difference is not statistically significant in AP$_{n-gram}$(Q$_A$-GPT-ANS) and AP$_{n-gram}$(Q$_A$-GEM-ANS), respectively. In both AP$_{n-gram}$(Q$_A$-G-SERP) and AP$_{n-gram}$(Q$_A$-B-SERP), the ratio of abstract words is significantly lower when $n \geq 4$ whereas in AP$_{n-gram}$(Q$_A$-G-RR) and AP$_{n-gram}$(Q$_A$-B-RR), the ratio is significantly lower when $n = 1$. This implies that longer queries result in responses that have significantly fewer abstract words in web snippets produced by Google and Bing, but the full content in the corresponding retrieved resource is more abstract in nature than the resources retrieved by the SE for shorter queries.

Inquiry Topic. Considering the impact of ASD in several autistic individuals' quality of life (Lin and Huang, 2019; Mason et al., 2018), inquiry topics related to their condition could be considered a crucial information need for autistic users. Hence, as an initial exploration of the impact of inquiry topic on the accessibility of IAS responses, we investigate if the considered IASs respond differently to autistic users' inquiries related to their condition in comparison to inquiries on more generic topics. For this, we categorise each query in Q$_A$ as an ASD-related inquiry (**D**) or a General inquiry (**G**) based on the constituent keywords in the query. In other words, if a query contains an ASD-related term, it is categorised as an ASD-related inquiry, and a General inquiry otherwise. Due to the lack of a standardised list of ASD-related terms at the time of this study, we rely on public educational resources for autistic people, specifically the glossary terms from ReframingAutism.org[14], a charity run by autistic people in Australia aimed to improve the understanding of Autism through education, resources, and research . The distribution of Q$_A$ queries between the two categories is illustrated in Figure 12. Based on these categories, we produce AP$_{topic}$(`log`) for

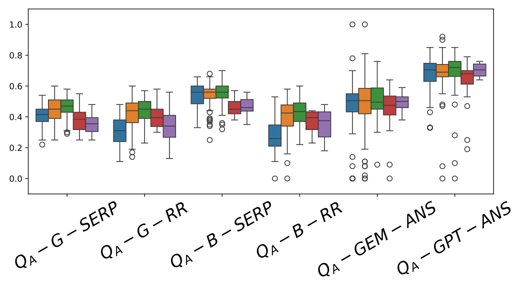
14. The terms were extracted from `https://reframingautism.org.au/service/glossary-terms/` in October 2024.



(a) Average Concreteness. The difference across n-gram lengths is not significant for any logs.



(b) Ratio of Concrete Words. The difference across n-gram lengths is not significant for $\mathtt{Q_A}$-B-SERP.



(c) Ratio of Abstract Words. The difference across n-gram lengths is not significant for $\mathtt{Q_A}$-GEM-ANS and $\mathtt{Q_A}$-GPT-ANS.

Figure 11: Text concreteness scores for $\mathtt{Q_A}$ based logs across query n-gram length categories. The difference in indicator scores across n-gram lengths is significant (one way ANOVA, $p < 0.05$) for a log unless stated otherwise.

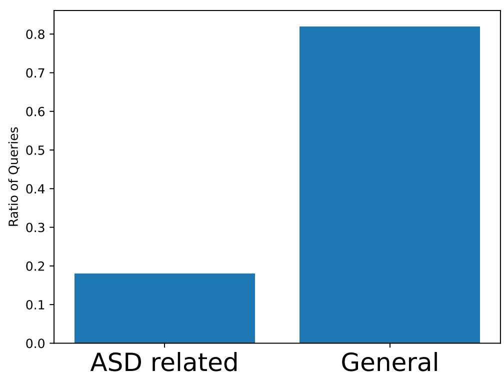


Figure 12: Distribution of $\mathtt{Q_A}$ queries between inquiry topic categories.

each $\mathtt{Q_A}$ based `log` with *topic* $= \{D, G\}$ and report the generated profiles in Table 6. The

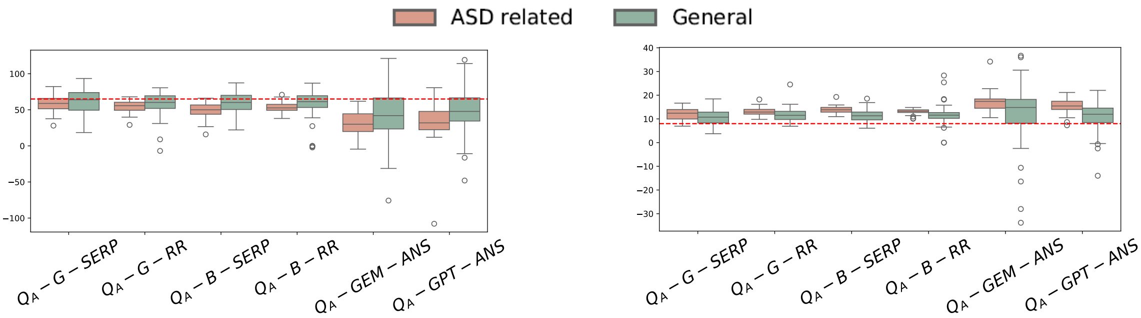
Synthetic Log	Inquiry Topic Category	Structure						Readability		Concreteness		
		Para Len ↓	SC ↓	Sent Len ↓	Heading Sents ↑	List Sents ↑	Para Sents ↓	FRES ↑	CLI ↓	Conc ↑	Concrete Words ↑	Abstract Words ↓
$\mathtt{Q_A}$-G-SERP	ASD	16.71	**3.99**	9.29	0.28	0.01	0.71	58.98	12	2.38	0.03	0.43
	General	**16.89**	4.1	**9.11**	0.28	0.01	0.71	**61.75**	**10.57 ***	**2.41**	**0.04 ***	0.43
$\mathtt{Q_A}$-G-RR	ASD	33.6	**115.97**	5.27	0.06	**0.44**	**0.5**	54.26	13	2.4	0.03	0.41
	General	**24.46 ***	117.97	**4.98**	0.06	0.32 *	0.62 *	**59.78 ***	**11.64 ***	2.44	**0.04 ***	**0.38**
$\mathtt{Q_A}$-B-SERP	ASD	24.45	4.48	12.83	0.25	0.02	0.73	49.48	13.76	2.33	0.04	0.54
	General	**22.9 ***	**4.42**	**12.01 ***	**0.26**	0.02	0.73	**59.18 ***	**11.29 ***	**2.36**	**0.05 ***	0.54
$\mathtt{Q_A}$-B-RR	ASD	41.58	139.8	5.29	**0.06**	**0.46**	**0.48**	53.77	13.13	2.4	0.03	0.4
	General	**21.12 ***	**131.7**	**4.35 ***	0.05 *	0.38 *	0.56 *	**60.17 ***	**11.49 ***	**2.42**	**0.04**	**0.36 ***
$\mathtt{Q_A}$-GEM-ANS	ASD	14.29	38.16	7.43	**0.11**	0.38	0.52	31.68	16.99	2.32	0.04	0.51
	General	**12.09**	**36.76**	**6.02 ***	0.09	**0.4**	**0.5**	**44.33 ***	**13.14 ***	**2.35**	**0.05**	**0.47**
$\mathtt{Q_A}$-GPT-ANS	ASD	**49.96**	5.67	17.24	0	**0.03**	**0.97**	32	16.52	2.26	0.04	**0.67**
	General	53.52	**3.84 ***	**17.03**	0	0.01	0.99	**50.32 ***	**11.17 ***	**2.41 ***	**0.07 ***	0.68

Table 6: $\text{AP}_D(\text{log})$ and $\text{AP}_G(\text{log})$ for all $\mathtt{Q_A}$-based logs. An asterisk (*) on the score for $\text{AP}_D(\text{log})$ denotes a significant difference (independent T-test $p < 0.05$) in the indicator score compared to the corresponding score in $\text{AP}_G(\text{log})$. The better indicator score per compared pair is **bolded**, where ↓ next to an indicator implies lower values to be better, and ↑ implies higher values to be better.

two profiles generated per `log` are compared at the component level using an independent T-test ($p < 0.05$).

From Table 6, we observe that particularly from the perspective of Readability $\text{AP}_G(\mathtt{log})$ is closer to autistic people's preferences than the corresponding $\text{AP}_D(\mathtt{log})$ (Figure 13) with the difference in both FRES and CLI scores statistically significant across all compared profile pairs, except for the difference in FRES score between $\text{AP}_G(\mathtt{Q_A\text{-}G\text{-}SERP})$ and $\text{AP}_D(\mathtt{Q_A\text{-}G\text{-}SERP})$. Even from the perspective of Concreteness, $\text{AP}_G(\mathtt{log})$ is better suited for autistic individuals than the corresponding $\text{AP}_D(\mathtt{log})$ (Figure 14), although not all indicator score differences between the compared pairs hold true in practice. The difference in average concreteness is statistically significant only between $\text{AP}_D(\mathtt{Q_A\text{-}GPT\text{-}ANS})$ and $\text{AP}_G(\mathtt{Q_A\text{-}GPT\text{-}ANS})$. The ratio of concrete words is *not* statistically significant between $\text{AP}_D(\mathtt{Q_A\text{-}B\text{-}RR})$ and $\text{AP}_G(\mathtt{Q_A\text{-}B\text{-}RR})$, and between $\text{AP}_D(\mathtt{Q_A\text{-}GEM\text{-}ANS})$ and $\text{AP}_G(\mathtt{Q_A\text{-}GEM\text{-}ANS})$, respectively. The ratio of abstract words is significantly different only between $\text{AP}_D(\mathtt{Q_A\text{-}B\text{-}RR})$ and $\text{AP}_G(\mathtt{Q_A\text{-}B\text{-}RR})$.

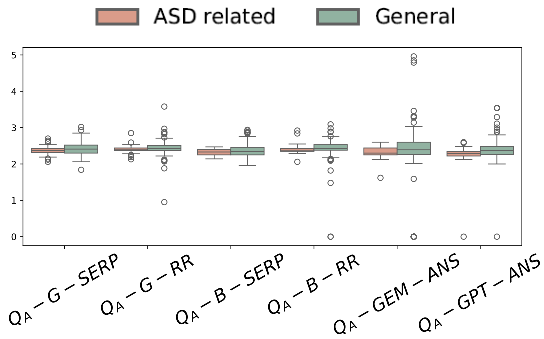
From the Structure perspective, we do not observe many significant differences between the compared profile pairs which is also prominent from the similar distributions across most components illustrated in Figure 15. The only exception is when we compare $\text{AP}_D(\mathtt{Q_A\text{-}B\text{-}RR})$ and $\text{AP}_G(\mathtt{Q_A\text{-}B\text{-}RR})$. In terms of text succinctness, $\text{AP}_G(\mathtt{Q_A\text{-}B\text{-}RR})$ has a significantly lower sentence count and significantly lower average paragraph and sentence lengths than $\text{AP}_D(\mathtt{Q_A\text{-}B\text{-}RR})$; whereas based on reading flow, $\text{AP}_D(\mathtt{Q_A\text{-}B\text{-}RR})$ has significantly higher ratios of headings and list items, respectively, along with a significantly lower ratio of paragraphs than $\text{AP}_G(\mathtt{Q_A\text{-}B\text{-}RR})$, implying that when autistic individuals' inquiries are related to ASD, Bing retrieves resources that are structured for easier reading flow, compared to resources retrieved for inquiries on general topics. Overall, we find the topic of inquiry does not influence the structure of IAS responses, except for the reading flow of resources retrieved by Bing.



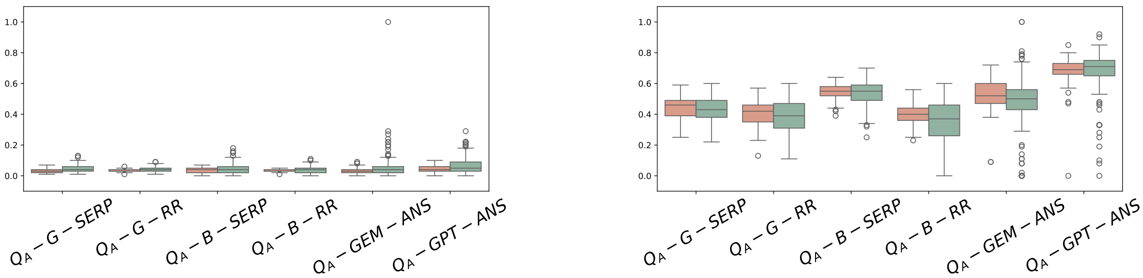
(a) FRES.

(b) CLI.

Figure 13: Text readability scores for $\mathtt{Q_A}$ based logs across inquiry topic categories. The dotted red line illustrates the minimum FRES score in (a) and maximum CLI score in (b) an IAS response must be considered readable for an autistic person.



(a) Average Concreteness.



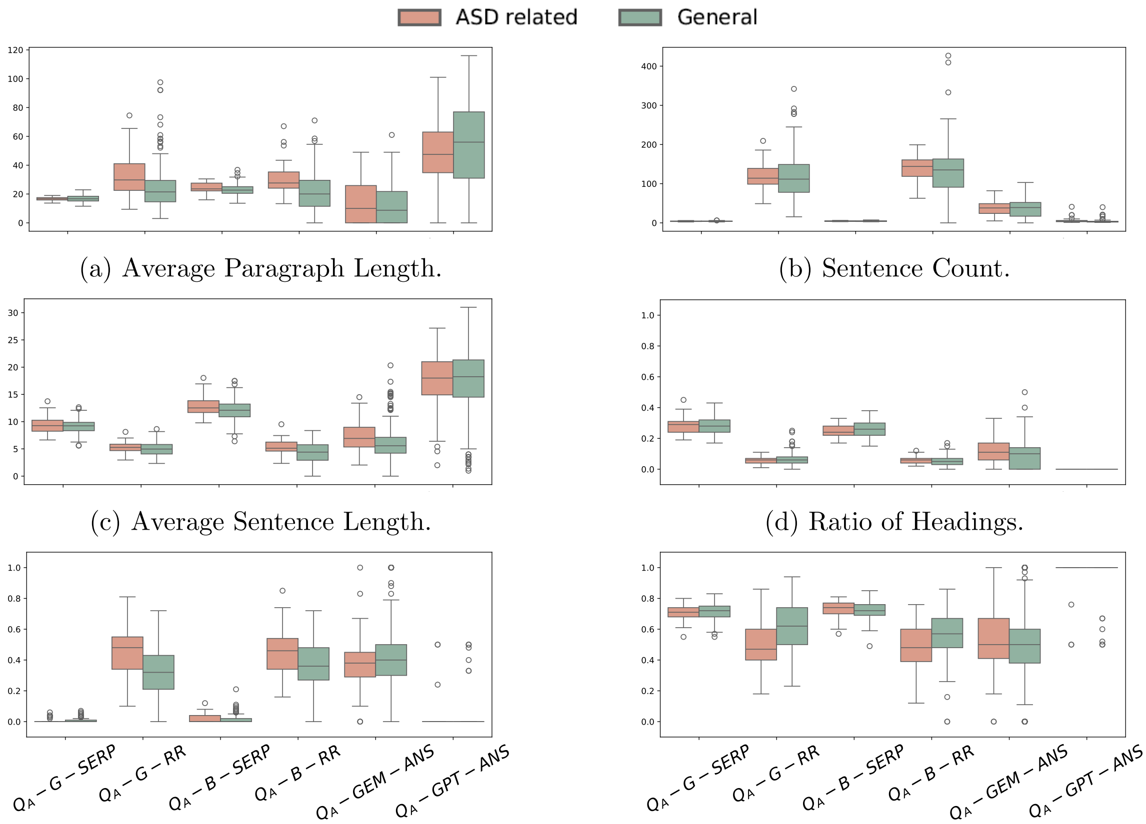
(b) Ratio of Concrete Words.

(c) Ratio of Abstract Words.

Figure 14: Text concreteness scores for $\mathtt{Q_A}$ based logs across inquiry topic categories.

4.4 Are the Responses of Mainstream SEs and LLM Agents Accessible to Autistic Users?

To answer RQ1, we scrutinised the Structure, Readability, and Concreteness of the textual content of IAS responses produced for autistic individuals' inquiries. We contextualised whether our observations are due to the systems' general algorithmic behaviour by also assessing the responses produced by the IAS under study for inquiries reflecting the in-



(a) Average Paragraph Length. (b) Sentence Count.

(c) Average Sentence Length. (d) Ratio of Headings.

(e) Ratio of List Items. (f) Ratio of Paragraphs.

Figure 15: Text structure scores for $\mathtt{Q_A}$ based logs across inquiry topic categories.

formation needs of the general population across the same three perspectives. Findings indicate that responses for autistic individuals' queries tend to be more verbose, harder to read, and more abstract in nature compared to responses to inquiries from the general population (§4.1 and §4.2). Further investigations whether the accessibility of an IAS response varies across different types of autistic users' inquiries reveal that responses are significantly harder to read when an autistic user's inquiry is related to ASD than when it is related to other topics (§4.3). Based on these observations, we argue that IAS responses for autistic individuals' inquiries are less closely aligned to the target audience's text accessibility expectations than responses produced for inquiries initiated by the general population.

5 Exploring the Impact of Diagnosis Disclosure in Query Phrasing

In this section, we discuss the results of the experiments we conducted to address **RQ2**.

5.1 Impact on IAS Responses

Knowing that autistic individuals who are aware of their diagnosis often state their condition in their online inquiry (Yechiam and Yom-Tov, 2021) and considering the influence of query formulation on biasing IAS responses (Alaofi et al., 2022; Pera et al., 2023), we investigate whether rephrasing the original synthetic queries in $\mathtt{Q_A}$ to explicitly state that the user is autistic (as described in Section 3.2) would notably change the response produced by the four IAS we consider in this study.

In the case of Bing and Google, we compare the resources retrieved for inquiries in $\mathtt{Q_A}$ and $\mathtt{Q'_A}$ respectively. For this, we use the Rank Biased Overlap (**RBO**) (Sarica et al., 2022), a measure that computes the similarity between two ranked lists of entities and returns a numeric value between 0 and 1, with 0 indicating that the lists are distinct and 1 indicating that the lists are the same. Specifically, we measure the similarity between SERP produced in response to each query in $\mathtt{Q_A}$ and its counterpart query in $\mathtt{Q'_A}$. Given that an SE may rank different landing pages of the same web resource differently for distinct query variations, we extract the homepage URL of each SERP result and compute the RBO for the two lists of homepages.

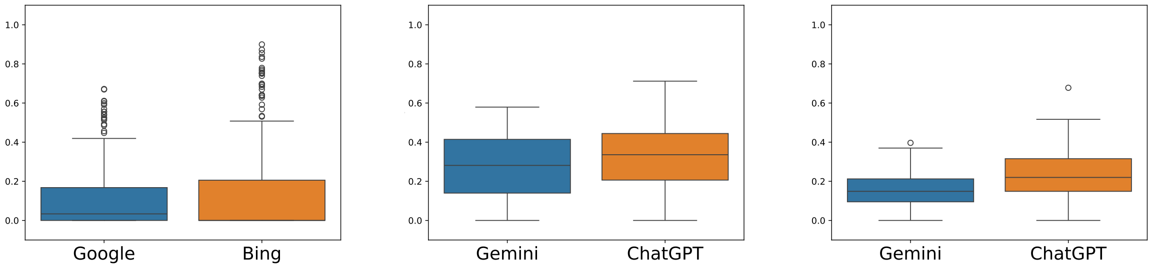
For Gemini and ChatGPT, we compare the direct answers generated by each agent for inquiries in $\mathtt{Q_A}$ and $\mathtt{P'_A}$ respectively. For this, we turn to the Recall-Oriented Understudy for Gisting Evaluation, i.e., ROUGE metrics (Lin, 2004) which measure the recall and precision of a generated summary compared to a (human-made) gold-standard summary, with the score ranging between 0 and 1, where a higher score implies higher similarity between the compared summaries. In our case, we consider the response generated by the LLM agents for queries in $\mathtt{Q_A}$ as the "gold standard" and compare it to the response generated by the agent for the corresponding query in $\mathtt{P'_A}$. Particularly, we measure the similarity between the compared responses—as a whole as well as at the sentence level—using the ROUGE-L metric due to its reported effectiveness on single document summarization tasks (Lin, 2004).

From Figure 16a, we observe that the RBO scores are distributed mostly between 0 and 0.2 for both Google and Bing, which implies that the resources retrieved for the rephrased queries are distinct from the resources retrieved for the original queries. In the case of LLM agent responses, ROUGE-L score at the summary level varies between 0.1 and 0.4 for Gemini and between 0.2 and 0.4 for ChatGPT responses (from Figure 16b), which also depicts low similarity between the LLM agent responses for the rephrased queries when compared to the responses for the original queries. Figure 16c further emphasises the difference at the sentence level, with ROUGE-L scores ranging between 0.1 and 0.2 for Gemini responses and between 0.2 and 0.3 for ChatGPT responses. Based on the similarity metric scores obtained, we infer that informing the IAS that the inquiry is from an autistic user yields notably different responses from all four considered IAS.

5.2 Influence on Text Accessibility

To investigate whether there is also a difference in the accessibility of the responses produced for the rephrased queries, we generate accessibility profiles for synthetic logs representing SE and LLM agent responses for queries in $\mathtt{Q'_A}$ and $\mathtt{P'_A}$, respectively. For each IAS, we juxtapose its profile, with respect to the counterpart for $\mathtt{Q_A}$ at the component level, as reported in Table 7 and illustrated in Figures 17-19; an independent T-test ($p < 0.05$) for statistical significance.

Google. To examine the impact of stating the user's condition in the inquiry on accessibility of Google responses, we compare AP($\mathtt{Q'_A}$-G-SERP) with AP($\mathtt{Q_A}$-G-SERP), and AP($\mathtt{Q'_A}$-G-RR)



(a) RBO score of Google and Bing.

(b) Summary-level ROUGE-L score for Gemini and ChatGPT.

(c) Sentence-level ROUGE-L score for Gemini and ChatGPT.

Figure 16: Similarity scores between IAS responses produced for different query phrasings.

IAS	Synthetic Log	Structure						Readability		Concreteness		
		Para Len ↓	SC ↓	Sent Len ↓	Heading Sents ↑	List Sents ↑	Para Sents ↓	FRES ↑	CLI ↓	Conc ↑	Concrete Words ↑	Abstract Words ↓
Google SERP	Q'_A-G-SERP	19.56	5.05	**8.55**	0.22	0	0.77	**75.56**	**7.2**	**2.54**	0.04	0.5
	Q_A-G-SERP	**16.86 ***	**4.08 ***	9.14 *	**0.28 ***	**0.01 ***	**0.71 ***	61.25 *	10.83 *	2.41 *	0.04	**0.43 ***
Google RR	Q'_A-G-RR	30.98	**96.4**	6.23	0.04	0.24	0.72	**69.08**	**9.75**	2.4	**0.05**	0.52
	Q_A-G-RR	**26.11 ***	117.61 *	**5.03 ***	**0.06 ***	**0.34 ***	**0.6 ***	58.78 *	11.89 *	**2.43 ***	0.04 *	**0.39 ***
Bing SERP	Q'_A-B-SERP	24.72	5.12	**11.58**	0.22	**0.03**	0.76	**58.98**	**11.44**	**2.42**	**0.05**	0.55
	Q_A-B-SERP	**23.18 ***	**4.43 ***	12.16 *	**0.26 ***	0.02 *	**0.73 ***	57.43	11.74	2.36 *	0.04	**0.54**
Bing RR	Q'_A-B-RR	34.91	139.96	6.12	0.05	0.37	0.57	57.89	11.94	2.41	0.04	0.46
	Q_A-B-RR	**24.80 ***	**133.16**	**4.52 ***	0.05	**0.39**	**0.55**	**59.01**	**11.79**	**2.42**	0.04 *	**0.37 ***
Gemini	P'_A-GEM-ANS	22.19	44.69	7.37	**0.13**	0.37	**0.5**	39.05	15.09	**2.36**	0.04	0.54
	Q_A-GEM-ANS	**12.49 ***	**37.01 ***	**6.27 ***	0.09 *	**0.4**	0.51	**42.06**	**13.83 ***	2.35	**0.05**	**0.48 ***
ChatGPT	P'_A-GPT-ANS	60.76	7.32	18.51	0	0.02	0.98	45.78	12.66	2.35	0.06	0.7
	Q_A-GPT-ANS	**52.87 ***	**4.17 ***	**17.06 ***	0	0.02	0.98	**47.03**	**12.13**	**2.39**	0.06	**0.68 ***

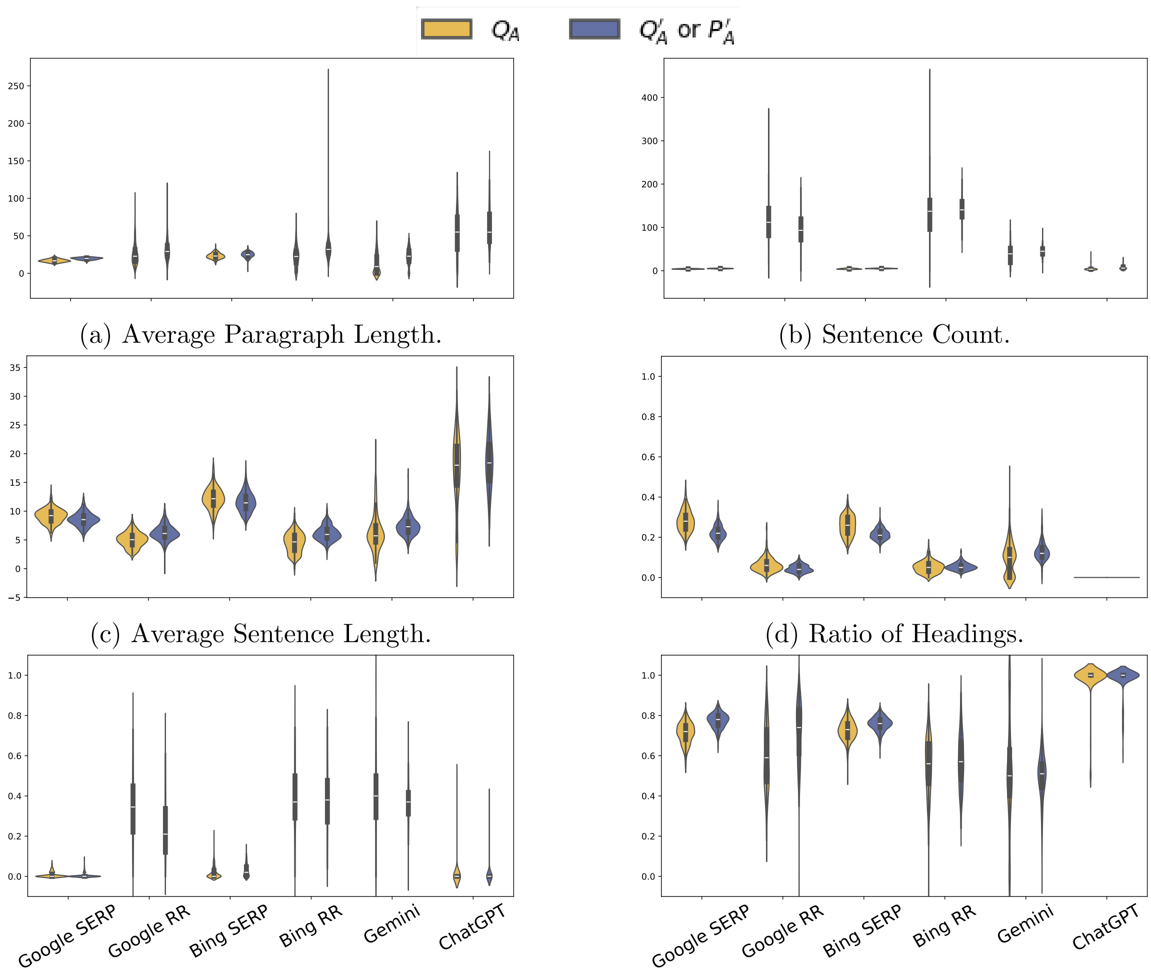
Table 7: Accessibility Profiles of Synthetic Logs based on Q'_A (or P'_A) and Q_A respectively. An asterisk (*) on the log based on Q_A denotes a significant difference (independent T-test $p < 0.05$) in indicator score compared to the corresponding log based on Q'_A or P'_A. The better indicator score per compared log pair is **bolded**, where ↓ next to an indicator implies lower values to be better, and ↑ implies higher values to be better.

with AP(Q_A-G-RR), as reported in Table 7. Note that all reported comparisons are statistically significant unless stated otherwise.

From the Structure perspective, in terms of reading flow, AP(Q'_A-G-SERP) and AP(Q'_A-G-RR) have a higher ratio of paragraphs and a lower ratio of headings and list items than AP(Q_A-G-SERP) and AP(Q_A-G-RR), respectively. Based on indicators related to text succinctness, AP(Q'_A-G-SERP) and AP(Q'_A-G-RR) have higher average paragraph length than AP(Q_A-G-SERP) and AP(Q_A-G-RR) respectively. Yet we observe an inverse relation between the sentence count and average sentence length in the compared profiles, i.e., sentence count is lower but average sentence length is higher in AP(Q'_A-G-SERP) than in AP(Q_A-G-SERP); whereas in AP(Q'_A-G-RR) the sentence count is significantly higher and average sentence length significantly lower than in AP(Q_A-G-RR).

Based on Readability, AP(Q'_A-G-SERP) has a higher FRES score and lower CLI score than AP(Q_A-G-SERP). In fact, FRES and CLI scores of AP(Q'_A-G-SERP) meet the readability criteria for autistic users, which implies that rephrasing the query resulted in Google SERP snippets that would actually be readable for autistic users. In the case of Google RR, AP($\mathtt{Q}'_\mathtt{A}$-G-RR) also has a higher FRES score and lower CLI score than AP($\mathtt{Q}_\mathtt{A}$-G-RR) but the higher scores don't meet the readability criteria.

In terms of Concreteness, the average concreteness in AP($\mathtt{Q}'_\mathtt{A}$-G-SERP) is higher than in AP($\mathtt{Q}_\mathtt{A}$-G-SERP) but the ratio of abstract words is also higher in AP($\mathtt{Q}'_\mathtt{A}$-G-SERP) than in AP($\mathtt{Q}_\mathtt{A}$-G-SERP). In case of Google RR on the other hand, AP($\mathtt{Q}_\mathtt{A}$-G-RR) is more concrete on average with a lower ratio of abstract words than AP($\mathtt{Q}'_\mathtt{A}$-G-RR); but the ratio of concrete words is higher in AP($\mathtt{Q}'_\mathtt{A}$-G-RR) than in AP($\mathtt{Q}_\mathtt{A}$-G-RR). From the perspective of Concreteness, it appears that the snippets in Google SERP are more concrete, on average, when the query specifies that the user is autistic, but the resources retrieved for such queries are less concrete on average compared to the resources retrieved when the queries do not state that the user is autistic.



(a) Average Paragraph Length. (b) Sentence Count.

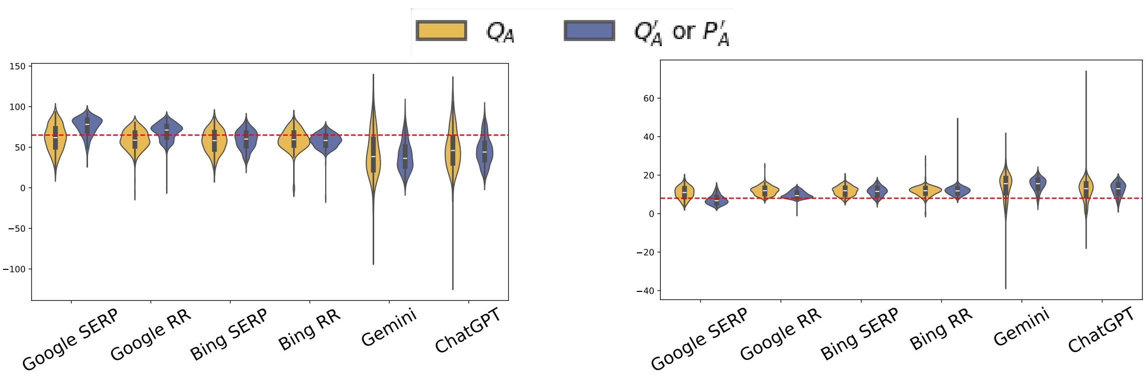
(c) Average Sentence Length. (d) Ratio of Headings.

(e) Ratio of List Items. (f) Ratio of Paragraphs.

Figure 17: Text structure scores for logs generated using $\mathtt{Q}_\mathtt{A}$, $\mathtt{Q}'_\mathtt{A}$, and $\mathtt{P}'_\mathtt{A}$.

Bing. Comparing AP($\mathtt{Q_A}$-B-SERP) and AP($\mathtt{Q'_A}$-B-SERP) reported in Table 7, we identify several prominent differences between the profiles, and all reported comparisons are significant unless stated otherwise. Examining the text structure of the profiles in terms of text succinctness, AP($\mathtt{Q'_A}$-B-SERP) has a lower average paragraph length and a higher sentence count than AP($\mathtt{Q_A}$-B-SERP); but the average sentence length in AP($\mathtt{Q'_A}$-B-SERP) is lower than in AP($\mathtt{Q_A}$-B-SERP). In terms of reading flow, AP($\mathtt{Q'_A}$-B-SERP) has a lower ratio of headings and a higher ratio of paragraphs but a significantly higher ratio of list items than AP($\mathtt{Q_A}$-B-SERP). From the Readability perspective, AP($\mathtt{Q'_A}$-B-SERP) has a higher FRES score and lower CLI score than AP($\mathtt{Q_A}$-B-SERP). However, the difference in the readability indicator scores is not statistically significant. Based on the Concreteness perspective, AP($\mathtt{Q'_A}$-B-SERP) has a lower average concreteness and a higher ratio of abstract words, but a lower ratio of concrete words than AP($\mathtt{Q_A}$-B-SERP); however, only the difference in average concreteness holds true in practice. From these findings, we infer that rephrasing the query to state that the user is autistic impacts the structure and average concreteness of Bing SERP snippets but not the readability of the snippets.

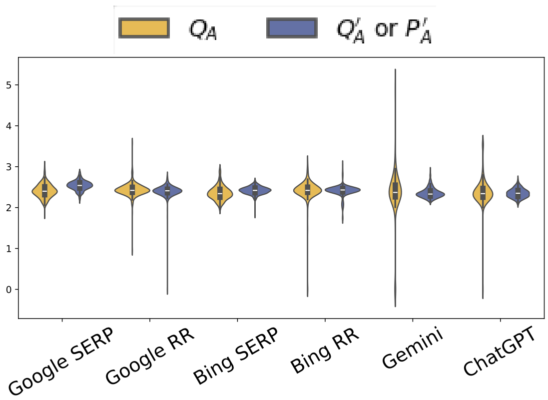
In case of Bing RR responses, juxtaposing AP($\mathtt{Q_A}$-B-RR) with AP($\mathtt{Q'_A}$-B-RR) as reported in Table 7 reveals very few significant differences. Based on Structure, AP($\mathtt{Q_A}$-B-RR) has significantly lower average paragraph and sentence lengths than AP($\mathtt{Q'_A}$-B-RR). From Readability and Concreteness perspectives, AP($\mathtt{Q_A}$-B-RR) has higher FRES and lower CLI score, as well as higher average concreteness and lower ratio of abstract words than AP($\mathtt{Q'_A}$-B-RR), but only the differences in the ratio of abstract words holds true in practice. Based on these outcomes, we infer that resources retrieved by Bing for queries in $\mathtt{Q'_A}$ are less succinct and less concrete on average compared to the resources retrieved for the queries in $\mathtt{Q_A}$, but the readability of the resources retrieved for either query variation is similar on average.



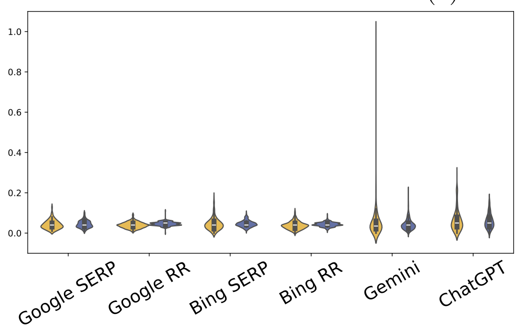
(a) FRES. (b) CLI.

Figure 18: Text readability scores for logs generated using $\mathtt{Q_A}$, $\mathtt{Q'_A}$, and $\mathtt{P'_A}$. The dotted red line illustrates the minimum FRES score in (a) and maximum CLI score in (b) an IAS response must be considered readable for an autistic person.

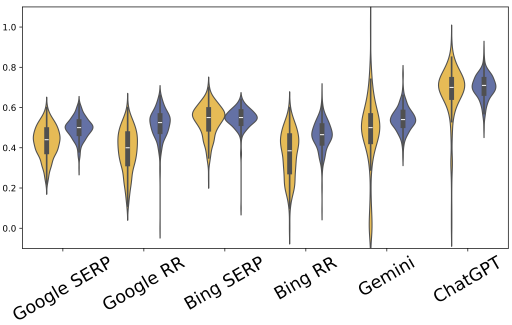
Gemini. Turning to the accessibility profiles of LLM agents reported in Table 7, we first compare AP($\mathtt{Q_A}$-GEM-ANS) with AP($\mathtt{P'_A}$-GEM-ANS), and find that AP($\mathtt{Q_A}$-GEM-ANS) is notably more closely aligned to autistic users' accessibility needs than AP($\mathtt{P'_A}$-GEM-ANS) but not all differences hold true in practice. Based on Structure, AP($\mathtt{Q_A}$-GEM-ANS) has significantly lower average paragraph and sentence lengths as well as significantly lower sentence count than AP($\mathtt{P}'_{\mathtt{A}}$-GEM-ANS). AP($\mathtt{P}'_{\mathtt{A}}$-GEM-ANS) has a higher ratio of headings and a lower ratio of paragraphs, but a lower ratio of list items than AP($\mathtt{Q}_{\mathtt{A}}$-GEM-ANS), although only the difference in the ratio of headings is statistically significant. In terms of Readability, AP($\mathtt{Q}_{\mathtt{A}}$-GEM-ANS) has a higher FRES score and significantly lower CLI score than AP($\mathtt{P}'_{\mathtt{A}}$-GEM-ANS) which implies that responses in $\mathtt{Q}_{\mathtt{A}}$-GEM-ANS are easier to read on average than the responses in $\mathtt{P}'_{\mathtt{A}}$-GEM-ANS. From the Concreteness perspective, AP($\mathtt{P}'_{\mathtt{A}}$-GEM-ANS) has a higher average concreteness than AP($\mathtt{Q}_{\mathtt{A}}$-GEM-ANS), but the difference is not significant. The ratio of abstract words, on the other hand, is significantly lower in AP($\mathtt{Q}_{\mathtt{A}}$-GEM-ANS) than in AP($\mathtt{P}'_{\mathtt{A}}$-GEM-ANS), from which we infer that rephrasing the query generally results in more abstract responses from Gemini.



(a) Average Concreteness.



(b) Ratio of Concrete Words.



(c) Ratio of Abstract Words.

Figure 19: Text concreteness scores for logs generated using $\mathtt{Q}_{\mathtt{A}}$, $\mathtt{Q}'_{\mathtt{A}}$ and $\mathtt{P}'_{\mathtt{A}}$.

ChatGPT. Examining the differences between AP($\mathtt{Q}_{\mathtt{A}}$-GPT-ANS) and AP($\mathtt{P}'_{\mathtt{A}}$-GPT-ANS) reported in Table 7 reveals that the scores pertaining to the reading flow of the responses remain unchanged. In terms of text succinctness, AP($\mathtt{Q}_{\mathtt{A}}$-GPT-ANS) has a significantly lower average paragraph and sentence lengths as well as significantly lower sentence count than AP($\mathtt{P}'_{\mathtt{A}}$-GPT-ANS). Even from the perspective of Readability, AP($\mathtt{Q}_{\mathtt{A}}$-GPT-ANS) has a higher FRES score and lower CLI score than AP($\mathtt{P}'_{\mathtt{A}}$-GPT-ANS), although the differences do not hold true in practice. In terms of Concreteness, AP($\mathtt{Q}_{\mathtt{A}}$-GPT-ANS) has a higher average concreteness and a lower ratio of abstract words than AP($\mathtt{P}'_\mathtt{A}$-GPT-ANS), but the difference in average concreteness is not statistically significant. Based on these findings, we infer that ChatGPT generates less succinct and less concrete answers on average for queries in $\mathtt{P}'_\mathtt{A}$, compared to the answers it generates for queries in $\mathtt{Q}_\mathtt{A}$.

5.3 Are the Responses of SE and LLM Agents More Accessible When Diagnosis Is Disclosed in a Query?

Based on our investigations juxtaposing responses produced by IASs to inquiries that do and do not explicitly state that the information needs are of an autistic user, we find that stating the user's condition in the inquiry does influence the type of resources retrieved by an SE as well as the answer generated by an LLM agent (§5.1). Comparing the accessibility profiles of each IAS generated based on responses for the two query variations (§5.2) reveals that from the perspective of text accessibility, SE responses are more prominently impacted compared to LLM agent responses. Particularly in the case of Google, the responses elicited using the rephrased queries are significantly easier to read with the SERP snippets even meeting the readability criteria for autistic users suggested in (Yaneva et al., 2015). However, both the snippets and the content in the retrieved resources are more verbose, and the structure of the responses, on average, is less suitable for an easy reading flow compared to the responses produced for the original queries. We observe a consistent trend across all the considered IASs that rephrasing the query to specify that the user is autistic results in responses that are less succinct than the responses produced for the original queries. The accessibility indicator scores captured in Figures 17, 18, and 19 reveal an interesting trend: the scores across all compared distributions are notably less varied for IAS responses to queries in $\mathtt{Q}'_\mathtt{A}$ (or $\mathtt{P}'_\mathtt{A}$) than the responses produced for queries in $\mathtt{Q}_\mathtt{A}$. Overall, we conclude that although informing the IAS of the user's diagnosis does not improve the general accessibility of IAS responses for autistic individuals, the rephrasing leads to responses that are more consistent in linguistic style in terms of Structure, Readability, and Concreteness.

6 Discussion

The outcomes of our study reveal important implications about the accessibility of mainstream IASs in catering to autistic individuals' information needs, which we discuss in this section; we also outline the limitations of our study and suggest potential future research directions.

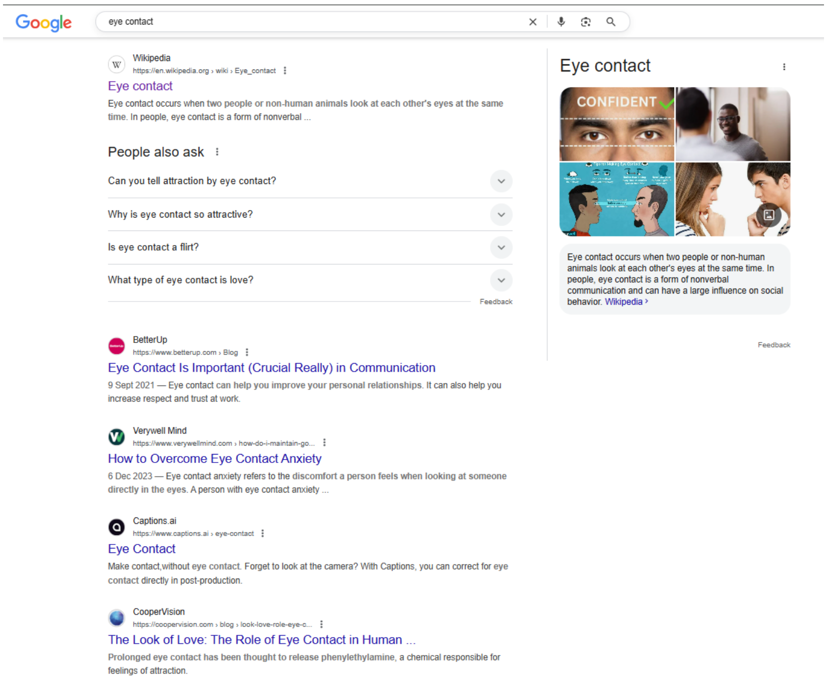
6.1 Text Accessibility of IAS Responses for Autistic Individuals

The results reported in Sections 4 and 5 highlight a potentially critical issue for autistic individuals who rely on popular IASs for information discovery. The accessibility profiles we generated of popular IASs to examine the text accessibility of IAS responses for autistic individuals' inquiries (RQ1) indicate that IAS responses to queries reflecting autistic individuals' common information needs are less accessible to the target audience than responses to inquiries from the general public. The issue is exacerbated when the autistic users' inquiry is related to their condition. Moreover, autistic individuals who are aware of their diagnosis may explicitly state their condition in their inquiry (Yechiam and Yom-Tov, 2021), most likely with the expectation that the IAS—now aware of the target audience—would respond to the inquiry in a manner suitable for autistic individuals. However, based on the results of our explorations on whether the text accessibility of IAS responses improves when the inquiry is rephrased to state the user's diagnosis (RQ2), we find that only Google meets these expectations, but only from the readability perspective. In all other cases, explicitly stating that the user is autistic leads to more verbose responses in general that are not structured to facilitate an easy reading flow compared to the responses produced when the queries do not include the user's diagnosis. Overall, our findings suggest that, in terms of text accessibility, mainstream IASs react distinctively to autistic individuals' information needs. Moreover, IAS responses to autistic users' queries are less aligned to these users' expectations, compared to system responses to the general population's inquiries. This could hinder autistic user from *utilising* the IAS response to satisfy their information need, suggesting a concerning trend of mainstream commercial IASs potentially impeding autistic individuals' access to online information.

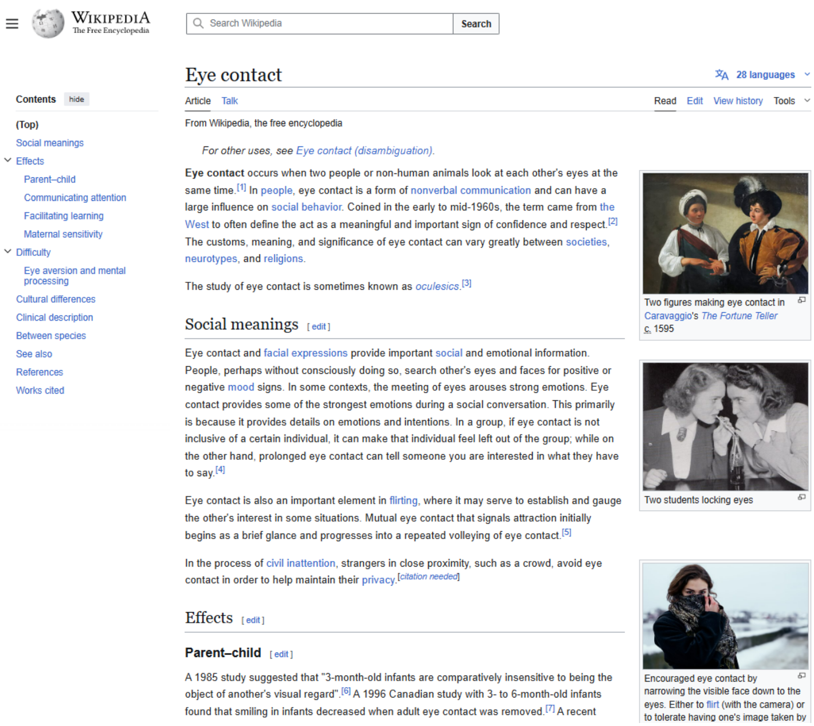
An unexpected outcome of our exploration was the impact of adding a diagnostic-disclosing phrase to the query on the linguistic *variability* of IAS responses. For an audience that struggles with adapting to unexpected changes as they prefer consistent and predictable interactions (WHO, 2023), the variation in the accessibility of IAS responses across query lengths and inquiry topics that we observed while examining the overall accessibility of IAS responses for autistic individuals (RQ1), can be anxiety-provoking. Specifying that the user is autistic in the inquiry to investigate its impact on the accessibility of IAS responses (RQ2) emerged as a solution to this issue. We found that the produced responses were notably more consistent in terms of accessibility across the rephrased inquiries compared to the accessibility of the responses for the original autistic individuals' inquiries.

6.2 Trade-Offs Across IAS

Our analysis highlights distinct text-based accessibility issues between SEs and LLM agents. LLM agents with their direct and succinct answers may appear to be an "autism-friendly" IAS, as they do not require the searcher to navigate through a list of resources, as is common in a traditional SE. Yet, from our analysis, we know that LLM agent responses, while succinct, are less concrete and harder to read than the web resources retrieved by an SE. This presents a trade-off, which we illustrate through an example. Consider an autistic user looking for information related to "eye contact". Entering this query in Google would yield a long list of snippets, as captured in Figure 20a. The user would then go through the list of short, easy-to-read snippets and choose a particular resource to explore further, like the Wikipedia article on eye contact illustrated in Figure 20b, for instance. The web resource would comprise a large amount of text compared to a snippet, but structured in a way that facilitates an easy reading flow. Moreover, based on the findings reported for RQ1, we could assume that the overall textual content would be much easier to read than the direct answer of an LLM agent. Using the same query on Gemini would yield a direct answer (as visualised in Figure 21). However, compared to the resources retrieved by Google, the answer would likely contain more complex sentences, so the user would have to spend extra effort to comprehend the text and extract the needed information. As mentioned previously, autistic individuals prefer consistency in their day-to-day interactions



(a) Top 5 ranked results in Google SERP.



(b) Excerpt from Wikipedia article on "eye contact", the top-ranked result in Google SERP from Figure 20a.

Figure 20: Google response for an autistic user query: "eye contact".

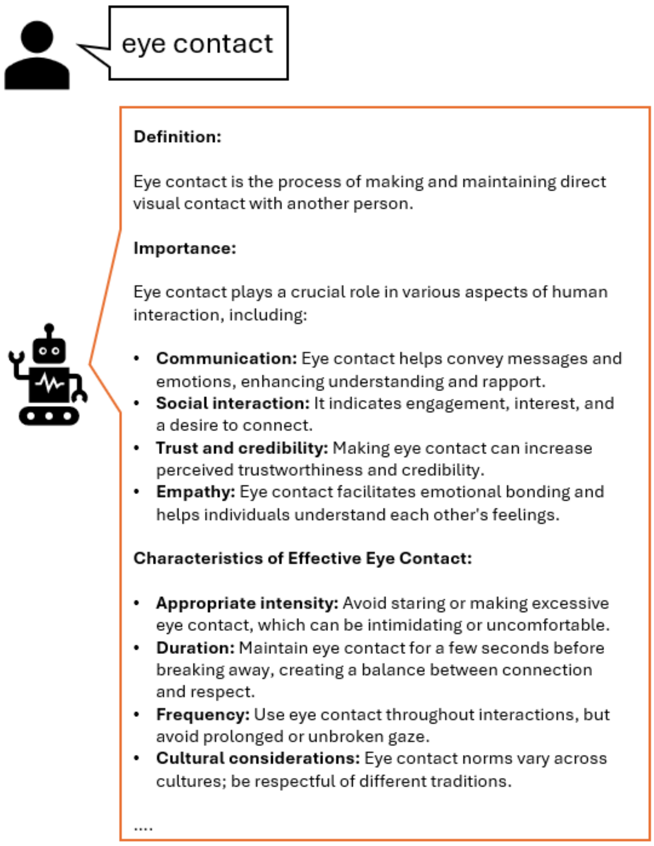


Figure 21: Excerpt from Gemini's generated response for an autistic user query: "eye contact".

(WHO, 2023). Therefore, the variation in the accessibility issues between SEs and LLM agents could hinder autistic users from using an IAS as they have grown accustomed to coping with the issues of the other system.

6.3 Designing for Accessibility

The individual profiles provide insights into the distinct strengths and limitations of each IAS in providing accessible responses to autistic individuals, which inspire strategies to enhance autistic users' overall online information-seeking process. Knowing that autistic individuals tend to scroll more on the SERP compared to non-autistic users when looking for information on an SE (Yechiam and Yom-Tov, 2021), it could be assumed that autistic users may not mind the extra effort of browsing through a SERP to find resources to answer their inquiry. Therefore, considering the easier readability and higher concreteness of SE responses, it may be useful to employ the better-structured response generation of LLM agents to further enhance the accessibility of SE responses for autistic users. For example, Gemini, in particular, generates responses structured to facilitate easier reading flow compared to the other IAS responses. Employing the LLM driving Gemini's response generation to restructure the verbose content in the resources retrieved by SE, as experimented in (Chhikara et al., 2024; Chang et al., 2023), could make it easier for autistic individuals to locate relevant components in an expansive web resource.

It is worth noting that SE responses, while easier to read than LLM responses, are yet to meet the readability criteria for autistic individuals. Stating the user is autistic in the inquiry leads to IAS responses that are even less aligned with autistic individuals' accessibility needs. Still, ChatGPT is capable of modifying its response to suit a specific need of a target audience, such as the reading level for a specified school grade (Murgia et al., 2023b). This could be utilised to simplify SE responses even for autistic individuals, by following the readability criteria suggested by Yaneva et al. (2015) and prompting ChatGPT to rephrase the SE response to suit the reading skill of an 8^{th} or 9^{th} grade student.

6.4 Limitations and Future Work

In the absence of publicly available data, we relied on synthetic queries to represent the information needs of autistic individuals. While our synthetic queries represent the plausible vocabulary and topics of interest common amongst autistic individuals, they may not represent other characteristics of how autistic users formulate their search queries, such as whether they formulate long or short queries, whether they formulate their queries using keywords only or opt for a more natural language approach, etc. Therefore, we advocate for the allocation of research efforts to collect representative query samples formulated by autistic individuals that can be used to examine whether their typical style of query formulation further influences the accessibility of IAS responses. Additionally, noting the difference in the readability of IAS responses for autistic users' inquiries directly related to ASD compared to more general inquiries (§4.3), future extensions of the study could also investigate the impact of autistic individuals' inquiry topic on the accessibility of the elicited IAS responses further based on more fine-grained categorisation of inquiry topics to reveal other trends. Regarding the experiments we conducted to answer RQ2, we explored the impact of including the diagnosis-reporting with a singular rephrasing strategy, but in the future, it may be beneficial to explore the potential of other known query rephrasing and prompt engineering strategies (e.g., Chen et al., 2021; Schmidt et al., 2023; Meskó, 2023; Xu et al., 2011) to yield IAS responses that are consistent in linguistic style but also accessible to autistic individuals.

Insights emerging from the accessibility profiles probed in our study showcase several potential accessibility limitations of popular IAS, even though we limited our analysis to the *textual* content of IAS responses to control the scope of our exploration. However, modern-day IAS are not limited to text modality. For instance, the online resources retrieved by an SE contain embedded videos, images, and sometimes even audio to keep the user engaged with the content. The latest versions of LLM agents are also capable of understanding and generating content with visual and audio elements (Islam and Moushi, 2024; Gemini Team Google et al., 2024). Multimodal content can be particularly beneficial for autistic individuals' access to online information—as long as each modality presents the same content—since it gives autistic individuals the freedom to interact with the same content in whichever modality they find the easiest to comprehend (WebAIM, 2025; Friedman and Bryen, 2007). Further, voice-controlled virtual assistants like Siri and Alexa, and video platforms like YouTube, enable new search paradigms as interactions with such tools and platforms often lack a textual component. The diversification of online search paradigms and modalities raises other accessibility concerns for autistic individuals. For example, autistic individuals are very sensitive to loud noises (Williams et al., 2021; Landon et al., 2016), therefore sudden disturbing noises like fireworks and sirens in embedded videos can negatively impact the accessibility of the online content for autistic individuals (Britto and Pizzolato, 2016; Sitdhisanguan et al., 2012). Similarly, autistic individuals benefit greatly from visual stimuli during information search, but if the visual component is too abstract, then it could instead hinder a autistic person from understanding the content they are exposed to (Yaneva et al., 2015; Rezae et al., 2020). Considering autistic people's accessibility needs that go beyond reading a piece of text, it is important to examine the accessibility of popular IASs from a multimodal perspective. For this, COGA's proposed guidelines based on autistic individuals' challenges with comprehending content in non-textual modalities could be used in the future to extend the design of our experimental setup for generating accessibility profiles that capture the suitability of IAS responses for autistic individuals from a diverse perspective of web accessibility.

We examined the accessibility of the first response of different IASs to independent autistic individuals inquiries. As noted earlier, AI overviews were not consistently available at the time of the study and were therefore excluded from our analysis. However, as mainstream SEs increasingly integrate AI-generated summaries (or direct responses) at the top of SERPs Reid (2026, 2024), users are likely to encounter and engage with these summaries before looking at the list of retrieved results. With that in mind, it is important for future extensions of this work to also consider AI-generated overviews when assessing the accessibility of SE responses for autistic individuals. Furthermore, according to the information search process (**ISP**) model proposed by Kuhlthau (1991), a user goes through several stages when searching for information online. Starting from *initiation*, which is when the user realises they have a knowledge gap. This is followed by the *selection* and *exploration* stages, during which the user conducts a preliminary search to collect and explore resources that can help understand the general topic related to their knowledge gap. Our study captures the accessibility of mainstream IASs in supporting this preliminary search. But after the exploration stage, once the user has acquired an understanding of the general topic, they enter the *formulation* stage, wherein they may further refine their initial query to collect resources that are more aligned to their specific inquiry, followed by *collection* and *presentation* stages only after which a user's ISP is said to be complete. In the future, our study could be extended to probe user-IAS interactions at different stages of the user's ISP to get a more holistic overview of the accessibility of an IAS for autistic individuals.

The accessibility guidelines we use to inform the design of our experimental framework present a generalised overview of the challenges faced by autistic individuals, but one of the most characteristic traits of ASD is the variability in the abilities of autistic people (WHO, 2023). In the future, a complementary study could further validate the reliability of the accessibility profiles generated for the IAS by incorporating the perception of actual autistic users conducting self-directed online search tasks. Furthermore, considering the low diagnosis rate for ASD in several countries, especially amongst women (WHO, 2023), existing guidelines may not account for the distinct needs and expectations of these under-represented communities within a population that already is a minority in general. This underscores the need for interdisciplinary research to broaden the understanding of ASD and its associated challenges to enable the creation of guidelines that represent the needs of *all* autistic individuals and not just an informed, stereotyped version of them.

7 Conclusions

In this work, we conducted an empirical exploration to gauge what kind of content autistic users are potentially exposed to when seeking information using mainstream IASs through the lens of Web Content Accessibility. In the absence of publicly available datasets, we constructed synthetic query collections using (i) key phrases from Reddit posts by self-reported autistic users—capturing topics, vocabulary, and phrasing common amongst autistic individuals—as proxy queries reflecting their information needs, along with (ii) a random sample from the Yahoo! Search Query Tiny dataset, to represent the information needs of the general population, respectively. We used these query collections to elicit responses from Google, Bing, ChatGPT, and Gemini. From the resulting IAS responses, we constructed an *accessibility profile* for each system. Rather than classifying a system as “good” or “bad” for autistic users, we use these profiles to probe IAS responses along three dimensions of text accessibility: structure, readability, and concreteness.

A nuanced exploration of the profiles (RQ1) reveals that IAS responses are less aligned to autistic individuals’ accessibility requirements when catering to autistic users information needs, compared to when these systems respond to inquiries from the general population. The issue is exacerbated when an autistic user’s inquiry is related to ASD. Rephrasing the autistic individuals’ inquiries to explicitly include the user’s condition in the inquiry to the IAS (RQ2)—an online search behaviour typical for diagnosed autistic individuals (Yechiam and Yom-Tov, 2021)—results in IAS responses that are even less accessible for this population. These findings suggest that autistic individuals relying on popular IASs for information discovery may struggle with finding the information they need to complete their information-seeking process. This highlights a critical limitation of popular IASs in catering to the information needs of autistic individuals and provides empirical evidence that could explain why autistic individuals so often feel frustrated while searching for information online (Gibson and Hanson-Baldauf, 2019).

Access to information is integral to a person’s freedom of expression (UNESCO). Considering autistic individuals’ tendency to give up on their search tasks when they are unable to find resources that fit their needs (Gibson and Hanson-Baldauf, 2019), we argue that the poor text accessibility of IAS responses could actively be *barring* autistic individuals from accessing crucial information online, which is a direct infringement of their fundamental rights as human beings. The European Accessibility Act also mandates all products and services—public or private—to comply with the accessibility requirements of people with disabilities (European Union, 2019). From both ethical and legal points of view, the outcomes of our study reveal a potentially concerning scenario of the accessibility of online information for autistic individuals, underscoring the need for researchers and developers alike to address the known issues as well as allocate efforts to recognise other limitations not only for autistic individuals but also other under-served communities, for which our experimental setup to generate accessibility profiles may serve as a useful reference.

Acknowledgments and Disclosure of Funding

The authors thank the reviewers and editor for their constructive feedback, which helped refine the final draft of the manuscript. This work was carried out without external funding or financial support from any funding agency, commercial entity, or other external organization.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Marwah Alaofi, Luke Gallagher, Dana Mckay, Lauren L. Saling, Mark Sanderson, Falk Scholer, Damiano Spina, and Ryen W. White. Where do queries come from? In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22, page 2850–2862, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450387323. doi: 10.1145/3477495.3531711.

Garrett Allen, Ashlee Milton, Katherine Landau Wright, Jerry Alan Fails, Casey Kennington, and Maria Soledad Pera. Supercalifragilisticexpialidocious: Why using the "right" readability formula in children's web search matters. In *European Conference on Information Retrieval*, pages 3–18. Springer, 2022.

Mona Alzahrani, Alexandra L Uitdenbogerd, and Maria Spichkova. Impact of animated objects on autistic and non-autistic users. In *Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software engineering in society*, pages 102–112, 2022.

American Psychiatric Association APA. *Diagnostic and statistical manual of mental disorders.* American Psychiatric Association Publishing, 2022.

Jesse G Bahrke. Mining for diagnostic criteria in the social behavior of people with asd: Using machine learning to explore aspects of reddit use. Master's thesis, Rosalind Franklin University of Medicine and Science, 2024.

Penny Benford and P Standen. The internet: a comfortable communication medium for people with asperger syndrome (as) and high functioning autism (hfa)? *Journal of Assistive Technologies*, 3(2):44–53, 2009.

Tiago Bianchi. Global search engine desktop market share 2024, May 2024. URL `https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/`.

Roi Blanco, Harry Halpin, Daniel M Herzig, Peter Mika, Jeffrey Pound, Henry S Thompson, and T Tran Duc. Entity search evaluation over structured web data. In *Proceedings of the 1st international workshop on entity-oriented search workshop (SIGIR 2011), ACM, New York*, volume 14, pages 2181–2187, 2011.

Alexis Bosseler and Dominic W Massaro. Development and evaluation of a computer-animated tutor for vocabulary and language learning in children with autism. *Journal of autism and developmental disorders*, 33:653–672, 2003.

Kristen Bottema-Beutel, Steven K Kapp, Jessica Nina Lester, Noah J Sasson, and Brittany N Hand. Avoiding ableist language: Suggestions for autism researchers. *Autism in adulthood*, 2021.

Talita Britto and Ednaldo Pizzolato. Towards web accessibility guidelines of interaction and interface design for people with autism spectrum disorder. In *ACHI 2016: the ninth international conference on advances in computer-human interactions*, pages 1–7, 2016.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods*, 46:904–911, 2014.

Yancheng Cao, Yangyang He, Yonglin Chen, Menghan Chen, Shanhe You, Yulin Qiu, Min Liu, Chuan Luo, Chen Zheng, Xin Tong, et al. Designing llm-simulated immersive spaces to enhance autistic children's social affordances understanding in traffic settings. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*, pages 519–537, 2025.

Emily Cary, Aparajita Rao, Erin Stephanie Misato Matsuba, and Natalie Russo. Barriers to an autistic identity: How rrbs may contribute to the underdiagnosis of females. *Research in Autism Spectrum Disorders*, 109:102275, 2023.

Centers for Disease Control and Prevention CDC. Signs and symptoms of autism spectrum disorder, 01 2024a. URL `https://www.cdc.gov/autism/signs-symptoms/index.html`.

Centers for Disease Control and Prevention CDC. About autism spectrum disorder, May 2024b. URL `https://www.cdc.gov/autism/about/index.html`.

Federica Cena, Noemi Mauro, Liliana Ardissono, Claudio Mattutino, Amon Rapp, Stefano Cocomazzi, Stefania Brighenti, and Roberto Keller. Personalized tourist guide for people with autism. In *Adjunct publication of the 28th ACM conference on user modeling, adaptation and personalization*, pages 347–351, 2020.

Yapei Chang, Kyle Lo, Tanya Goyal, and Mohit Iyyer. Booookscore: A systematic exploration of book-length summarization in the era of llms. *arXiv preprint arXiv:2310.00785*, 2023.

Jia Chen, Jiaxin Mao, Yiqun Liu, Fan Zhang, Min Zhang, and Shaoping Ma. Towards a better understanding of query reformulation behavior in web search. In *Proceedings of the web conference 2021*, pages 743–755, 2021.

Garima Chhikara, Anurag Sharma, V Gurucharan, Kripabandhu Ghosh, and Abhijnan Chakraborty. Lamsum: A novel framework for extractive summarization of user generated content using llms. *arXiv preprint arXiv:2406.15809*, 2024.

Dasom Choi, Sunok Lee, Sung-In Kim, Kyungah Lee, Hee Jeong Yoo, Sangsu Lee, and Hwajung Hong. Unlock life with a chat(gpt): Integrating conversational ai with large language models into everyday lives of autistic individuals. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. doi: 10.1145/3613904.3641989. URL https://doi.org/10.1145/3613904.3641989.

Lei Chu, Hongyan Wu, and Yi Pan. Chatasd: A dialogue framework for llms enhanced by autism knowledge graph retrieval. In *Proceedings of the 15th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 1–8, 2024.

Meri Coleman and Ta Lin Liau. A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60(2):283, 1975.

Cherie Courseault Trumbach and Dinah Payne. Identifying synonymous concepts in preparation for technology mining. *Journal of Information Science*, 33(6):660–677, 2007.

Antonina Dattolo, Flaminia L. Luccio, and Elisa Pirone. Webpage accessibility and usability for autistic users: a case study on a tourism website. In *International Conference on Advances in Computer-Human Interaction*, 2016a. URL https://api.semanticscholar.org/CorpusID:55048059.

Antonina Dattolo, Flaminia L Luccio, Elisa Pirone, et al. Web accessibility recommendations for the design of tourism websites for people with autism spectrum disorders. *International Journal on Advances in Life Sciences*, 8(3 & 4):297–308, 2016b.

Munmun De Choudhury, Scott Counts, Eric J Horvitz, and Aaron Hoff. Characterizing and predicting postpartum depression from shared facebook data. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 626–638, 2014.

Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 2098–2110, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450333627. doi: 10.1145/2858036.2858207.

Emani Dotch. Audiobuddy: An assistive technology for people with noise sensitivity and their care networks. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '24, page 267–271, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400710582. doi: 10.1145/3675094.3678365.

Emani Dotch, Jazette Johnson, Rebecca W. Black, and Gillian R Hayes. Understanding noise sensitivity through interactions in two online autism forums. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400702204. doi: 10.1145/3597638.3608413.

B Driver and V Chester. The recognition and diagnosis of autism in women and girls: A literature review. *Advances in Autism*, 7(3):194–207, 2021.

Sukru Eraslan, Victoria Yaneva, Yeliz Yesilada, and Simon Harper. Do web users with autism experience barriers when searching for information within web pages? In *Proceedings of the 14th International Web for All Conference*, W4A '17, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450349000. doi: 10.1145/3058555.3058566.

Sukru Eraslan, Yeliz Yesilada, Victoria Yaneva, and Le An Ha. "keep it simple!": An eye-tracking study for exploring complexity and distinguishability of web pages for people with autism. *Universal Access in the Information Society*, 20(1):69–84, 2020. doi: 10.1007/s10209-020-00708-9.

Curts Eric. Chrome extensions for struggling students and special needs, Nov 2019. URL `https://www.controlaltachieve.com/2016/10/special-needs-extensions.html`.

European Union. Directive (eu) 2019/882 of the european parliament and of the council of 17 april 2019 on the accessibility requirements for products and services (text with eea relevance), 2019. URL `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L0882`.

Rudolf Flesch. How to write plain english, 07 2016. URL `https://web.archive.org/web/20160712094308/http://www.mang.canterbury.ac.nz/writing_guide/writing/flesch.shtml`.

Seraphina Fong, Alessandro Carollo, Giacomo Vivanti, Daniel S Messinger, Dagmara Dimitriou, and Gianluca Esposito. Autism spectrum disorders discourse on social media platforms: A topic modeling study of reddit posts. *Autism research*, 2025.

Mark G Friedman and Diane Nelson Bryen. Web accessibility design recommendations for people with cognitive disabilities. *Technology and disability*, 19(4):205–212, 2007.

Mark G. Friedman and Diane Nelson Bryen. Web accessibility design recommendations for people with cognitive disabilities. *Technology and Disability*, 19(4):205–212, 2008. doi: 10.3233/tad-2007-19406.

Uta Frith and Maggie Snowling. Reading for meaning and reading for sound in autistic and dyslexic children. *British journal of developmental psychology*, 1(4):329–342, 1983.

Gemini Team Google, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.

Gemini Team Google, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, et al. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*, 2024.

Amelia N Gibson and Dana Hanson-Baldauf. I want it the way i need it": Modality, readability, and format control for autistic information seekers online. *International Journal on Innovations in Online Education*, 2(4):1–12, 2019.

Kristen Gillespie-Lynch, Steven K Kapp, Christina Shane-Simpson, David Shane Smith, and Ted Hutman. Intersections between the autism spectrum and the internet: Perceived benefits and preferred functions of computer-mediated communication. *Intellectual and developmental Disabilities*, 52(6):456–469, 2014.

Maarten Grootendorst. Keybert: Minimal keyword extraction with bert., 2020. URL `https://doi.org/10.5281/zenodo.4461265`.

Francesca Happé and Uta Frith. The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of autism and developmental disorders*, 36:5–25, 2006.

Francesca GE Happé. Central coherence and theory of mind in autism: Reading homographs in context. *British journal of developmental psychology*, 15(1):1–12, 1997.

Rukhshan Haroon and Fahad Dogar. Twips: A large language model powered texting application to simplify conversational nuances for autistic users. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–18, 2024.

Arno Hartholt, Sharon Mozgai, Ed Fast, Matt Liewer, Adam Reilly, Wendy Whitcup, and Albert" Skip" Rizzo. Virtual humans in augmented reality: A first step towards real-world embedded virtual roleplayers. In *Proceedings of the 7th international conference on human-agent interaction*, pages 205–207, 2019.

Amanda J Haskins, Thomas L Botch, Brenda D Garcia, Jennifer McLaren, and Caroline E Robertson. Gaze patterns modeled with a llm can be used to classify autistic vs. non-autistic viewers. *Journal of Vision*, 24(10):1190–1190, 2024.

Patricia Howlin and Philippa Moss. Adults with autism spectrum disorders. *The Canadian Journal of Psychiatry*, 57(5):275–283, 2012.

Chuanbo Hu, Wenqi Li, Mindi Ruan, Xiangxu Yu, Lynn K Paul, Shuo Wang, and Xin Li. Exploiting chatgpt for diagnosing autism-associated language disorders and identifying distinct features. *Research Square*, pages rs–3, 2024.

Raisa Islam and Owana Marzia Moushi. Gpt-4o: The cutting-edge advancement in multimodal llm. *Authorea Preprints*, 2024.

JiWoong Jang, Sanika Moharana, Patrick Carrington, and Andrew Begel. "it's the only thing i can trust": Envisioning large language model use by autistic workers for communication assistance. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2024.

Yi Jiang, Qingyang Shen, Shuzhong Lai, Shunyu Qi, Qian Zheng, Lin Yao, Yueming Wang, and Gang Pan. Copiloting diagnosis of autism in real clinical scenarios via llms. *arXiv preprint arXiv:2410.05684*, 2024.

MV Judy, U Krishnakumar, and AG Hari Narayanan. Constructing a personalized e-learning system for students with autism based on soft semantic web technologies. In *2012 IEEE International Conference on Technology Enhanced Education (ICTEE)*, pages 1–5. IEEE, 2012.

Tamar Kalandadze, Courtenay Norbury, Terje Nærland, and Kari-Anne B Næss. Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism*, 22(2):99–117, 2018.

Michelle R Kandalaft, Nyaz Didehbani, Daniel C Krawczyk, Tandra T Allen, and Sandra B Chapman. Virtual reality social cognition training for young adults with high-functioning autism. *Journal of autism and developmental disorders*, 43:34–44, 2013.

Gabriella Kazai, Paul Thomas, and Nick Craswell. The emotion profile of web search. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, pages 1097–1100, 2019.

Lorcan Kenny, Caroline Hattersley, Bonnie Molins, Carole Buckley, Carol Povey, and Elizabeth Pellicano. Which terms should be used to describe autism? perspectives from the uk autism community. *Autism*, 20(4):442–462, 2016.

Jina Kim, Jieon Lee, Eunil Park, and Jinyoung Han. A deep learning model for detecting mental illness from user content on social media. *Scientific reports*, 10(1):11846, 2020.

J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. 1975.

N Koteyko. Understanding autistic adults' use of social media. *Proceedings of the ACM on Human-Computer Interaction*, 2023.

Carol Collier Kuhlthau. Inside the search process: Information seeking from the user's perspective. *J. Am. Soc. Inf. Sci.*, 42:361–371, 1991. URL `https://api.semanticscholar.org/CorpusID:14416802`.

Meng-Chuan Lai and Simon Baron-Cohen. Identifying the lost generation of adults with autism spectrum conditions. *The Lancet Psychiatry*, 2(11):1013–1027, 2015.

Jason Landon, Daniel Shepherd, and Veema Lodhia. A qualitative study of noise sensitivity in adults with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 32: 43–52, 2016.

Monica Landoni, Maria Soledad Pera, Emiliana Murgia, and Theo Huibers. Inside out: Exploring the emotional side of search engines in the classroom. In *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization*, pages 136–144, 2020.

Michael Liedtke and Matt O'Brien. Google launches gemini, upping the stakes in the global ai race, Dec 2023. URL `https://apnews.com/article/google-gemini-artificial-intelligence-launch-95d05d02051e75e20b574614ae720b8b`.

Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81, 2004.

Chu-Sui Lin, Shu-Hui Chang, Wen-Ying Liou, and Yu-Show Tsai. The development of a multimedia online language assessment tool for young children with autism. *Research in developmental disabilities*, 34(10):3553–3565, 2013.

Ling-Yi Lin and Pai-Chuan Huang. Quality of life and its related factors for adults with autism spectrum disorder. *Disability and rehabilitation*, 41(8):896–903, 2019.

Rachel Tsz-Wai Lo, Ben He, and Iadh Ounis. Automatically building a stopword list for an information retrieval system. In *Journal on Digital Information Management: Special Issue on the 5th Dutch-Belgian Information Retrieval Workshop (DIR)*, volume 5, pages 17–24, 2005.

Jill Locke, Eric H Ishijima, Connie Kasari, and Nancy London. Loneliness, friendship quality and the social networks of adolescents with high-functioning autism in an inclusive school setting. *Journal of research in special educational needs*, 10(2):74–81, 2010.

O Ivar Lovaas and Laura Schreibman. Stimulus overselectiv1ty of autistic children in a two stimulus situation. *Behaviour research and therapy*, 9(4):305–310, 1971.

David Mason, Helen McConachie, Deborah Garland, Alex Petrou, Jacqui Rodgers, and Jeremy R Parr. Predictors of quality of life for autistic adults. *Autism Research*, 11(8): 1138–1147, 2018.

Alexey N. Medvedev, Renaud Lambiotte, and Jean-Charles Delvenne. The anatomy of reddit: An overview of academic research. *Springer Proceedings in Complexity*, page 183–204, May 2019. doi: 10.1007/978-3-030-14683-2_9.

Bertalan Meskó. Prompt engineering as an important emerging skill for medical professionals: tutorial. *Journal of medical Internet research*, 25:e50638, 2023.

Dan Milmo. Google says new ai model gemini outperforms chatgpt in most tests, Dec 2023. URL `https://www.theguardian.com/technology/2023/dec/06/google-new-ai-model-gemini-bard-upgrade`.

Ashlee Milton and Maria Soledad Pera. Into the unknown: Exploration of search engines' responses to users with depression and anxiety. *ACM Trans. Web*, 17(4), 07 2023. ISSN 1559-1131. doi: 10.1145/3580283. URL `https://doi.org/10.1145/3580283`.

Emiliana Murgia, Zahra Abbasiantaeb, Mohammad Aliannejadi, Theo Huibers, Monica Landoni, and Maria Soledad Pera. Chatgpt in the classroom: A preliminary exploration on the feasibility of adapting chatgpt to support children's information discovery. In *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '23 Adjunct, page 22–27, New York, NY, USA, 2023a. Association for Computing Machinery. ISBN 9781450398916. doi: 10.1145/3563359.3597399. URL `https://doi.org/10.1145/3563359.3597399`.

Emiliana Murgia, Maria Soledad Pera, Monica Landoni, and Theo Huibers. Children on chatgpt readability in an educational context: Myth or opportunity? UMAP '23 Adjunct, page 311–316, New York, NY, USA, 2023b. Association for Computing Machinery. ISBN 9781450398916. doi: 10.1145/3563359.3596996. URL `https://doi.org/10.1145/3563359.3596996`.

Johan Nyrenius, Jonas Eberhard, Mohammad Ghaziuddin, Christopher Gillberg, and Eva Billstedt. The 'lost generation'in adult psychiatry: psychiatric, neurodevelopmental and sociodemographic characteristics of psychiatric patients with autism unrecognised in childhood. *BJPsych open*, 9(3):e89, 2023.

Irene M O'Connor and Perry D Klein. Exploration of strategies for facilitating the reading comprehension of high-functioning students with autism spectrum disorders. *Journal of autism and developmental disorders*, 34:115–127, 2004.

OpenAI. Introducing chatgpt, Nov 2022. URL `https://openai.com/index/chatgpt/`.

Elizabeth O'Nions, Irene Petersen, Joshua E.J. Buckman, Rebecca Charlton, Claudia Cooper, Anne Corbett, Francesca Happé, Jill Manthorpe, Marcus Richards, Rob Saunders, and et al. Autism in england: Assessing underdiagnosis in a population-based cohort study of prospectively collected primary care data. *The Lancet Regional Health - Europe*, 29:100626, 06 2023. doi: 10.1016/j.lanepe.2023.100626.

Allan Paivio. Dual coding theory, word abstractness, and emotion: a critical review of kousta et al.(2011). 2013.

Anna Y Park, Andy Jin, Jeremy Zhengqi Huang, Jesse Carr, and Dhruv Jain. Masksound: Exploring sound masking approaches to support people with autism in managing noise sensitivity. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–12, 2024.

Michael Paul and Mark Dredze. You are what you tweet: Analyzing twitter for public health. In *Proceedings of the international AAAI conference on web and social media*, volume 5, pages 265–272, 2011.

Nikolay Pavlov. User interface for people with autism spectrum disorders. *Journal of Software Engineering and Applications*, 2014, 2014.

Maria Soledad Pera, Emiliana Murgia, Monica Landoni, Theo Huibers, and Mohammad Aliannejadi. Where a little change makes a big difference: a preliminary exploration of children's queries. In *European Conference on Information Retrieval*, pages 522–533. Springer, 2023.

Bertram O Ploog. Stimulus overselectivity four decades later: A review of the literature and its implications for current research in autism spectrum disorder. *Journal of autism and developmental disorders*, 40:1332–1349, 2010.

Cynthia Putnam and Lorna Chong. Software and technologies designed for people with autism: what do users want? In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '08, page 3–10, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781595939760. doi: 10.1145/1414471.1414475. URL `https://doi.org/10.1145/1414471.1414475`.

Amon Rapp, Federica Cena, Guido Boella, Alessio Antonini, Alessia Calafiore, Stefania Buccoliero, Maurizio Tirassa, Roberto Keller, Romina Castaldo, and Stefania Brighenti. Interactive urban maps for people with autism spectrum disorder. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1987–1992, 2017.

Vijayalaxmi N Rathod, RH Goudar, Anjanabhargavi Kulkarni, Dhananjaya GM, and Geetabai S Hukkeri. A survey on e-learning recommendation systems for autistic people. *IEEE Access*, 12:11723–11732, 2024.

Dora M Raymaker, Steven K Kapp, Katherine E McDonald, Michael Weiner, Elesia Ashkenazy, and Christina Nicolaidis. Development of the aaspire web accessibility guidelines for autistic web users. *Autism in Adulthood*, 1(2):146–157, 2019.

Elizabeth Reid. Generative ai in search: Let google do the searching for you, May 2024. URL https://blog.google/products-and-platforms/products/search/generative-ai-google-search-may-2024/.

Elizabeth Reid. A new era for ai search, May 2026. URL https://blog.google/products-and-platforms/products/search/search-io-2026/#powerful-ai.

Mortaza Rezae, Nigel Chen, David McMeekin, Tele Tan, Aneesh Krishna, and Hoe Lee. The evaluation of a mobile user interface for people on the autism spectrum: An eye movement study. *International Journal of Human-Computer Studies*, 142:102462, 2020. doi: 10.1016/j.ijhcs.2020.102462.

Esteban A Ríssola, Mohammad Aliannejadi, and Fabio Crestani. Mental disorders on online social media through the lens of language and behaviour: Analysis and visualisation. *Information processing & management*, 59(3):102890, 2022.

Sergio Rubio-Martín, María Teresa García-Ordás, Martín Bayón-Gutiérrez, Natalia Prieto-Fernández, and José Alberto Benítez-Andrades. Early detection of autism spectrum disorder through ai-powered analysis of social media texts. In *2023 IEEE 36th International symposium on computer-based medical systems (CBMS)*, pages 235–240. IEEE, 2023.

Ginny Russell, Sal Stapley, Tamsin Newlove-Delgado, Andrew Salmon, Rhianna White, Fiona Warren, Anita Pearson, and Tamsin Ford. Time trends in autism diagnosis over 20 years: a uk population-based cohort study. *Journal of Child Psychology and Psychiatry*, 63(6):674–682, 2022.

Siba Sankar Sahu, Debrup Dutta, Sukomal Pal, and Imran Rasheed. Effect of stopwords and stemming techniques in urdu ir. *SN Computer Science*, 4(5):547, 2023.

Alessia Sarica, Andrea Quattrone, and Aldo Quattrone. Introducing the rank-biased overlap as similarity measure for feature importance in explainable machine learning: a case study on parkinson's disease. In *International Conference on Brain Informatics*, pages 129–139. Springer, 2022.

Douglas C Schmidt, Jesse Spencer-Smith, Quchen Fu, and Jules White. Cataloging prompt patterns to enhance the discipline of prompt engineering. 2023. URL https://www.dre.vanderbilt.edu/~schmidt/PDF/ADA_Europe_Position_Paper.pdf.

Lisa Seeman and Michael Cooper. Cognitive accessibility user research, 12 2024. URL https://www.w3.org/TR/coga-user-research/#autism.

Dushyant Sharma, Rishabh Shukla, Anil Kumar Giri, and Sumit Kumar. A brief review on search engine optimization. *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pages 687–692, 2019. URL https://api.semanticscholar.org/CorpusID:199059492.

Paul T Shattuck, Gael I Orsmond, Mary Wagner, and Benjamin P Cooper. Participation in social activities among adolescents with an autism spectrum disorder. *PloS one*, 6(11): e27176, 2011.

Safa Shubbar. *Advancing Autism Spectrum Disorder Diagnosis: A Phenotype-Genotype Co-Analysis and Retrieval-Augmented LLM Framework for Clinical Decision Support*. PhD thesis, Kent State University, 2025.

Karanya Sitdhisanguan, Nopporn Chotikakamthorn, Ajchara Dechaboon, and Patcharaporn Out. Using tangible user interfaces in computer-based training systems for low-functioning autistic children. *Personal and Ubiquitous Computing*, 16:143–155, 2012.

Amanda Hoover Samantha Spengler. For some autistic people, chatgpt is a lifeline, May 2023. URL `https://www.wired.com/story/for-some-autistic-people-chatgpt-is-a-lifeline/`.

Sandra Thom-Jones, Imogen Melgaard, and Chloe S Gordon. Autistic women's experience of motherhood: A qualitative analysis of reddit. *Journal of Autism and Developmental Disorders*, pages 1–11, 2024.

UNESCO. Right to information. URL `https://www.unesco.org/en/right-information`.

United Nations. Universal declaration of human rights, Dec 1948. URL `https://www.un.org/en/about-us/universal-declaration-of-human-rights`.

Cornelis J Van Rijsbergen. Information retrieval, 2nd edn. newton, ma, 1979.

Sharon Varghese, Emily Carlson, and Vinoo Alluri. Listening to the spectrum: Exploring music preferences and associations in autism spectrum disorder communities on reddit, 2024.

Mila Vulchanova and Valentin Vulchanov. Rethinking figurative language in autism: What evidence can we use for interventions? *Frontiers in Communication*, 7:910850, 2022.

World Wide Web Consortium W3C. Web accessibility initiative, 2019. URL `https://www.w3.org/WAI/`.

World Wide Web Consortium W3C. Cognitive and learning disabilities accessibility task force, 2021. URL `https://www.w3.org/WAI/about/groups/task-forces/coga/`.

Tao Wang, Monica Garfield, Pamela Wisniewski, and Xinru Page. Benefits and challenges for social media users on the autism spectrum. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing*, pages 419–424, 2020.

WebAIM. Evaluating cognitive web accessibility with wave, 2025. URL `https://wave.webaim.org/cognitive`.

Kevin White and Shadi Abou-Zahra. Ian, data entry clerk with autism, 06 2024. URL `https://www.w3.org/WAI/people-use-web/user-stories/story-two/`.

Ryen W. White. Advancing the search frontier with ai agents. *Commun. ACM*, 67(9): 54–65, August 2024. ISSN 0001-0782. doi: 10.1145/3655615.

World Health Organization WHO. Autism, 03 2023. URL `https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders`.

Diane L Williams, Nancy J Minshew, and Gerald Goldstein. Further understanding of complex information processing in verbal adolescents and adults with autism spectrum disorders. *Autism*, 19(7):859–867, 2015.

Zachary J Williams, Jason L He, Carissa J Cascio, and Tiffany G Woynaroski. A review of decreased sound tolerance in autism: Definitions, phenomenology, and potential mechanisms. *Neuroscience & Biobehavioral Reviews*, 121:1–17, 2021.

Zheng Xu, Xiangfeng Luo, Jie Yu, and Weimin Xu. Mining web search engines for query suggestion. *Concurrency and Computation: Practice and Experience*, 23(10):1101–1113, 2011.

Yahoo! Search query tiny sample, 2009. URL `https://web.archive.org/web/20250517143901/https://webscope.sandbox.yahoo.com/catalog.php?datatype=l`.

Victoria Yaneva and Richard Evans. Six good predictors of autistic text comprehension. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 697–706, 2015.

Victoria Yaneva, Irina Temnikova, and Ruslan Mitkov. Accessible texts for autism: An eye-tracking study. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '15, page 49–57, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450334006. doi: 10.1145/2700648.2809852.

Eldad Yechiam and Elad Yom-Tov. Unique internet search strategies of individuals with self-stated autism: Quantitative analysis of search engine users' investigative behaviors. *Journal of Medical Internet Research*, 23(7):e23829, 2021.

Jinan Zeidan, Eric Fombonne, Julie Scorah, Alaa Ibrahim, Maureen S Durkin, Shekhar Saxena, Afiqah Yusuf, Andy Shih, and Mayada Elsabbagh. Global prevalence of autism: A systematic review update. *Autism research*, 15(5):778–790, 2022.

Huan Zhao, Zhaobo Zheng, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar. Design of a haptic-gripper virtual reality system (hg) for analyzing fine motor behaviors in children with autism. *ACM Transactions on Accessible Computing (TACCESS)*, 11(4):1–21, 2018.

Z. Kevin Zheng, Nandan Sarkar, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar. Cheerbrush: A novel interactive augmented reality coaching system for toothbrushing skills in children with autism spectrum disorder. *ACM Trans. Access. Comput.*, 14(4), October 2021. doi: 10.1145/3481642.

George Kingsley Zipf. *Human behavior and the principle of least effort: An introduction to human ecology.* Ravenio books, 2016.

Appendix A. Subreddits scraped for data collection

Name	Official Description	#Members
r/autism	Autism. Autism news, information and support. Please feel free to submit articles to enhance the knowledge, acceptance, understanding and research of Autism and ASD.	480K
r/aspergers	Aspergers. For those affected by Autism Spectrum Disorder, providing a space for support, discussion, and sharing experiences.	175K
r/AutismInWomen	AutismInWoman. A community for autists that are not cisgender males. We discuss the struggles, triumphs, and mundane life events that come with our autistic experience. We're LGBTQIA+ inclusive. TERFS not welcome. We are open to those who are self-suspecting autism, self-diagnosed as autistic, and formally diagnosed autistic (regardless of age), as we recognize the barriers around formal diagnosis and assessment. Please engage kindly, read our Rules and Wiki pages, and modmail if you have any questions ¡3	229K
r/AutismTranslated	AutismTranslated. If you think you might be autistic - or even if you're on The Quest, to figure out why life seems so much stranger and harder for you than it does for other people - then we made this space for you. It's one thing to read DSM diagnostic criteria or an Autism Parent's lamentations, and another to really hear us as we describe what it feels like to _be_ autistic. Welcome, and please feel free to ask questions! :)	64K
r/AutisticAdults	Adults on the Autistic Spectrum. For adults who are on the autistic spectrum, or think they might be on the autistic spectrum, regardless of diagnostic status. This is a relaxed, rules-light community, preferring discussion rather than memes and media. Non-autistic people are welcome if they are here for the benefit of autistic people (see Rule 1 before posting).	100K
r/AutisticPride	Autistic Pride. A bunch of proud autistic folks. No cure necessary—we're not diseased. We support societal reforms that accommodate both autistic and neurotypical communities equally.	56K

r/aspergirls	Life skills and healthy coping mechanisms for the ASD community. Aspergirls is a place to share advice and tips for topics related to autism and self-improvement. We help with INTERPERSONAL questions/struggles related to autism and life skills, personal growth, healthy coping mechanisms, etc. We are a support community for autists, please remain civil at all times when posting here. Thank you!	100K
r/AskAutism	Ask Autistics. A forum to hear the experiences of autistic individuals, and educate about neurodiversity, human rights concerns, and how to be a better human being both to autistic people and all citizens of the world.	10K
r/autism_controversial	Autism_controversial. This subreddit is for autistic people to talk about controversial topics revolving autism. People with close relations to autistic people are welcome to join the conversation, but please do not speak over actually autistic people.	585
r/AutismCertified	AutismCertified. A community for clinically-dxed autistics. Its purpose is to create a space where autistic representation is not diluted by those who are not officially confirmed to have ASD. — Disclaimer — - If you wish to speak on the autistic experience without having been professionally evaluated and found to have ASD, please go elsewhere.	3K
r/TrueEvilAutism	TrueEvilAutism. A venting and ranting subreddit intended for diagnosed autistics to express their experiences and frustrations	1K
r/AutisticPeeps	AutisticPeeps. This community gives a space for professionally diagnosed autistic people to thrive. Self-diagnosed people are explicitly not allowed to participate in this subreddit, but self-suspicion is fine. Self-suspicion is okay, but self-diagnosis is not.	7K
r/SpicyAutism	SpicyAutism. This is a subreddit for level 2/3/otherwise higher support needs autists, where we are the majority and feel understood and validated. This subreddit is a safe space for all autistic people, family members, doctors, teachers, etc., with the understanding that the priority is the comfort and inclusion of higher support needs autists and our experiences. Here you can ask questions, share experiences, talk about your interests, and more.	21K

r/austismlevel2and3	Autismlevel2and3. A place not overrun with level 1's telling us that Autism "isn't a disability"	1K
r/AutisticLiberation	AutisticLiberation. Welcome to our subreddit! We are a safe space for autistic leftists, regardless of background.	2K
r/Autism_Pride	Autism_Pride. A bunch of autistic leftists fighting against discrimination and celebrating our authentic selves.	4K

Table 8: List of subreddits and their official descriptions used for data collection.

Appendix B. Table of Abbreviations

Acronym	**Full Form**
ASD	Autism Spectrum Disorder
IAS	Information Access System
SE	Search Engine
LLM	Large Language Model
WAI	Web Accessibility Initiative
W3C	World Wide Web Consortium
WCAG	Web Content Accessibility Guidelines
COGA	Cognitive and Learning Disabilities Accessibility Task Force
SERP	Search Engine Result Page
RR	Retrieved Resources
ANS	Answer generated by an LLM in response to a user prompt

Table 9: List of abbreviations used in the manuscript.