JMIR Publications
The Leading eHealth Publisher

From Click to Chat: A Comparative, Bilingual Study of Autism Related Information on Websites and General Purpose AI Chatbots

Valentin Nădășan, Ciprian-Rareș Păroiu, Alexandra Neguțescu, Elena-Gabriela Strete

Submitted to: Journal of Medical Internet Research
on: October 02, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript **5**
Supplementary Files **17**
Figures **18**
Figure 1 ..... **19**
Figure 2 ..... **20**

Preprint JMIR Publications

From Click to Chat: A Comparative, Bilingual Study of Autism Related Information on Websites and General Purpose AI Chatbots

Valentin N?d??an[1*]; Ciprian-Rare? P?roiu[2*]; Alexandra Negu?escu[3]; Elena-Gabriela Strete[4]

[1]Hygiene Department Universitatea de Medicin?, Farmacie, ?tiin?e ?i Tehnologie „George Emil Palade” din Târgu Mure? Târgu Mure? RO
[2]Paychiatry Clinic No 2 Spitalul Clinic Judetean Mures Târgu Mure? RO
[3] Faculty of Medicine Universitatea de Medicin?, Farmacie, ?tiin?e ?i Tehnologie „George Emil Palade” din Târgu Mure? Târgu Mure? RO
[4]Psychiatry Department Universitatea de Medicin?, Farmacie, ?tiin?e ?i Tehnologie „George Emil Palade” din Târgu Mure? Târgu Mure? RO
[*]these authors contributed equally

Corresponding Author:
Ciprian-Rare? P?roiu
Paychiatry Clinic No 2
Spitalul Clinic Judetean Mures
Str. Gheorghe Marinescu, 38
Târgu Mure?
RO

Abstract

Background: Parents increasingly consult the internet —both websites and, more recently, AI chatbots—for information on autism spectrum disorder (ASD). Yet the comparative quality of these two source types, especially across languages, remains under-explored.

Objective: We aimed to assess the completeness and accuracy of ASD information delivered by websites and five popular AI chatbots, and determine whether performance differs between English and Romanian content.

Methods: In a cross-sectional design, we evaluated 25 English-language and 25 Romanian-language websites and the responses of ChatGPT, Gemini, Claude, Copilot, and DeepSeek. Content was benchmarked against a 24-item checklist, yielding completeness and accuracy scores. Chatbots were tested in two scenarios: a single broad query and 24 item-specific queries.

Results: Websites achieved higher completeness in English than in Romanian (6.9 vs 5.1; p=0.007) and marginally higher accuracy (6.9 vs 6.1; p=0.045). In Scenario A, chatbots' completeness (English 5.3 vs Romanian 6.2; p= 0.151), and accuracy (English 6.0 vs Romanian 5.6; p= 0.322) did not show significant differences by language. In the single-query scenario, websites showed higher accuracy than chatbots in both English (6.9 vs 6.0, p=0.19) and Romanian (6.1 vs 5.6, p=0.532), with neither difference reaching statistical significance. Conversely, item-specific questioning favored chatbots, which outperformed websites in English (8.3 vs 6.9, p=0.053) and—significantly—in Romanian (8.0 vs 6.1, p=0.007). Accuracy rose significantly from the single-query to the item-specific scenario (English: 6.0 vs 8.3, p=0.002; Romanian: 5.6 vs 8.0, p=0.001). Between-chatbot variation was statistically significant (One-way repeated measures ANOVA p=0.005), with Gemini and DeepSeek showing superior scores (p=0.016 and 0.023 respectively).

Conclusions: The quality of online ASD information varies by source type and language. English-language websites were more complete than Romanian ones. Targeted questioning of chatbots yielded significantly more accurate answers in both languages.

(JMIR Preprints 02/10/2025:85196)
DOI: https://doi.org/10.2196/preprints.85196

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Preprint JMIR Publications

Original Manuscript

Preprint JMIR Publications

Paper type: Original Paper

From Click to Chat: A Comparative, Bilingual Study of Autism Related Information on Websites and General Purpose AI Chatbots

Valentin Nădășan [1], Ciprian-Rareș Păroiu [2]*, Alexandra Neguțescu[3], Elena-Gabriela Strete [2,4]

[1] Hygiene Department, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania; valentin.nadasan@umfst.ro

[2] Psychiatry Clinic No 2, Mureș County Clinical Hospital, Targu Mures, Romania; raresparoiu@yahoo.com

[3] Faculty of Medicine, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania; alexandranegutescu@yahoo.com

[4] Psychiatry Department, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania; elena.buicu@umfst.ro

* Corresponding Author : email: raresparoiu@yahoo.com phone: 0040752569236

Abstract

Background/Objectives: Parents increasingly consult the internet —both websites and, more recently, AI chatbots—for information on autism spectrum disorder (ASD). Yet the comparative quality of these two source types, especially across languages, remains under-explored. We aimed to assess the completeness and accuracy of ASD information delivered by websites and five popular AI chatbots, and determine whether performance differs between English and Romanian content.

Methods: In a cross-sectional design, we evaluated 25 English-language and 25 Romanian-language websites and the responses of ChatGPT, Gemini, Claude, Copilot, and DeepSeek. Content was benchmarked against a 24-item checklist, yielding completeness and accuracy scores. Chatbots were tested in two scenarios: a single broad query and 24 item-specific queries.

Results: Websites achieved higher completeness in English than in Romanian (6.9 vs 5.1; p=0.007) and marginally higher accuracy (6.9 vs 6.1; p=0.045). In Scenario A, chatbots' completeness (English 5.3 vs Romanian 6.2; p= 0.151), and accuracy (English 6.0 vs Romanian 5.6; p= 0.322) did not show significant differences by language. In the single-query scenario, websites showed higher accuracy than chatbots in both English (6.9 vs 6.0, p=0.19) and Romanian (6.1 vs 5.6, p=0.532), with neither difference reaching statistical significance. Conversely, item-specific questioning favored chatbots, which outperformed websites in English (8.3 vs 6.9, p=0.053) and—significantly—in Romanian (8.0 vs 6.1, p=0.007). Accuracy rose significantly from the single-query to the item-specific scenario (English: 6.0 vs 8.3, p=0.002; Romanian: 5.6 vs 8.0, p=0.001). Between-chatbot variation was statistically significant (One-way repeated measures ANOVA p=0.005), with Gemini and DeepSeek showing superior scores (p=0.016 and 0.023 respectively).

Conclusions: The quality of online ASD information varies by source type and language. English-language websites were more complete than Romanian ones. Targeted questioning of chatbots yielded significantly more accurate answers in both languages.

Keywords: autism spectrum disorder, online health information, chatbots, large language models, information quality, completeness and accuracy, multilingual assessment

1. Introduction

Autism spectrum disorder (ASD) is one of the most common neurodevelopmental disorders, with a steadily increasing prevalence over the past two decades [1]. This diagnosis significantly affects the quality of life of both patients and their caregivers or parents, requiring early detection and a complex, individualized treatment plan. The diagnosis often generates an emotional shock for the family and creates a highly uncertain context [2], leading to the need for additional, clear information and psychoeducational support. Collaboration between the family, professionals (physicians, therapists, psychologists), the educational system, and the social system is essential [3,4].

Consequently, it is natural for parents to seek information from the most accessible sources. As such, the majority of parents turn to the internet for this purpose [1,5,6]. This is partly due to the limited access to medical services, the short duration of specialist consultations, the potential lack of empathy from professionals, and the use of overly technical language that is not easily understood by the general public [1]. On the other hand, online information sources offer several advantages: accessibility, anonymity, and a variety of formats including specialized websites, social networks, forums, mobile applications, and, in the last 2–3 years, AI-based conversational agents, which have become increasingly used for this purpose, even though no official recommendations currently endorse their use [7,8].

The ease of access to this type of information carries a number of significant concerns, particularly regarding the quality of information available to parents—which remains insufficiently studied or regulated. As a result, parents may be exposed to inaccurate, harmful, or even dangerous content [9]. Equally troubling are the myths and misinformation propagated in this space, which can undermine parents' trust in professionals and indirectly lead to delayed or incorrect diagnoses, with potentially devastating consequences throughout the life of both the child and the family [10].

ChatGPT, along with other AI-based conversational models (LLMs), represents one of the current challenges that has drawn attention from healthcare professionals as well [11]. The novel mode of interaction with these easily accessible models—capable of offering human-like answers to almost any question—has led to their rapid, widespread adoption. However, there is considerable uncertainty regarding the impact of this mass usage in the context of medical information. Assessments of ChatGPT's performance have raised concerns about the confidence with which it provides responses, as it can deliver inaccurate answers with persuasive confidence [12]. Moreover, it has underperformed in tasks involving differential diagnosis and, in some cases, has offered potentially harmful treatment recommendations [8].

To date, few studies have assessed the credibility and quality of ASD-related information available on websites [13,14], and the gaps are even more significant concerning the evaluation of autism-related content provided by general-purpose AI conversational agents [1].

Our study aimed to investigate the quality of autism-related information available on two major categories of online sources: websites and general-use AI conversational agents. The primary objective was to assess the quality of ASD-related information provided by online sources, namely websites and chatbots. The secondary objective was to identify potential differences between various websites and chatbots, as well as between the languages in which the information was delivered.

2. Materials and Methods

This research was designed as an observational, cross-sectional study with two core components: the first focusing on the quality of autism-related information available on websites, and the second on the quality of autism-related information provided by AI chatbot interfaces. The study is part of a broader project on the hygiene of medical information that aims to evaluate the quality of online information about the most common diseases and applies a standard methodology described in previously published work [14-16]. The essential elements of the methodology are described below.

Study of Website Articles

The study was conducted in August–September 2019 and included websites that provide information

on ASD. We selected English-language sites—the most widely used language globally—and Romanian-language sites—a language of local importance—to explore possible particularities of online information in a language with limited circulation. To identify the websites, we employed Google search engine, which is used by the vast majority of users. Although studies of user behavior show a clear preference for results in the first 5–10 positions, we included 25 websites for each language to capture less common search patterns, such as extending the search beyond the top 10 Google results [17].

Eligible sites were identified by entering the query term “autism” in English (on the global page, www.google.com) and in Romanian (on the local page, www.google.ro). The term was chosen based on its popularity as indicated by Google Trends, a tool that analyzes the frequency of top searches conducted in Google Search across regions and languages.

For each language, the first 100 results were exported to a PDF file, from which 25 sites were selected according to predefined inclusion and exclusion criteria. To be included, a site had to address ASD, present information in the target language, and contain at least 300 words, with content intended for lay users without medical training. We excluded pages dealing with topics other than autism, sponsored results displayed at the top of the list, and pages that were compromised or inaccessible. We also excluded sites consisting exclusively of audio or video content and those requiring registration or payment for access. Pages that treated the topic only as news pieces or forum/social-media comments—rather than providing an in-depth explanation—were left out of the sample. Eligible sites were analyzed consecutively following a rigorous protocol.

After sampling, each text was evaluated for content quality against a 24-item checklist covering essential ASD concepts. The checklist was developed using peer-reviewed literature and evidence-based guidelines on the subject of interest as sources of information, in such a way that it covered the topic of ASD to an extent considered sufficient and comprehensible for nonprofessionals. It included information on the following aspects: What is autism, and how prevalent is it? What signs and symptoms should prompt suspicion? Which factors raise the risk of developing the condition? How does autism typically progress, and what complications may arise? What treatment options are available? The checklist was critically and independently reviewed by two practicing psychiatrists.

For each item, the presence (coded as 1) or absence (coded as 0) of the information was recorded; the sum represented the absolute completeness score (ACS), which was transformed into a relative completeness score (RCS) ranging from 0 to 10 using the formula:

$$RCS = 10 \times ACS / 24$$

The accuracy of each concept presented on the websites was rated on an ordinal scale from 0 to 2, where 0 indicated incorrect information with minimal or absent details; 1 indicated partially correct information with partial details; and 2 indicated correct information with complete details. The sum of item scores yielded the absolute accuracy score (AAS), which was transformed into a relative accuracy score (RAS) ranging from 0 to 10 using the formula:

$$RAS = 10 \times AAS / MAS$$

where MAS represents the maximum accuracy score, which was a site-specific maximum accuracy score, the highest possible number of points that a specific site could be awarded assuming all the items addressed were totally accurate.

Completeness and accuracy were assessed by two independent raters following identical instructions. Interrater agreement was checked using Cohen’s κ test, with an acceptable agreement threshold of 0.8. For websites with $\kappa < 0.8$, raters re-examined divergent scores and assigned a final score by discussion and consensus.

Study of Chatbot-Generated Responses

The second component of the study was conducted in June–July 2025 and assessed the quality of ASD information provided in English and Romanian by five general-purpose conversational agents: ChatGPT-4.0 (OpenAI), Gemini 2.5 Flash (Google DeepMind), Claude Sonnet 4 (Anthropic), Copilot-ChatGPT 4 Turbo (Microsoft), and DeepSeek-R1 (DeepSeek). Selection criteria for AI chatbots were: relevance and popularity among the general public (ChatGPT, Gemini, and Copilot are among the most widely used conversational interfaces worldwide, while Claude and DeepSeek represent rapidly growing models); diversity of providers and ecosystems (including multiple major and independent companies to reduce the risk of bias associated with a single technological infrastructure); multiplatform accessibility (Windows and Apple operating systems) with web/mobile interfaces accessible to the public; functionality in multiple languages (English and Romanian); and free, globally available access.

Each chatbot was queried under two complementary scenarios: in Scenario A, a single broad question was asked (“What should a parent know about ASD?”), while Scenario B used one specific question for each of the 24 items in the checklist described in the previous section. The questions were drafted by one study author and reviewed by the other authors to simulate how a layperson without medical training would ask.

Quality indicators (completeness and accuracy) and the evaluation procedure were identical to those applied to websites to allow direct comparisons. The only notable difference was that completeness assessment of chatbot responses was relevant only in Scenario A considering that in Scenario B—where one question was posed for each checklist item—the completeness score would have been automatically and invariably 10.

As with the websites, the two dimensions of information quality provided by chatbots were rated by two independent evaluators, who recorded the presence of items and judged accuracy according to the checklist. Interrater agreement was verified using Cohen’s κ, and discrepancies were resolved by the same consensus procedure.

Statistical Analysis

Completeness and accuracy scores were calculated using the formulas mentioned above. Normality of score distributions was tested using the Kolmogorov–Smirnov test. For two-group comparisons with normal distributions, we used the independent or paired t-test, as appropriate; when normality was not met, we used the Mann–Whitney U test. To compare the performance of the five individual chatbots on six quality of information indicators, we applied the one-way repeated-measures ANOVA (within-subjects factor = chatbot). Statistical significance was set at $\alpha = 0.05$.

3. Results

Figure 1 displays the overall completeness and accuracy of the information about ASD available to general users from the investigated sources.

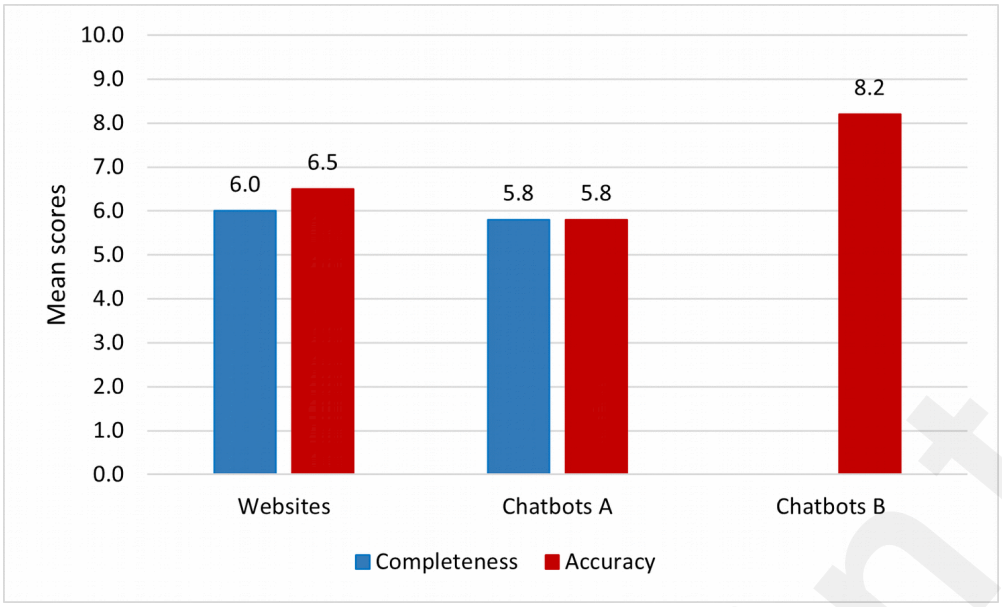


Figure 1. Overall completeness and accuracy scores for information about autism across websites and general-purpose chatbots (A: scenario with a single, general question; B: scenario with multiple, specific questions).

Figure 2 shows the level of completeness and accuracy of information about ASD depending on the language in which the information was presented on the websites and AI chatbots respectively.

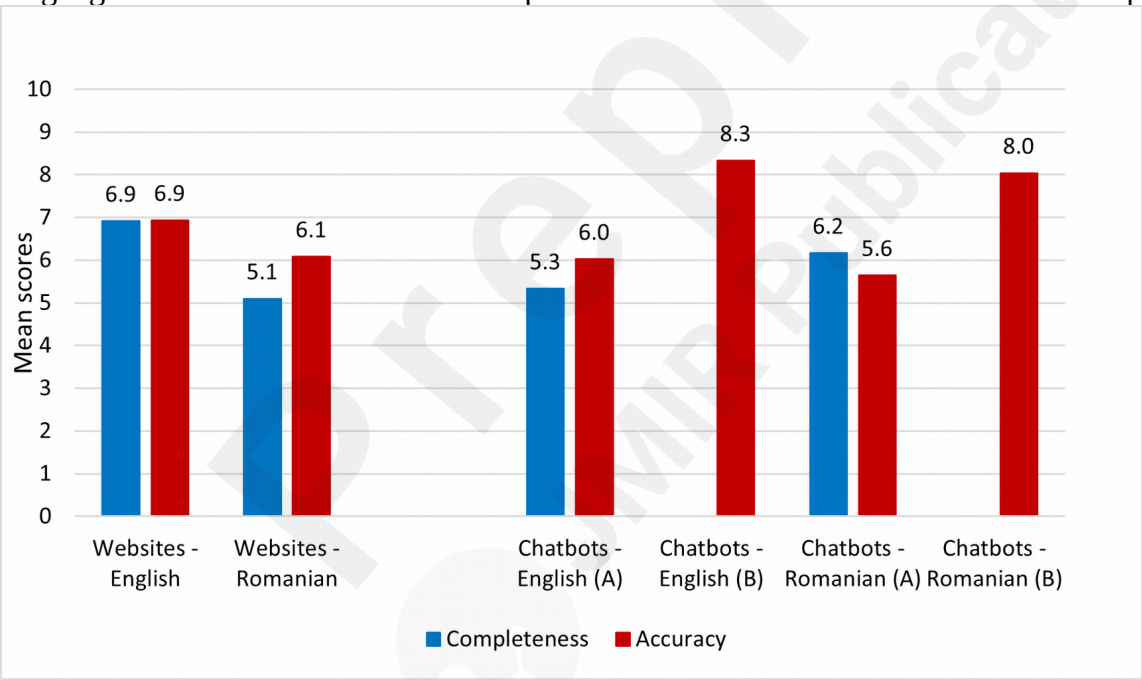


Figure 2. Mean completeness and accuracy scores for information about autism across sources, languages, and querying scenarios (A: scenario with a single, general question; B: scenario with multiple, specific questions).

Table 1 reports completeness and accuracy scores—together with the corresponding statistical tests—comparing English- and Romanian-language results for ASD information, separately for traditional websites and generative-AI chatbots.

Source and scenario	Parameter	English Mean (SD)	Romanian Mean (SD)	p-value
Websites	Completeness	6.9 (2.3)	5.1 (2.2)	0.007
	Accuracy	6.9 (1.5)	6.1 (1.4)	0.045
Chatbots	Completeness	5.3 (1.3)	6.2 (1.3)	0.151

(Scenario A)				
	Accuracy	6.0 (0.3)	5.6 (0.7)	0.322
Chatbots (Scenario B)	Accuracy	8.3 (0.5)	8.0 (1.0)	0.392

Table 1. Comparison of completeness and accuracy scores by language and information source (websites and chatbots).

Scores are expressed in points on a 0–10 scale (0 = minimum, 10 = maximum).

Table 2 shows the results of comparative tests aiming to discover information quality differences between websites and AI chatbots.

Table 2. Comparison of completeness and accuracy scores between websites and chatbots (Scenarios A and B), by language.

Language	Parameters	Websites Mean (SD)	Chatbots Mean (SD)	p-value
English	Completeness	6.9 (2.3)	Scenario A: 5.3 (1.3)	0.152
	Accuracy	6.9 (1.5)	Scenario A: 6.0 (0.3)	0.19
		6.9 (1.5)	Scenario B: 8.3 (0.5)	0.053
Romanian	Completeness	5.1 (2.2)	Scenario A: 6.2 (1.3)	0.313
	Accuracy	6.1 (1.4)	Scenario A: 5.6 (0.7)	0.532
		6.1 (1.4)	Scenario B: 8.0 (1.0)	0.007

Scores are expressed in points on a 0–10 scale (0 = minimum, 10 = maximum).

For English, the five chatbots showed a mean accuracy of 6.0 in Scenario A and 8.3 in Scenario B (p=0.002). For Romanian, the corresponding means were 5.6 in Scenario A and 8.0 in Scenario B (p=0.001).

Table 3 presents the completeness and accuracy scores obtained by each of the five chatbots, reported separately for English and Romanian queries in both test scenarios, with a final row that aggregates these results to show each chatbot's overall performance across languages and scenarios.

Table 3. Completeness, accuracy, and combined performance scores of five chatbots by language and scenario.

Scenario	Parameters/language	Gemini	ChatGPT	DeepSeek	Claude	Copilot
Scenario A	Completeness-English	6.3	6.7	5.0	5.4	3.3
	Completeness-Romanian	7.5	7.1	6.7	4.6	5.0
	Accuracy-English	6.3	5.9	5.8	5.8	6.3
	Accuracy-Romanian	6.4	5.6	6.3	5.0	5.0
Scenario	Accuracy-English	8.3	8.5	8.5	8.8	7.5

B						
	Accuracy-Romanian	9.0	7.9	8.5	8.3	6.5
Overall mean (SD)		7.3 (1.2)	7.0 (1.1)	6.8 (1.5)	6.3 (1.8)	5.6 (1.5)

Scores are expressed in points on a 0–10 scale (0 = minimum, 10 = maximum). The one-way repeated-measures ANOVA showed a p-value of 0.005 in the within-subjects effects test. Pairwise comparisons showed Gemini and DeepSeek were superior to Copilot if unadjusted p values were considered (p=0.016 and p=0.023 respectively). When Bonferroni correction for multiple comparisons was applied, differences between chatbots were no more statistically significant in pairwise comparisons.

4. Discussion

This cross-sectional evaluation compared the completeness and accuracy of autism related information retrieved from traditional websites and five general purpose large language model (LLM) chatbots, in both English and Romanian. Overall, ASD-related information on both English- and Romanian-language websites was moderately complete and accurate, while chatbot responses were of similar or higher quality depending on the prompting strategy. English-language websites offered more complete coverage than Romanian ones. The roughly 2-point gap in completeness between languages likely reflects several missing topic areas, potentially reducing the practical usefulness of the guidance for parents searching in a less-resourced language. Chatbots matched website accuracy in the single-query scenario. However, when prompted item by item, the conversational agents exceeded the websites by more than 20%—enough to turn several partially correct statements into fully accurate ones—with the greatest improvement observed in Romanian. Between chatbot variability was modest; only Gemini and DeepSeek outperformed Copilot on unadjusted pairwise tests, and no differences persisted after conservative correction of the significance threshold. These patterns indicate a shifting information landscape in which conversational AI can partly fill linguistic gaps but still delivers uneven quality when prompted broadly.

Several factors likely explain the observed findings. Websites represent static, human curated repositories; English sites benefit from larger editorial teams, wider peer to peer review, and more abundant source material, hence their higher completeness. Romanian sites probably rely on much smaller teams and budgets, leading to fewer topic sections and slower updates. Chatbots, by contrast, draw on vast multilingual corpora, and transformer models exhibit strong cross lingual transfer [18]. However, their generative nature means that spontaneous responses must compress large knowledge bases into a short answer, which can omit checklist items [19]. When prompted item by item, the model can focus on a narrower retrieval bandwidth, boosting factual density and accuracy. Our observation complements the work of other authors who have shown that different prompts lead large language models to generate responses with varying levels of quality [20-22].

Our findings on the quality of ASD-related online information are consistent with previous studies indicating that large language models—such as GPT-4 and Gemini—generally produce information of moderate to good quality about autism and, more broadly, other mental health disorders [1, 23]. Earlier assessments of website quality for ASD reported moderate reliability using the DISCERN tool (46.5 of a possible 80) compatible, although not closely comparable to our results, due to the different nature of the assessment methodology [13].

Our cross-lingual analysis, which found that English-language sources consistently scored higher than Romanian sites on both completeness and accuracy, mirrors and extends several earlier lines of evidence. First, Chu et al. showed that non-English speakers—particularly Chinese- and Spanish-

language respondents—were more than twice as likely to report frustration, doubt the credibility, and have difficulty understanding online health information, underscoring a perceived quality gap linked to language preference [24]. Second, the classic audit by Berland et al. demonstrated a concrete content deficit: only 22 % of key clinical elements were adequately covered on Spanish pages compared with 45 % on English pages, although accuracy per se was similar [25]. Third, previous studies in various areas, more or less adjacent to psychiatry, suggest that general-purpose AI chatbots are already improving cross-language patient education and communication. In acute ENT settings, for example, ChatGPT has delivered guideline-concordant advice on sudden sensorineural hearing loss in both English and Spanish with similar clinical accuracy [26]. Within dermatology, board-certified clinicians have judged Spanish and Russian versions of patient-education materials generated by the same model to be clinically usable, showing progress toward language-concordant care even if performance for languages such as Mandarin remains uneven [27]. Likewise, GPT-4 has produced multilingual pediatric discharge instructions that faithfully preserve meaning and introduce markedly fewer critical errors than standard machine-translation services, demonstrating its capacity to support caregivers across linguistic boundaries [28].

Parents typically look online soon after suspecting ASD, often before a formal evaluation [1,2]. Incomplete or inaccurate content heightens the risk of delayed diagnosis, adoption of unproven therapies, or abandonment of evidence-based interventions [9,10]. Our finding that Romanian websites scored lowest on completeness underscores the vulnerability of families searching in a less-resourced language. On the other hand, the substantial accuracy gains observed when parents asked chatbots specific questions suggest that AI tools may supplement conventional resources— on condition that users know how to phrase granular prompts. Other prompt engineering techniques documented by various authors may further enhance the quality of AI chatbot answers. However, clinical teams, specifically pediatric doctors and nurses, may need to counsel caregivers on effective, checklist style questioning while emphasizing cross checking with professional advice [29-31].

This study has several limitations that should be considered when interpreting the findings. First, there is a temporal mismatch between the two data sources: websites were sampled in 2019, whereas chatbot responses were collected in 2025. Although web content may have changed in the intervening years, previous longitudinal research suggests that health-related website content tends to remain relatively stable over time [14]. Second, the scope of our language analysis was limited to English and Romanian. While this allowed for a focused comparison between a dominant global language and a less-resourced Central-Eastern European language, the findings may not necessarily extend to other low-resource or non-Indo-European languages. Finally, the assessment of information quality was conducted by clinicians, who evaluated accuracy and completeness from a professional standpoint. As a result, our analysis does not capture how well this information would be understood or perceived by lay users, such as parents or caregivers of autistic children—a perspective that remains essential for evaluating real-world accessibility and utility.

Future research should aim to build on our findings by addressing several important directions. A key next step would be to sample websites at the same time as chatbot queries, allowing for a more accurate comparison of the current digital information landscape and its evolution over time. Expanding the study to include additional languages—particularly low-resource ones—and applying culturally adapted evaluation tools would enhance the generalizability and relevance of results across diverse populations. It is also critical to involve lay stakeholders, such as parents and autistic adults, in the assessment process to better capture perceived trustworthiness, usability, and overall accessibility of the information. Future work should further examine the performance of other large language models and domain-specific medical chatbots, as well as their potential influence on real-world decision-making. In this context, controlled trials comparing structured versus unstructured chatbot use could provide valuable insights into the clinical utility of AI-guided information seeking, particularly in caregiving and support contexts.

5. Conclusions

Overall, autism-related information on English- and Romanian-language websites achieved moderate levels of completeness and accuracy. Chatbot answers matched—and, when guided by detailed, topic-specific prompts, often exceeded—this baseline. Nevertheless, a language gap remained: English pages covered more checklist items than Romanian ones, potentially putting families who search in a low-resource language at a relative disadvantage.

These findings point to two complementary priorities: expanding high-quality web content in under-served languages and helping parents craft precise questions that elicit dependable chatbot responses. Importantly, online resources—whether websites or AI tools—should complement, not replace, an early consultation with a qualified healthcare professional when autism is suspected. Addressing these needs will move us toward a health-information ecosystem that is safer, more equitable, and truly responsive to users.

Author Contributions: Conceptualization, V.N. and R.C.P.; methodology, V.N. and R.C.P.; validation, R.C.P., V.N and A.N.; formal analysis, V.N. and R.C.P.; investigation, V.N., R.C.P. and A.N.; data curation, V.N. and R.C.P.; writing—original draft preparation, V.N. and R.C.P.; writing—review and editing, V.N. and R.C.P., A.N. and EGS; visualization, V.N and R.C.P.; supervision, EGS. All authors have read and agreed to the published version of the manuscript

Acknowledgments: During the preparation of this manuscript, the author(s) used ChatGPT 4o for the purposes of translating and improve the original manuscript for the purposes of translating the original Romanian text into English and refining language, style, and grammar. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available from the corresponding authors upon reasonable request.

Abbreviations

The following abbreviations are used in this manuscript:

ASD	Autism Spetcrum Disorder
AI	Artificial Intelligence
LLM	Large Language Model
ACS	Absolute Completeness Score
RCS	relative completeness score
AAS	Absolute Accuracy Score
RAS	Relative Accuracy Score
MAS	Maximum Accuracy Score
SD	Standard Deviation

References

1. McFayden T.C.; Bristol S.; Putnam O.; Harrop C. ChatGPT: Artificial intelligence as a potential tool for parents seeking information about autism. Cyberpsychol. Behav. Soc. Netw. 2024, 27, 135–148. https://doi.org/10.1089/cyber.2023.0202.
2. Rabba A.S.; Dissanayake C.; Barbaro J. Development of a web-based resource for parents of young children newly diagnosed with autism: participatory research design. JMIR Pediatr. Parent. 2020, 3, e15786. https://doi.org/10.2196/15786.
3. Beauchamp T.L.; Childress J.F. Principles of biomedical ethics, 5th ed.; Oxford University

Press: Oxford, UK, 2001. ISBN: 9780195143324.
4. Alencar D.D.C.; Sousa Neto A.R.; Ibiapina A.R.S.; de Oliveira J.D. Information needs for parents of children with autism spectrum disorder: a qualitative study. Cogitare Enferm. 2025, 30, e97345pt. https://doi.org/10.5380/ce.v30i0.97345.
5. Slavych B.; Oliver C.; McIntyre N.S.; Zraick R.I. Assessment of online resources and education materials relating to autism spectrum disorder. Health Educ. J. 2024, 83, 101–120. https://doi.org/10.1177/00178969251328913)
6. Esmaeilzadeh P.; Maddah M.; Mirzaei T. Using AI chatbots (e.g., CHATGPT) in seeking health-related information online: the case of a common ailment. Comput. Hum. Behav. Artif. Hum. 2025, 3, 100127. https://doi.org/10.1016/j.chbah.2024.100127.
7. Sinha J.; Serin N. Online health information seeking and preventative health actions: cross-generational online survey study. J. Med. Internet Res. 2024, 26, e48977. https://doi.org/10.2196/48977.
8. Li J.; Dada A.; Puladi B.; Kleesiek J.; Egger J. ChatGPT in healthcare: a taxonomy and systematic review. Comput. Methods Programs Biomed. 2024, 245, 108013. https://doi.org/10.1016/j.cmpb.2023.108013.
9. Coughler C.; Burke S.M.; Oram Cardy J. Analysis of the quality of online resources for parents of children who are late to talk. Autism Dev. Lang. Impair. 2020, 5, 2396941520917940. https://doi.org/10.1177/2396941520917940.
10. French B.; Daley D.; Groom M.; Cassidy S. Risks associated with undiagnosed ADHD and/or autism: a mixed-method systematic review. J. Atten. Disord. 2023, 27, 1393–1410. https://doi.org/10.1177/10870547231176862.
11. Younis H.A.; Eisa T.A.E.; Nasser M.; Sahib T.M.; Noor A.A.; Alyasiri O.M.; Salisu S.; Hayder I.M.; Younis H.A. A systematic review and meta-analysis of artificial intelligence tools in medicine and healthcare: applications, considerations, limitations, motivation and challenges. Diagnostics 2024, 14, 109. https://doi.org/10.3390/diagnostics14010109.
12. Dayawansa S.; Mantziaris G.; Sheehan J. Chat GPT versus human touch in stereotactic radiosurgery. J. Neuro-Oncol. 2023, 163, 481–483. https://doi.org/10.1007/s11060-023-04353-z.
13. Grant N.; Rodger S.; Hoffmann T. Evaluation of autism-related health information on the web. J. Appl. Res. Intellect. Disabil. 2015, 28, 276–282. https://doi.org/10.1111/jar.12127.
14. Nădăşan V.; Roşca A.N.; Tarcea M.; Ábrám Z.; Măruşteri M. The quality of Romanian breast cancer websites: a five-year longitudinal assessment. J. Cancer Educ. 2018, 33, 703–707. https://doi.org/10.1007/s13187-016-1145-8 .
15. Șulea C.M.; Nădășan V.; Ursachi T.; Toboltoc P.C.; Benedek T. What patients find on the internet when looking for information about percutaneous coronary intervention: multilanguage cross-sectional assessment. J. Med. Internet Res. 2022, 24, e41219. https://doi.org/10.2196/41219.
16. Popescu S.D.; Popescu A.O.; Dănilă M.; Dobria M.; Maior D.; Nădășan V. General characteristics and quality of stroke-related online information—a cross-sectional assessment of the Romanian and Hungarian websites. Acta Med. Marisiensis 2018, 64, 116–120. https://doi.org/10.2478/amma-2018-0023.
17. Eysenbach G.; Köhler C. How do consumers search for and appraise health information on the World Wide Web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002, 324, 573–577. https://doi.org/10.1136/bmj.324.7337.573.
18. Raiaan M.A.K.; Mukta M.S.H.; Fatema K.; Fahad N.M.; Sakib S.; Mim M.M.J.; Ahmad J.; Ali M.E.; Azam S. A review on large language models: architectures, applications, taxonomies, open issues and challenges. IEEE Access 2024, 12, 26839–26874. https://doi.org/10.1109/ACCESS.2024.3365742.
19. Tie J.; Yao B.; Li T.; Ahmed S.I.; Wang D.; Zhou S. LLMs are imperfect, then what? An empirical study on LLM failures in software engineering. arXiv 2024, arXiv:2411.09916. https://doi.org/10.48550/arXiv.2411.09916.

20. Wang L.; Chen X.; Deng X.; Wen H.; You M.; Liu W.; Li Q.; Li J. Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. NPJ Digit. Med. 2024, 7, 41. https://doi.org/10.1038/s41746-024-01029-4.
21. Nguyen D.; MacKenzie A.; Kim Y.H. Encouragement vs. liability: how prompt engineering influences ChatGPT-4's radiology exam performance. Clin. Imaging 2024, 115, 110276. https://doi.org/10.1016/j.clinimag.2024.110276.
22. Masur L.; Driller M.; Suppiah H.; Matzka M.; Sperlich B.; Düking P. Assessment of recommendations provided to athletes regarding sleep education by GPT-4o and Google Gemini: comparative evaluation study. JMIR Form Res. 2025, 9, e71358. https://doi.org/10.2196/71358.
23. Porwal G.; Jeenger J. Evaluating the quality of mental health information generated by large language model chatbots. medRxiv 2024, preprint. https://doi.org/10.1101/2024.12.20.24319373.
24. Chu J.N.; Sarkar U.; Rivadeneira N.A.; Hiatt R.A.; Khoong E.C. Impact of language preference and health literacy on health information-seeking experiences among a low-income, multilingual cohort. Patient Educ. Couns. 2022, 105, 1268–1275. https://doi.org/10.1016/j.pec.2021.08.028.
25. Berland G.K.; Elliott M.N.; Morales L.S.; Algazy J.I.; Kravitz R.L.; Broder M.S.; Kanouse D.E.; Muñoz J.A.; Puyol J.A.; Lara M.; Watkins K.E.; Yang H.; McGlynn E.A. Health information on the internet: accessibility, quality, and readability in English and Spanish. JAMA 2001, 285, 2612–2621. https://doi.org/10.1001/jama.285.20.2612.
26. Ajit-Roger E.; Moise A.; Peralta C.; Orishchak O.; Daniel S.J. Enhancing multilingual patient education: ChatGPT's accuracy and readability for SSNHL queries in English and Spanish. OTO Open 2024, 8, e70048. https://doi.org/10.1002/oto2.70048.
27. Dzuali F.; Seiger K.; Novoa R.; Aleshin M.; Teng J.; Lester J.; Daneshjou R. ChatGPT may improve access to language-concordant care for patients with non-English language preferences. JMIR Med. Educ. 2024, 10, e51435. https://doi.org/10.2196/51435.
28. Brewster R.C.L.; Gonzalez P.; Khazanchi R.; Butler A.; Selcer R.; Chu D.; Aires B.P.; Luercio M.; Hron J.D. Performance of ChatGPT and Google Translate for pediatric discharge instruction translation. Pediatrics 2024, 154, e2023065573. https://doi.org/10.1542/peds.2023-065573.
29. Yau J.Y.S.; Saadat S.; Hsu E.; Murphy L.S.L.; Roh J.S.; Suchard J.; Tapia A.; Wiechmann W.; Langdorf M.I. Accuracy of prospective assessments of four large language model chatbot responses to patient questions about emergency care: experimental comparative study. J. Med. Internet Res. 2024, 26, e60291. https://doi.org/10.2196/60291.
30. Aqavil-Jahromi S.; Eftekhari M.; Akbari H.; Aligholi-Zahraie M. Evaluation of correctness and reliability of GPT, Bard, and Bing chatbots' responses in basic life support Scenarios. Sci. Rep. 2025, 15, 11429. https://doi.org/10.1038/s41598-024-82948-w.
31. Neo J.R.E.; Ser J.S.; Tay S.S. Use of large language model-based chatbots in managing the rehabilitation concerns and education needs of outpatient Stroke survivors and caregivers. Front. Digit. Health 2024, 6, 1395501. https://doi.org/10.3389/fdgth.2024.1395501.

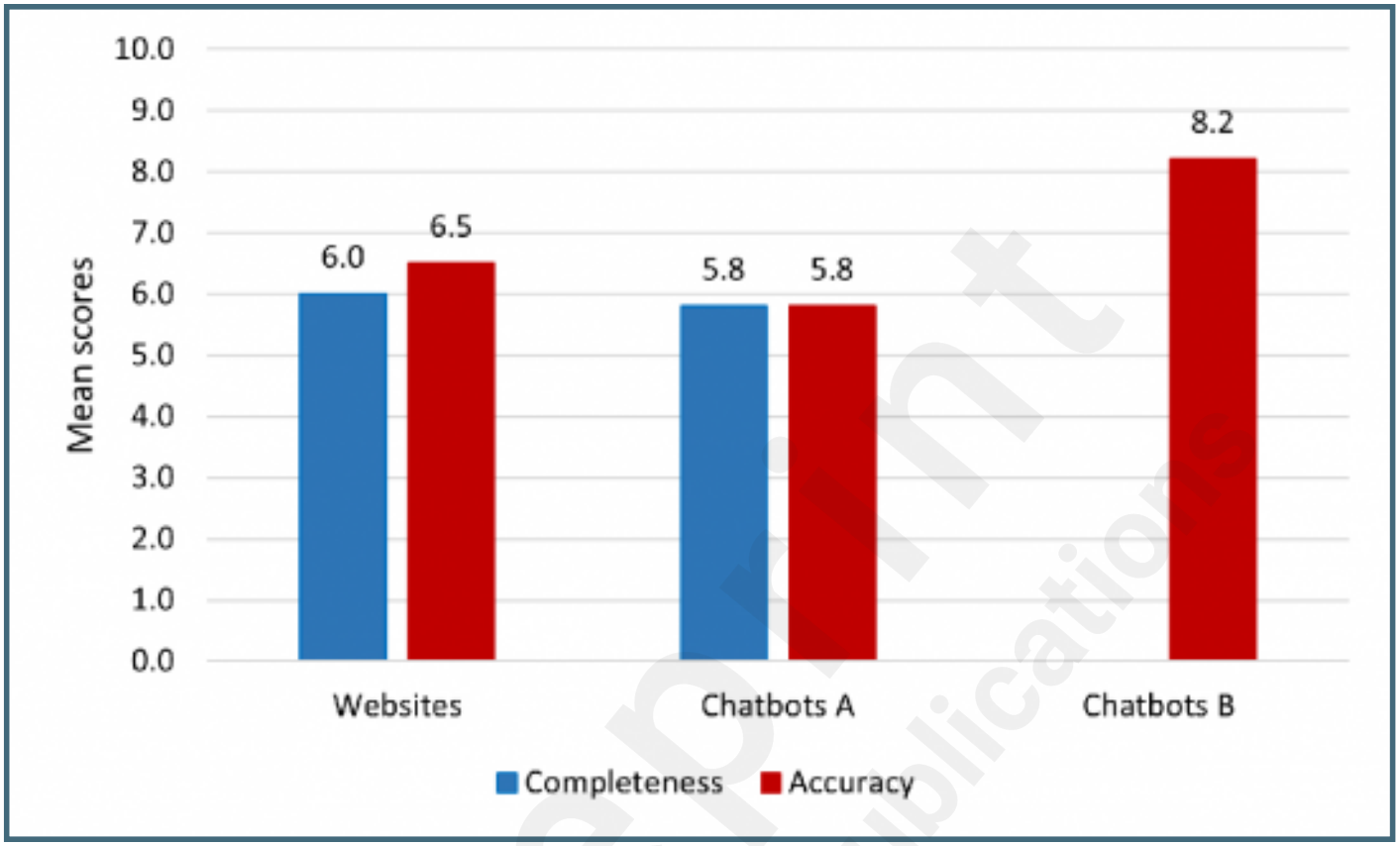
Supplementary Files

Preprint JMIR Publications

Figures

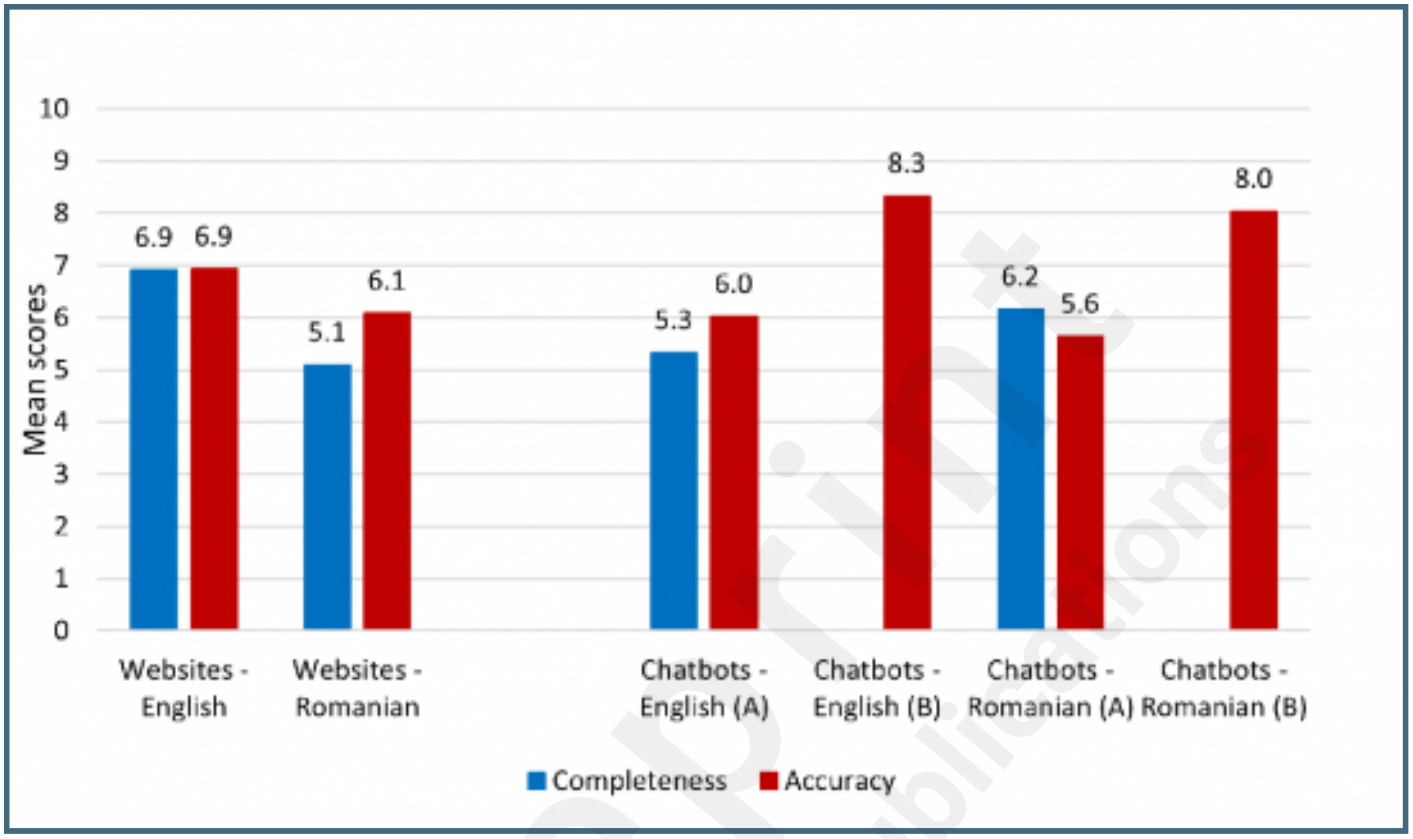
Preprint JMIR Publications

Overall completeness and accuracy scores for information about autism across websites and general-purpose chatbots (A: scenario with a single, general question; B: scenario with multiple, specific questions).



Preprint JMIR Publications

Mean completeness and accuracy scores for information about autism across sources, languages, and querying scenarios (A: scenario with a single, general question; B: scenario with multiple, specific questions).



Preprint JMIR Publications