



ChatGPT vs. neurologists: a cross-sectional study investigating preference, satisfaction ratings and perceived empathy in responses among people living with multiple sclerosis

Elisabetta Maida¹ · Marcello Moccia^{2,3} · Raffaele Palladino^{4,5} · Giovanna Borriello⁶ · Giuseppina Affinito⁴ · Marinella Clerico⁷ · Anna Maria Repice⁸ · Alessia Di Sapio⁹ · Rosa Iodice¹⁰ · Antonio Luca Spiezia³ · Maddalena Sparaco¹ · Giuseppina Miele¹ · Floriana Bile¹ · Cristiano Scandurra¹⁰ · Diana Ferraro^{11,12} · Maria Laura Stromillo¹³ · Renato Docimo¹⁴ · Antonio De Martino¹⁵ · Luca Mancinelli¹⁶ · Gianmarco Abbadessa^{1,17} · Krzysztof Smolik¹² · Lorenzo Lorusso¹⁸ · Maurizio Leone¹⁹ · Elisa Leveraro^{20,21} · Francesca Lauro¹⁰ · Francesca Trojsi¹ · Lidia Mislin Streito⁷ · Francesca Gabriele²² · Fabiana Marinelli²³ · Antonio Ianniello⁶ · Federica De Santis²⁴ · Matteo Foschi^{22,25} · Nicola De Stefano¹³ · Vincenzo Brescia Morra¹⁰ · Alvino Bisecco¹⁴ · Giancarlo Coghe²⁶ · Eleonora Cocco²⁶ · Michele Romoli¹⁶ · Francesco Corea²⁷ · Letizia Leocani^{28,29} · Jessica Frau²⁶ · Simona Sacco²² · Matilde Inglese^{20,21} · Antonio Carotenuto¹⁰ · Roberta Lanzillo¹⁰ · Alessandro Padovani^{30,31} · Maria Triassi⁴ · Simona Bonavita¹  · Luigi Lavorgna¹ on behalf of Digital Technologies, Web, Social Media Study Group of the Italian Society of Neurology (SIN)

Received: 15 January 2024 / Revised: 11 March 2024 / Accepted: 12 March 2024
© The Author(s) 2024

Abstract

Background ChatGPT is an open-source natural language processing software that replies to users' queries. We conducted a cross-sectional study to assess people living with Multiple Sclerosis' (PwMS) preferences, satisfaction, and empathy toward two alternate responses to four frequently-asked questions, one authored by a group of neurologists, the other by ChatGPT.

Methods An online form was sent through digital communication platforms. PwMS were blind to the author of each response and were asked to express their preference for each alternate response to the four questions. The overall satisfaction was assessed using a Likert scale (1–5); the Consultation and Relational Empathy scale was employed to assess perceived empathy.

Results We included 1133 PwMS (age, 45.26 ± 11.50 years; females, 68.49%). ChatGPT's responses showed significantly higher empathy scores (Coeff = 1.38; 95% CI = 0.65, 2.11; $p > z < 0.01$), when compared with neurologists' responses. No association was found between ChatGPT' responses and mean satisfaction (Coeff = 0.03; 95% CI = - 0.01, 0.07; $p = 0.157$). College graduate, when compared with high school education responder, had significantly lower likelihood to prefer ChatGPT response (IRR = 0.87; 95% CI = 0.79, 0.95; $p < 0.01$).

Conclusions ChatGPT-authored responses provided higher empathy than neurologists. Although AI holds potential, physicians should prepare to interact with increasingly digitized patients and guide them on responsible AI use. Future development should consider tailoring AIs' responses to individual characteristics. Within the progressive digitalization of the population, ChatGPT could emerge as a helpful support in healthcare management rather than an alternative.

Keywords Artificial intelligence · Machine learning · Multiple sclerosis · Large language model

Introduction

Clinical practice is quickly changing following digital and technological advances, including artificial intelligence (AI) [1–4]. ChatGPT [5] is an open-source generative processing AI software that proficiently replies to users' queries, in a human-like modality [6]. Since its first

Elisabetta Maida and Marcello Moccia contributed equally to this work as co-first authors.

Extended author information available on the last page of the article

iteration (November 2022), ChatGPT's popularity has been growing exponentially. In its most recent release (January 2023), the number of users exceeded the threshold of 100,000 [7]. ChatGPT was not specifically developed to provide healthcare opinion, but replies to a wide range of questions, including those health-related [8]. Thus, ChatGPT could represent an easily-accessible resource to seek health information and advice [9, 10].

The search for health information is particularly relevant in people living with chronic diseases [11], inevitably facing innumerable challenges, including communication with their healthcare providers. We decided to focus on people living with Multiple Sclerosis (PwMS), as a model of chronic disease that can provide insights applicable to the broader landscape of chronic diseases. The young age of onset of MS results in high patients' digitalization, including the use of mobile health apps, remote monitoring devices, and AI-based tools [12]. The increasing engagement by patients with AI platforms to ask questions related to their MS is a reality that clinicians will probably need to confront.

We conducted a comparative analysis to investigate the perspective of PwMS towards two alternate responses to four frequently-asked health-related questions. The responses were authored by a group of neurologists and by ChatGPT, with PwMS unaware of whether they were formulated by neurologists or generated by ChatGPT. The aim was to assess patients' preferences, overall satisfaction, and perceived empathy between the two options.

Methods

Study design and form preparation

This is an Italian multicenter cross-sectional study, conducted from 09/01/2023 to 09/20/2023. The study conduction and data presentation followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statements [13].

The study was conducted within the activities of the "Digital Technology, Web and Social Media" study group [14], which includes 205 neurologists, affiliated with the Italian Society of Neurology (SIN). The study invitation was disseminated to all members of the group via the official mailing list.

Following the invitation to participate, neurologists were required to meet the following criteria:

- i. Dedicate over 50% of their clinical time to MS care and be active outpatient practitioners;

- ii. Regularly receive and respond to patient-generated emails or engage with patients on web platform or other social media.

Only the 34 neurologists who met the specified criteria were included and, using Research Randomizer [15], were randomly assigned to four groups. Demographic information is presented in Table 1 of Online Resources.

- Group A, including 5 neurologists, was required to identify a list of frequently-asked questions based on the actual queries of PwMS received via e-mail in the preceding 2 months. The final list, drafted by Group A, was composed of fourteen questions.
- Group B, including 19 neurologists, had to identify the four questions they deemed the most common and relevant for clinical practice (from the fourteen elaborated by Group A). Group B was deliberately designed as the largest group to ensure a more comprehensive and representative selection of the questions. The four identified questions were 1. "I feel more tired during summer season, what shall I do?"; 2. "I have had new brain MRI, and there is one new lesion. What should I do?"; 3. "Recently, I've been feeling tired more easily when walking long distances. Am I relapsing?"; 4. "My primary care physician has given me a prescription for an antibiotic for a urinary infection. Is there any contraindication for me?".
- Group C, including 5 neurologists, focused on elaborating the responses to the four questions identified as the most common by Group B. The responses were collaboratively formulated through online meetings. Any discrepancies were addressed through discussion and consensus.

Afterwards, the same questions were submitted to ChatGPT 3.5, which provided its own version of the answers. Hereby:

- Group D, including 5 neurologists, carefully reviewed the responses generated by ChatGPT to identify any inaccuracies in medical information or discrepancies from the recommendations before submitting to PwMS (none were identified and, thus, no changes were required).

Questions and answers are presented in the full version of the form in Online Resources.

Subsequently, we designed an online form to explore the perspective of PwMS on the two alternate responses to the four common questions, those authored by Group C neurologists and the others by the open-source AI tool (ChatGPT). PwMS were unaware of whether the responses were formulated by neurologists or generated by ChatGPT. The workflow process is illustrated in Fig. 1 in Online Resources.

The study was conducted in accordance with the guidelines of the Declaration of Helsinki involving human

subjects and the patient's informed consent was obtained at the outset of the survey. The Ethical Committee of the University of Campania "Luigi Vanvitelli" approved the study (protocol number 0014460/i).

MS population and variables

PwMS were invited to participate to the study by their neurologists, through different communication tools, such as institutional e-mail and instant messaging platform. A total of 2854 invites were sent from 09/01/2023 to 09/20/2023.

The study covariates included demographic information (year of birth, sex, area of residence in Italy, and level of education, defined as elementary school, middle school, high school graduate or college graduate, the latter encompassing both a bachelor's degree and post-secondary education) and clinical characteristics (depressed mood, subjective memory and attention deficits, year of MS diagnosis, MS clinical descriptors, such as relapsing–remitting—RRMS, secondary–progressive—SPMS, primary–progressive—PPMS, or “I don't know”). The occurrence of depressive mood was surveyed using the Patient Health Questionnaire (PHQ)-2 scale [16]. The rationale behind the decision to employ this rating scale was the rapidity of completion and its widespread use in the previous online studies including PwMS [17, 18]. Subjective memory and attention impairment deficits were investigated by directly asking patients about their experience on these symptoms (yes/no).

Preference between alternate responses was investigated by asking patients to express their choice. Furthermore, for each response, a Likert scale ranging from 1 to 5 was provided to assess overall satisfaction (higher scores indicating higher satisfaction). The Consultation and Relational Empathy (CARE) scale [19] was employed to evaluate the perceived empathy of the different responses (higher scores indicating higher empathy). The CARE scale measures empathy within the context of the doctor–patient relationship, and was ultimately selected for its intuitiveness, easiness of completion, and because already used in online studies [20, 21]. Given the digital nature of our research, we made a single wording adjustment to the CARE scale to better align to our study. Further details on the form (Italian original version and English translated version) and measurement scales are presented in Online Resources.

Prior to submitting the form to the patients, overall readability level was assessed. All responses elaborated by the neurologists and by ChatGPT were analysed by two validated tools for the Italian language: the Gulpease index [22] and the Read-IT scores (version 2.1.9) [23]. This step was deemed meaningful for a thorough and comprehensive appraisal of all possible factors that could influence patients' perceptions.

Statistical analysis

The study variables were described as mean (standard deviation), median (range), or number (percent), as appropriate.

The likelihood of selecting the ChatGPT response for each question was evaluated through a stratified analysis employing logistic regression models; this approach was adopted to address variations in the nature of the questions. The selection rate of answers generated by ChatGPT was assessed using Poisson regression models. The continuous outcomes (average satisfaction and average CARE scale scores for ChatGPT responses) were assessed using mixed linear models with robust standard errors accounting for heteroskedasticity across patients. Covariates were age, sex, treatment duration, clinical descriptors, presence of self-reported cognitive deficit, presence of depressive symptoms and educational attainment. The software consistently selected the first level of the categorical variable, alphabetically or numerically, as the reference group to ensure straightforward interpretation of coefficients or effects.

The results were reported as adjusted coefficient (Coeff), odds ratio (OR), incidence rate ratio (IRR), 95% confidence intervals (95% CI), and *p* values, as appropriate. The results were considered statistically significant for $p < 0.05$. Statistical analyses were performed using Stata 17.0.

Results

The study included 1133 PwMS (age, 45.26 ± 11.50 years; females, 68.49%), with an average response rate of 39.70%. Demographic and clinic characteristics are summarized in Table 1.

Table 2 provides an overview of participant preferences, mean satisfaction (rated on a Likert scale ranging from 1 to 5), and CARE scale scores for each response by ChatGPT and neurologists to the four questions.

Univariate analyses did not show significant differences in preferences. However, after adjusting for factors potentially influencing the outcome, emerged that the likelihood of selecting ChatGPT response was lower for college graduates when compared with respondent with high school education (IRR = 0.87; 95% CI = 0.79, 0.95; $p < 0.01$).

Further analysis of each singular question resulted in additional findings summarized in Table 3.

Although there was no association between the ChatGPT responses and satisfaction (Coeff = 0.03; 95% CI = -0.01, 0.07; $p = 0.157$), they exhibited higher CARE scale scores (Coeff = 1.38; 95% CI = 0.65, 2.11; $p > z < 0.01$), as compared to the responses processed by neurologists. The findings are summarized in Table 4.

The readability of the answers provided by ChatGPT and neurologists was medium, as assessed with the Gulpease

Table 1 Demographic and clinical characteristics of the study population ($N=1133$)

Age, mean (SD), years	45.26 (11.50)
Female, No (%)	776 (68.49)
Education, No (%)	
Elementary school	7 (0.62)
Middle school	134 (11.83)
High school	575 (50.75%)
College graduates	417 (36.80%)
Geographical origin, No (%)	
Northern-Italy	247 (21.80)
Centre-Italy	527 (46.51)
Southern-Italy	282 (24.89)
Island	77 (6.80)
Duration of MS disease, mean (SD), years	12.28 (9.61)
MS clinical descriptors, No (%)	
RRMS	800 (70.61)
SPMS	100 (8.83)
PPMS	61 (5.38)
“I don’t know”	172 (15.18)

MS Multiple sclerosis, n : number, PPMS primary progressive multiple sclerosis, RRMS relapsing–remitting multiple sclerosis, SD standard deviation, SPMS secondary-progressive multiple sclerosis

Table 2 Participant preferences, mean satisfaction (rated on a Likert scale ranging from 1 to 5), and CARE scale mean scores for each response

	ChatGPT	Neurologist
Question 1		
Preference, n (%)	779 (68.76%)	354 (31.24%)
Likert scale (1–5), mean (SD)	3.63 (1.03)	3.02 (1.19)
CARE scale, mean (SD)	34.24 (8.09)	28.91 (9.47)
Question 2		
Preference, n (%)	672 (59.31%)	461 (40.69%)
Likert scale (1–5), mean (SD)	3.62 (1.22)	3.30 (1.09)
CARE scale, mean (SD)	31.25 (9.71)	30.89 (8.53)
Question 3		
Preference, n (%)	437 (38.57%)	696 (61.43%)
Likert scale (1–5), mean (SD)	3.22 (1.08)	3.63 (1.19)
CARE scale, mean (SD)	32.38 (9.02)	31.58 (9.51)
Question 4		
Preference, n (%)	430 (37.95%)	703 (62.05%)
Likert scale (1–5), mean (SD)	3.17 (1.24)	3.57 (1.04)
CARE scale, mean (SD)	29.54 (9.74)	30.87 (8.61)

n Number, SD standard deviation

Index. Although similar, Gulpease indices were slightly higher for ChatGPT’s responses than for neurologists (ChatGPT: from 47 to 52; neurologists: from 40 to 44). The results were corroborated by ReadIT scores, which are

inversely correlated with Gulpease Index. Table 5 shows the readability of each response.

Discussion

The Internet and other digital tools, such as AI, have become a valuable source of health information [24, 25]. Seeking answers online requires minimal effort and guarantees immediate results, making it more convenient and faster than contacting healthcare providers. AIs, like ChatGPT can be viewed as a new, well-structured search engine with a simplified, intuitive interface. This allows patients to submit questions and receive direct answers, eliminating the need to navigate multiple websites [26]. However, there is the risk that internet and AI-based health information provides incomplete or incorrect information, along with potentially reduced empathy of communication [27–29]. Our study examined participant preferences, satisfaction ratings and perceived empathy regarding responses generated by ChatGPT as compared to those from neurologists. Interestingly, although ChatGPT responses did not significantly affect satisfaction levels, they were perceived as more empathetic compared to responses from neurologists. Furthermore, after adjusting for confounding factors, including education level, our results revealed that college graduates showed less inclination to choose ChatGPT responses compared to individuals with a high school education. This highlights how individual preferences are not deterministic but could instead be influenced by a variety of factors, including age, education level, and others [30, 31].

In line with the previous study [32], ChatGPT provided sensible responses, which were deemed more empathetic than those authored by neurologists. A plausible explanation for this outcome may lie in the observation that ChatGPT’s responses showed a more informal tone¹ when addressing patients’ queries. Furthermore, ChatGPT tended to include empathetic remarks and language implying solidarity (i.e., a welcoming remark of gratitude and a sincere-sounding invitation for further communication). Thus, PwMS, especially those with lower level of education, might perceive confidentiality and informality as empathy. Moreover, the lower empathy shown by neurologists could be related to job well-being factors, including feelings of overwhelming and work overload (i.e., allocation of time to respond to patients’ queries). Even though our study did not aim to identify the reasons behind participants’ ratings,

¹ For clarification, in Italian grammar, the use of the second or third person is based on levels of formality and familiarity. While the second person singular is more informal, the third person singular is employed in more formal or professional contexts, reflecting a respectful or polite tone.

Table 3 Multivariate logistic regressions

	All questions		Question 1		Question 2		Question 3		Question 4	
	IRR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value
Age, years	0.99 (0.99–1.00)		0.381 0.98 (0.97–0.99)	0.021	1.00 (0.99–1.02)	0.273	0.98 (0.97–1.00)	0.066	1.00 (0.98–1.01)	0.796
Sex (Ref. other)										
Female	1.48 (0.68–3.19)		0.316 3.57 (0.50–25.60)		0.204 0.59 (0.06–5.83)	0.659	2.31 (0.20–26.06)	0.497	0.92 (0.70–1.21)	0.575
Male	1.53 (0.71–3.31)		0.277 4.23 (0.58–30.47)		0.152 0.55 (0.05–5.46)	0.617	2.73 (0.24–31.03)	0.416	0.82 (0.81–1.35)	0.574
Education (Ref. high school graduate)										
Elementary school	0.85 (0.49–1.48)		0.576 1.46 (0.26–8.15)		0.662 0.29 (0.06–1.39)	0.124	2.64 (0.54–12.85)	0.229	0.17 (0.02–1.47)	0.108
Middle school	1.02 (0.90–1.17)		0.675 1.16 (0.74–1.79)		0.506 0.84 (0.56–1.25)	0.398	1.19 (0.80–1.76)	0.379	1.11 (0.75–1.66)	0.087
College graduates	0.87 (0.79–0.95)		< 0.01 0.64 (0.48–0.85)		< 0.01 0.72 (0.55–0.95)	0.019	0.80 (0.61–1.05)	0.108	0.78 (0.59–1.03)	0.584
Duration of MS disease, years	1.00 (0.99–1.00)		0.711 1.00 (0.98–1.01)		0.736 0.99 (0.98–1.01)	0.711	1.01 (0.99–1.02)	0.066	0.99 (0.98–1.01)	0.502
MS clinical descriptors (Ref. “PPMS”)										
RRMS	0.86 (0.72–1.03)		0.114 0.45 (0.22–0.90)		0.025 0.66 (0.37–1.19)	0.169	0.78 (0.44–1.37)	0.460	0.93 (0.52–1.64)	0.804
SPMS	0.90 (0.72–1.13)		0.383 0.36 (0.16–0.80)		0.012 0.83 (0.41–1.68)	0.610	0.47 (0.23–0.96)	0.038	2.19 (1.10–4.35)	0.025
“I don’t know”	0.91 (0.74–1.11)		0.365 0.44 (0.20–0.93)		0.032 0.78 (0.41–1.48)	0.451	0.79 (0.42–1.47)	0.460	1.31 (0.70–2.47)	0.393
Depression	0.94 (0.86–1.02)		0.173 0.67 (0.51–0.89)		< 0.01 1.08 (0.84–1.40)	0.526	0.81 (0.63–1.05)	0.127	0.93 (0.71–1.27)	0.624
Subjective memory and attention deficit	1.03 (0.94–1.13)		0.423 0.79 (0.56–1.04)		0.097 1.39 (1.06–1.81)	0.014	0.84 (0.64–1.09)	0.203	1.45 (1.11–1.90)	< 0.01

The main outcome (likelihood of selecting the ChatGPT response) was evaluated through a stratified analysis employing logistic regression models, by assessing the frequency of selecting the ChatGPT option divided by the total number of questions. This approach was adopted to address variations in the nature of the questions. Covariates were age, sex, treatment duration, clinical descriptors, self-reported cognitive deficit, depressive symptoms and educational attainment

CI Confidence interval, IRR incidence rate ratio, MS multiple sclerosis, OR odds ratio, PPMS primary progressive multiple sclerosis, RRMS relapsing–remitting multiple sclerosis, SPMS secondary progressive multiple sclerosis; the use of bold formatting within the table was employed to highlight significant values

Table 4 Mixed linear regressions

	Satisfaction		CARE scale	
	Coeff (95% CI)	<i>p</i> value	Coeff (95% CI)	<i>p</i> value
ChatGPT (Ref. neurologists)	0.03 (− 0.01 to 0.07)	0.157	1.37 (0.64 to 2.11)	< 0.01
Age, years	0.00 (− 0.00 to 0.00)	0.441	0.00 (− 0.03 to 0.04)	0.867
Sex				
Female	− 0.33 (− 0.91 to 0.25)	0.265	− 5.56 (− 11.61 to 0.48)	0.071
Male	− 0.35 (− 0.93 to 0.23)	0.241	− 6.05 (− 12.11 to 0.00)	0.050
Education (Ref. high school graduate)				
Elementary school	− 0.63 (− 1.11 to − 0.15)	0.01	− 2.24 (− 7.21 to 2.72)	0.376
Middle school	0.021 (− 0.10 to 0.14)	0.732	2.27 (0.494 to 0.05)	0.369
College graduates	− 0.007 (− 0.089 to 0.074)	0.857	− 1.08 (− 1.93 to − 0.24)	0.012
Duration of MS disease, years	− 0.00 (− 0.00 to 0.001)	0.251	− 0.02 (− 0.07 to 0.01)	0.206
MS clinical descriptors (Ref. PPMS)				
RRMS	0.19 (0.00 to 0.38)	0.024	2.27 (0.49 to 4.05)	0.012
SPMS	0.14 (0.02 to 0.36)	0.171	1.83 (− 0.32 to 3.99)	0.097
“I don’t know”	0.19 (− 0.06 to 0.35)	0.048	1.90 (− 0.06 to 3.87)	0.058
Depression	0.04 (− 0.04 to 0.11)	0.355	− 0.68 (− 1.50 to 0.12)	0.098
Subjective memory and attention deficit	0.03 (− 0.05 to 0.11)	0.431	0.19 (− 0.63 to 1.03)	0.641

The continuous outcomes (average satisfaction and average CARE scale scores for ChatGPT responses) were assessed using mixed linear models with robust standard errors accounting for heteroskedasticity across patients. Covariates were age, sex, treatment duration, clinical descriptors, presence of self-reported cognitive deficit, presence of depressive symptoms and educational attainment

CI Confidence interval, IRR incidence rate ratio, MS multiple sclerosis, OR odds ratio, PPMS primary progressive multiple sclerosis, RRMS relapsing–remitting multiple sclerosis, SPMS secondary-progressive multiple sclerosis; *y* years; the use of bold formatting within the table was employed to highlight significant values

Table 5 Mean readability indices for each question

Readability index		Question 1	Question 2	Question 3	Question 4	Minimum–maximum—ChatGPT	Minimum–maximum—neurologist
Gulpease index	ChatGPT	52	49	48	47	47–52	40–44
	Neurologist	40	44	42	40		
ReadIt—base	ChatGPT	32.3%	55.5%	56.7%	65.9%	32.3–65.9	88.9–98.7
	Neurologist	96.6%	88.9%	98.7%	98.0%		
ReadIt—lexical	ChatGPT	74.4%	24.9%	96.0%	70.4%	24.9–96	80.2–100
	Neurologist	100%	89.3%	93.3%	80.2%		
ReadIt—syntax	ChatGPT	21.7%	36.3%	95.7%	92.1%	21.7–95.7	15.1–99.6
	Neurologist	78.3%	15.1%	99.6%	95.5%		
ReadIt—global	ChatGPT	2.0%	0.8%	71.9%	50.8%	0.8–71.9	90.7–100
	Neurologist	100%	90.7%	100%	100%		

The Gulpease index evaluates the overall readability of a text on a scale of 0 to 100, with higher scores indicating better ease of reading. A Gulpease index between 40 and 60 denotes a text that is poorly understandable to individuals with an elementary or middle school license, but easily comprehensible to high school graduates and those with higher education. Conversely, the Read-IT scores assess the different layers of a text, namely the structural, syntactic, and lexical plan, thereby rendering four distinct indices, Read-IT Base, Read-IT Lexical, Read-IT Syntax and Read-IT Global

these findings might represent a potential direction for future research.

Another relevant finding was that PwMS with higher levels of education showed lower satisfaction towards

the responses developed by ChatGPT, this suggest that educational level could be a key factor in health communication. Several studies suggest that having a higher degree of education is associated with a better predisposition

towards AIs [33], and in general, towards online information seeking [34, 35]. Although it may seem a contradiction, the predisposition toward digital technology doesn't necessarily align with the perception of communicative messages within the doctor-patient relationship. Moreover, people with higher levels of education may have developed greater critical skills, enabling them to better appreciate the appropriateness and precision of the language employed by neurologists [30].

In addition, in our study, the responses provided by ChatGPT have shown adequate overall readability, using simple words and clear language. This could be one of the reasons that make them potentially more favourable for individuals with lower levels of education and for younger people.

When examining individually the four questions, we observed varying results without a consistent pattern; however, no contradictory findings emerged.

In questions N° 2 and N° 4, PwMS who reported subjective memory and attention deficits were more likely to select the AI response. Still, in question N° 1 and N° 4, a higher likelihood to prefer the ChatGPT response emerged for subjects with PPMS. This result could be attributed to the distinct cognitive profile showed in PPMS [36], which is characterized by moderate-to-severe impairments.

In addition, for question N° 1, there was a decrease in the probability of preferring the response generated by ChatGPT with the increasing age of the participants. This result is in line with some previous findings [34, 35], and further highlights the digital divide between "Digital Natives," those who were born into digital age, and "Digital Immigrants", those who experienced the transition to digital [37, 38].

Finally, in question N° 1, PwMS with depressive symptoms showed a lower propensity to select the response generated by ChatGPT. This result could suggest that ChatGPT employs a type of language and vocabulary that is perhaps less well-received by individuals with depressive symptoms than the one used by neurologists, leading them to prefer the latter. Further research with a combination of quantitative and qualitative tools is needed to deepen this insight.

Our results point to the need to tailor digital resources, including ChatGPT, to render them more accessible and user-friendly for all users, considering their needs and skills. This could help bridge the present gap and enable digital resources to be effective for a wide range of users, regardless of their age, education, and medical and digital background. Indeed, a significant issue is that ChatGPT lacks knowledge of the details and background data of PwMS medical record, as it could lead to inaccurate or incorrect advice.

Furthermore, as our findings showed greater empathy of ChatGPT towards PwMS queries, the concern is that they may over-rely on AI rather than consulting their neurologist. Given the potential risks associated with the unsupervised use of AIs, physicians are encouraged to adapt to the progressive digitization of patients. This includes not only

providing proper guidance on the use of digital resources for health information seeking, but also addressing the potential drawbacks associated with relying solely on AI-generated results. Moreover, future research should address the possible integration of chatbots into a mixed-mode framework (AI-assisted approach). Integrating generative AI software into the neurologist's clinical practice could facilitate efficient communication while maintaining the human element.

Our study has several strengths and limitations. The objective was to explore the potential of AIs, such as ChatGPT, in the interaction with PwMS by engaging patients themselves in the evaluation [9, 10]. To this aim, we have replicated a patient-neurologist or patient-AIs interaction scenario. ChatGPT open access version 3.5 was preferred over newer and more advanced version 4.0 (pay per use). In fact, nonmajor users of online services will likely seek information free of charge [39, 40].

The main limitation of our study is the relatively small number of questions, however we deliberately selected a representative sample to enhance compliance and avoid discouraging PwMS from responding [41], as length can be a factor affecting response rates. Moreover, while our study adopted MS as a model of chronic disease for its core features, findings are not generalizable to other chronic conditions. We employed subjective self-report of cognitive impairment, and more studies adopting objective measures of cognitive screening will be needed to confirm our findings. Because a high level of education does not always correspond to a high digital literacy, it will be essential to assess digital literacy in future studies. Moreover, we preferred the use of institutional e-mail and instant messaging platform for recruitment, over in-person participation, given the predominantly digital nature of the research; still, the average response rate was in line with previous research [42]. We acknowledged that using stratified models may entail the risk of incurring Type I errors. However, stratification has been applied within homogeneous subgroups and on outcomes that are contextually differentiated due to the nature of different questions. This targeted approach allows for more accurate associations between predictors and outcomes to be discovered, thereby minimizing the risk of Type I error. Given the nonuniform distribution of patients across education classes within the categorical variable and the direct relationship between statistical significance and sample size, it's conceivable that this influenced the outcomes for the lower education classes (elementary and high school). However, the already observed statistically significant difference among the higher education classes implies that similar results could be achieved by standardizing the distribution of patients within the "education" categorical variable. This highlights the potential significance of education level in health communication, a crucial aspect warranting

further exploration in scientific research. Finally, we tested ChatGPT in the perceived quality of communication, and did not assess its ability to make actual clinical decisions, which would require further specific studies.

Future development should include to (a) guide the development of AI-based systems that better meet the needs and preferences of patients, taking into consideration their cultural, social and digital backgrounds; (b) educate healthcare professionals and patients on AI's role and capabilities for an informed and responsible use; (c) implement research methodology in the field of remote healthcare communication.

Conclusion

Our study showed that PwMS find ChatGPT's responses more empathetic than those of neurologists. However, it seems that ChatGPT is not completely ready to fully meet the needs of some categories of patients (i.e., high educational attainment). While physicians should prepare themselves to interact with increasingly digitized patients, ChatGPT's algorithms needs to focus on tailoring its responses to individual characteristics. Therefore, we believe that AI tools may pave the way for new perspectives in chronic disease management, serving as valuable support elements rather than alternatives.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00415-024-12328-x>.

Author contributions Conception and design of the study: EM, LL, SB and MM. Material Preparation: EM, LL, MM and SB. The first draft of the manuscript was written by EM, LL, SB and MM and all authors commented on previous versions of the manuscript. All authors read and approved the final version to be published. All authors agree to be accountable for all aspects of the work. SB is the guarantor.

Funding Open access funding provided by Università degli Studi della Campania Luigi Vanvitelli within the CRUI-CARE Agreement. The author(s) received no financial support for the research, authorship, and/or publication of this article.

Data availability The data that support the findings of this study are available from the corresponding author, SB, upon reasonable request.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Ethical approval The Ethical Committee of the University of Campania "Luigi Vanvitelli" approved the study (protocol number 0014460/i).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source,

provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y (2017) Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2(4):230–243. <https://doi.org/10.1136/svn-2017-000101>
- Ortiz M, Mallen V, Boquete L, Sánchez-Morla EM, Cordón B, Vilades E, Dongil-Moreno FJ, Miguel-Jiménez JM, Garcia-Martin E (2023) Diagnosis of multiple sclerosis using optical coherence tomography supported by artificial intelligence. *Mult Scler Relat Disord* 74:104725. <https://doi.org/10.1016/j.msard.2023.104725>
- Afzal HMR, Luo S, Ramadan S, Lechner-Scott J (2022) The emerging role of artificial intelligence in multiple sclerosis imaging. *Mult Scler* 28(6):849–858. <https://doi.org/10.1177/1352458520966298>
- Zivadinov R, Bergsland N, Jakimovski D, Weinstock-Guttman B, Benedict RHB, Riolo J, Silva D, Dwyer MG (2022) DeepGRAI registry study group. Thalamic atrophy measured by artificial intelligence in a multicentre clinical routine real-word study is associated with disability progression. *J Neurol Neurosurg Psychiatry* jnnp. <https://doi.org/10.1136/jnnp-2022-329333>
- ChatGPT. <https://openai.com/blog/chatgpt>. Accessed Dec 2023
- Shah NH, Entwistle D, Pfeffer MA (2023) Creation and adoption of large language models in medicine. *JAMA* 330(9):866–869. <https://doi.org/10.1001/jama.2023.14217>
- ChatGPT Statistics 2023: Trends and the Future Perspectives. <https://blog.gitnux.com/chat-gpt-statistics/>. Accessed Nov 2023
- Goodman RS, Patrinely JR, Stone CA Jr et al (2023) Accuracy and reliability of chatbot responses to physician questions. *JAMA Netw Open* 6(10):e2336483. <https://doi.org/10.1001/jamanetwopen.2023.36483>
- Ali SR, Dobbs TD, Hutchings HA, Whitaker IS (2023) Using ChatGPT to write patient clinic letters. *Lancet Digit Health* 5(4):e179–e181. [https://doi.org/10.1016/S2589-7500\(23\)00048-1](https://doi.org/10.1016/S2589-7500(23)00048-1)
- Inojosa H, Gilbert S, Kather JN, Proschmann U, Akgün K, Ziemssen T (2023) Can ChatGPT explain it? Use of artificial intelligence in multiple sclerosis communication. *Neurol Res Pract* 5(1):48. <https://doi.org/10.1186/s42466-023-00270-8>
- Madrigal L, Escoffery C (2019) Electronic health behaviors among us adults with chronic disease: cross-sectional survey. *J Med Internet Res* 21(3):e11240. <https://doi.org/10.2196/11240>
- Charness N, Boot WR (2023) A grand challenge for psychology: reducing the age-related digital divide. *Curr Dir Psychol Sci* 31(2):187–193. <https://doi.org/10.1177/09637214211068144>
- Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M (2007) STROBE initiative. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Epidemiology* 18(6):805–835. <https://doi.org/10.1097/EDE.0b013e3181577511>
- Digital Technology, Web and Social Media Study Group. https://www.neuro.it/web/eventi/NEURO/gruppi.cfm?p=DIGITAL_WEB_SOCIAL. Accessed Dec 2023

15. Research Randomizer. <https://www.randomizer.org>. Accessed July 2023
16. Kroenke K, Spitzer RL, Williams JB (2003) The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 41(11):1284–1292. <https://doi.org/10.1097/01.MLR.0000093487.78664.3C>
17. Beswick E, Quigley S, Macdonald P, Patrick S, Colville S, Chandran S, Connick P (2022) The Patient Health Questionnaire (PHQ-9) as a tool to screen for depression in people with multiple sclerosis: a cross-sectional validation study. *BMC Psychol* 10(1):281. <https://doi.org/10.1186/s40359-022-00949-8>
18. Patten SB, Burton JM, Fiest KM, Wiebe S, Bulloch AG, Koch M, Dobson KS, Metz LM, Maxwell CJ, Jetté N (2015) Validity of four screening scales for major depression in MS. *Mult Scler* 21(8):1064–1071. <https://doi.org/10.1177/1352458514559297>
19. Mercer SW, Maxwell M, Heaney D, Watt GC (2004) The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. *Fam Pract* 21(6):699–705. <https://doi.org/10.1093/fampra/cmh621>
20. Wang Y, Wang P, Wu Q, Wang Y, Lin B, Long J, Qing X, Wang P (2023) Doctors' and patients' perceptions of impacts of doctors' communication and empathy skills on doctor-patient relationships during COVID-19. *J Gen Intern Med* 38(2):428–433. <https://doi.org/10.1007/s11606-022-07784-y>
21. Martikainen S, Falcon M, Wikström V, Peltola S, Saarikivi K (2022) Perceptions of doctors' empathy and patients' subjective health status at an online clinic: development of an empathic Anamnesis Questionnaire. *Psychosom Med* 84(4):513–521. <https://doi.org/10.1097/PSY.0000000000001055>
22. Lucisano P, Piemontese ME (1988) Gulpease: a formula to predict readability of texts written in Italian Language. *Scuola Città* 3:110–124
23. Dell'orletta F, Montemagni S, Venturi G (2011) READ-IT: assessing readability of italian texts with a view to text simplification, in Proceedings of the Workshop on Speech and Language Processing for Assistive Technologies. Edinburgh, pp 73–83
24. Zhao YC, Zhao M, Song S (2022) Online health information seeking among patients with chronic conditions: integrating the health belief model and social support theory. *J Med Internet Res* 24(11):e42447. <https://doi.org/10.2196/42447>
25. Brigo F, Lattanzi S, Bragazzi N, Nardone R, Moccia M, Lavorgna L (2018) Why do people search wikipedia for information on multiple sclerosis? *Mult Scler Relat Disord* 20:210–214. <https://doi.org/10.1016/j.msard.2018.02.001>
26. Ayoub NF, Lee YJ, Grimm D, Balakrishnan K (2023) Comparison between ChatGPT and google search as sources of postoperative patient instructions. *JAMA Otolaryngol Head Neck Surg* 149(6):556–558. <https://doi.org/10.1001/jamaoto.2023.0704>
27. Lavorgna L, De Stefano M, Sparaco M, Moccia M, Abbadessa G, Montella P, Buonanno D, Esposito S, Clerico M, Cenci C, Trojsi F, Lanzillo R, Rosa L, Morra VB, Ippolito D, Maniscalco G, Biseco A, Tedeschi G, Bonavita S (2018) Fake news, influencers and health-related professional participation on the web: a pilot study on a social-network of people with multiple sclerosis. *Mult Scler Relat Disord* 25:175–178. <https://doi.org/10.1016/j.msard.2018.07.046>
28. Herzer KR, Pronovost PJ (2021) Ensuring quality in the era of virtual care. *JAMA* 325(5):429–430. <https://doi.org/10.1016/j.msard.2018.07.046>
29. Mello MM, Guha N (2023) ChatGPT and physicians' malpractice risk. *JAMA Health Forum* 4(5):e231938. <https://doi.org/10.1001/jamahealthforum.2023.1938>
30. van Laar E, van Deursen AJAM, van Dijk JAGM, de Haan J (2020) Determinants of 21st-century skills and 21st-century digital skills for workers: a systematic literature review. *SAGE Open*. <https://doi.org/10.1177/2158244019900176>
31. National Research Council (2000) How people learn: brain, mind, experience, and school expanded edition. The National Academies Press, Washington, DC
32. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, Faix DJ, Goodman AM, Longhurst CA, Hogarth M, Smith DM (2023) Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med* 183(6):589–596. <https://doi.org/10.1001/jamainternmed.2023.1838>
33. Kaya F, Aydin F, Schepman A et al (2022) The roles of personality traits, AI anxiety, and demographic factors in attitudes toward artificial intelligence. *Int J Hum-Comput Int*. <https://doi.org/10.1080/10447318.2022.2151730>
34. Jia X, Pang Y, Liu LS (2021) Online health information seeking behavior: a systematic review. *Healthcare (Basel)* 9(12):1740. <https://doi.org/10.3390/healthcare9121740>
35. D'Andrea A, Grifoni P, Ferri F (2023) Online health information seeking: an italian case study for analyzing citizens' behavior and perception. *Int J Environ Res Public Health* 20(2):1076. <https://doi.org/10.3390/ijerph20021076>
36. De Meo E, Portaccio E, Giorgio A et al (2021) Identifying the distinct cognitive phenotypes in multiple sclerosis. *JAMA Neurol* 78(4):414–425. <https://doi.org/10.1001/jamaneurol.2020.4920>
37. Hatcher-Martin JM, Busis NA, Cohen BH, Wolf RA, Jones EC, Anderson ER, Fritz JV, Shook SJ, Bove RM (2021) American academy of neurology telehealth position statement. *Neurology* 97(7):334–339. <https://doi.org/10.1212/WNL.00000000000012185>
38. Haluza D, Naszay M, Stockinger A, Jungwirth D (2017) Digital natives versus digital immigrants: influence of online health information seeking on the doctor-patient relationship. *Health Commun* 32(11):1342–1349. <https://doi.org/10.1080/10410236.2016.1220044>
39. Chua V, Koh JH, Koh CHG, Tyagi S (2022) The willingness to pay for telemedicine among patients with chronic diseases: systematic review. *J Med Internet Res* 24(4):e33372. <https://doi.org/10.2196/33372>
40. Xie Z, Chen J, Or CK (2022) Consumers' willingness to pay for ehealth and its influencing factors: systematic review and meta-analysis. *J Med Internet Res* 24(9):e25959. <https://doi.org/10.2196/25959>
41. Fan W, Yan Z (2010) Factors affecting response rates of the web survey: a systematic review. *Comput Hum Behav* 26:132–139. <https://doi.org/10.1016/j.chb.2009.10.01>
42. Wu MJ, Zhao K, Fils-Aime F (2022) Response rates of online surveys in published research: a meta-analysis. *Comput Hum Behav*. <https://doi.org/10.1016/j.chbr.2022.100206>

Authors and Affiliations

Elisabetta Maida¹ · Marcello Moccia^{2,3} · Raffaele Palladino^{4,5} · Giovanna Borriello⁶ · Giuseppina Affinito⁴ · Marinella Clerico⁷ · Anna Maria Repice⁸ · Alessia Di Sapio⁹ · Rosa Iodice¹⁰ · Antonio Luca Spiezia³ · Maddalena Sparaco¹ · Giuseppina Miele¹ · Floriana Bile¹ · Cristiano Scandurra¹⁰ · Diana Ferraro^{11,12} · Maria Laura Stromillo¹³ · Renato Docimo¹⁴ · Antonio De Martino¹⁵ · Luca Mancinelli¹⁶ · Gianmarco Abbadessa^{1,17} · Krzysztof Smolik¹² · Lorenzo Lorusso¹⁸ · Maurizio Leone¹⁹ · Elisa Leveraro^{20,21} · Francesca Lauro¹⁰ · Francesca Trojsi¹ · Lidia Mislin Streito⁷ · Francesca Gabriele²² · Fabiana Marinelli²³ · Antonio Ianniello⁶ · Federica De Santis²⁴ · Matteo Foschi^{22,25} · Nicola De Stefano¹³ · Vincenzo Brescia Morra¹⁰ · Alvino Bisecco¹⁴ · Giancarlo Coghe²⁶ · Eleonora Cocco²⁶ · Michele Romoli¹⁶ · Francesco Corea²⁷ · Letizia Leocani^{28,29} · Jessica Frau²⁶ · Simona Sacco²² · Matilde Inglese^{20,21} · Antonio Carotenuto¹⁰ · Roberta Lanzillo¹⁰ · Alessandro Padovani^{30,31} · Maria Triassi⁴ · Simona Bonavita¹  · Luigi Lavorgna¹ on behalf of Digital Technologies, Web, Social Media Study Group of the Italian Society of Neurology (SIN)

✉ Simona Bonavita
simona.bonavita@unicampania.it

¹ Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Via Pansini 5, 80131 Naples, Italy

² Department of Molecular Medicine and Medical Biotechnology, Federico II University of Naples, Naples, Italy

³ Multiple Sclerosis Unit, Policlinico Federico II University Hospital, Naples, Italy

⁴ Department of Public Health, University “Federico II” of Naples, Naples, Italy

⁵ Department of Primary Care and Public Health, Imperial College of London, London, UK

⁶ Department of Human Neuroscience, Sapienza University of Rome, Rome, Italy

⁷ Dipartimento di Scienze Cliniche e Biologiche, Università Di Torino, Turin, Italy

⁸ Department of Neurology 2 and Tuscan Region Multiple Sclerosis Referral Centre, Careggi University Hospital, Florence, Italy

⁹ Department of Neurology and Multiple Sclerosis Regional Referral Centre, AOU San Luigi Gonzaga, Orbassano, Turin, Italy

¹⁰ Department of Neurosciences, Reproductive Sciences and Odontostomatology, University of Naples Federico II, Naples, Italy

¹¹ Department of Neuroscience, Azienda Ospedaliero-Universitaria di Modena, Modena, Emilia-Romagna, Italy

¹² Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy

¹³ Department of Medicine, Surgery and Neuroscience, University of Siena, Siena, Italy

¹⁴ Multiple Sclerosis Center, Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy

¹⁵ Institute of Neurology, University Magna Graecia of Catanzaro, Catanzaro, Italy

¹⁶ Neurology Unit, Bufalini Hospital, Local Health Agency of Romagna, Cesena, Italy

¹⁷ Department of Brain Sciences, Imperial College London, London W120BZ, UK

¹⁸ Neurology Unit-Neuroscience Department A.S.S.T.Lecco, Merate Hospital, 23807 Merate, Italy

¹⁹ Neurology Unit, Fondazione IRCCS Casa Sollievo della Sofferenza, 71013 San Giovanni Rotondo, Italy

²⁰ Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DiNOGMI), University of Genoa, Genoa, Italy

²¹ Department of Neurology, IRCCS Ospedale Policlinico San Martino, Genoa, Italy

²² Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, L'Aquila, Italy

²³ Neurology Unit, Multiple Sclerosis Center, Fabrizio Spaziani Hospital, Frosinone, Italy

²⁴ Department of Neurology and Stroke Unit of Avezzano-Sulmona, ASL 1 Avezzano-Sulmona-L'Aquila, L'Aquila, Italy

²⁵ Department of Neuroscience, Multiple Sclerosis Center, S. Maria delle Croci Hospital, AUSL Romagna, Ravenna, Italy

²⁶ Department of Medical Sciences and Public Health, Multiple Sclerosis Center, Binaghi Hospital, ASL Cagliari, University of Cagliari, Cagliari, Italy

²⁷ Dipartimento di Neurologia, Ospedale di Foligno, Azienda USL Umbria 2, Terni, Italy

²⁸ Vita-Salute San Raffaele University, Milan, Italy

²⁹ Experimental Neurophysiology Unit, Institute of Experimental Neurology-INSPE, IRCCS Scientific Institute San Raffaele, Milan, Italy

³⁰ Unit of Neurology, Azienda Socio-Sanitaria Territoriale Spedali Civili, Brescia, Italy

³¹ Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy