

Article

Evaluation of ChatGPT as a Counselling Tool for Italian-Speaking MASLD Patients: Assessment of Accuracy, Completeness and Comprehensibility

Nicola Pugliese ^{1,2} , Davide Polverini ^{1,2}, Rosa Lombardi ^{3,4} , Grazia Pennisi ⁵, Federico Ravaoli ^{6,7} , Angelo Armandi ^{8,9} , Elena Buzzetti ^{10,11} , Andrea Dalbeni ^{12,13} , Antonio Liguori ¹⁴ , Alessandro Mantovani ¹⁵ , Rosanna Villani ¹⁶ , Ivan Gardini ¹⁷, Cesare Hassan ^{1,18}, Luca Valenti ^{4,19}, Luca Miele ^{14,20} , Salvatore Petta ⁵, Giada Sebastiani ²¹ , Alessio Aghemo ^{1,2,*} and NAFLD Expert Chatbot Working Group [†]

- ¹ Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, 20072 Milan, Italy; nicola.pugliese@humanitas.it (N.P.); davide.polverini@humanitas.it (D.P.); cesare.hassan@hunimed.eu (C.H.)
- ² Division of Internal Medicine and Hepatology, Department of Gastroenterology, IRCCS Humanitas Research Hospital, Rozzano, 20089 Milan, Italy
- ³ Unit of Internal Medicine and Metabolic Disease, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico of Milan, 20122 Milan, Italy; rosa.lombardi@unimi.it
- ⁴ Department of Pathophysiology and Transplantation, Università degli Studi di Milano, 20122 Milan, Italy; luca.valenti@unimi.it
- ⁵ Section of Gastroenterology and Hepatology, PROMISE, University of Palermo, 90127 Palermo, Italy; grazia.pennisi@unipa.it (G.P.); salvatore.petta@unipa.it (S.P.)
- ⁶ Department of Medical and Surgical Sciences (DIMEC), University of Bologna, 40138 Bologna, Italy; f.ravaoli@unibo.it
- ⁷ Division of Internal Medicine, Hepatobiliary and Immunoallergic Diseases, IRCCS Azienda Ospedaliero Universitaria di Bologna, 40138 Bologna, Italy
- ⁸ Division of Gastroenterology and Hepatology, Department of Medical Sciences, University of Turin, Corso Dogliotti 14, 10126 Turin, Italy; angelo.armandi@unito.it
- ⁹ Metabolic Liver Disease Research Program, I. Department of Internal Medicine, University Medical Center of Mainz, 55131 Mainz, Germany
- ¹⁰ Internal Medicine and Centre for Hemochromatosis and Hereditary Liver Diseases, ERN-EuroBloodNet Center for Iron Disorders, Azienda Ospedaliero-Universitaria di Modena-Policlinico, 41125 Modena, Italy; elena.buzzetti@unimore.it
- ¹¹ Department of Medical and Surgical Sciences, Università degli Studi di Modena e Reggio Emilia, 41125 Modena, Italy
- ¹² Division of General Medicine C, Department of Medicine, University and Azienda Ospedaliera Universitaria Integrata of Verona, University of Verona, 37134 Verona, Italy; andrea.dalbeni@univr.it
- ¹³ Liver Unit, Department of Medicine, University and Azienda Ospedaliera Universitaria Integrata of Verona, University of Verona, 37134 Verona, Italy
- ¹⁴ DiSMeC—Department of Scienze Mediche e Chirurgiche, Fondazione Policlinico Gemelli IRCCS, 00168 Rome, Italy; antonio.liguori@guest.policlinicogemelli.it (A.L.); luca.miele@policlinicogemelli.it (L.M.)
- ¹⁵ Section of Endocrinology, Diabetes and Metabolism, Department of Medicine, University and Azienda Ospedaliera Universitaria Integrata of Verona, Piazzale Stefani, 37126 Verona, Italy; alessandro.mantovani@univr.it
- ¹⁶ C.U.R.E. (University Center for Liver Disease Research and Treatment), Liver Unit, Department of Medical and Surgical Sciences, University of Foggia, 71122 Foggia, Italy; rosanna.villani@unifg.it
- ¹⁷ EpaC Onlus, Italian Liver Patient Association, 10141 Turin, Italy; gardini@epac.it
- ¹⁸ Division of Gastroenterology and Digestive Endoscopy, Humanitas Research Hospital, IRCCS, Rozzano, 20089 Milan, Italy
- ¹⁹ Precision Medicine Lab, Biological Resource Center, Department of Transfusion Medicine, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, 20122 Milan, Italy
- ²⁰ Department of Medicina e Chirurgia Traslazionale, Università Cattolica Del Sacro Cuore, 00168 Rome, Italy
- ²¹ Division of Gastroenterology and Hepatology, McGill University Health Centre, Montreal, QC H4A 3J1, Canada; giada.sebastiani@mcgill.ca
- * Correspondence: alessio.aghemo@hunimed.eu; Tel.: +39-02-8224-7006; Fax: +39-02-8224-2298
- † NAFLD Expert Chatbot Working Group (Collaborators names are provided in the Acknowledgments).



Citation: Pugliese, N.; Polverini, D.; Lombardi, R.; Pennisi, G.; Ravaoli, F.; Armandi, A.; Buzzetti, E.; Dalbeni, A.; Liguori, A.; Mantovani, A.; et al. Evaluation of ChatGPT as a Counselling Tool for Italian-Speaking MASLD Patients: Assessment of Accuracy, Completeness and Comprehensibility. *J. Pers. Med.* **2024**, *14*, 568. <https://doi.org/10.3390/jpm14060568>

Academic Editor: Abdul Rahman Asif

Received: 9 May 2024

Revised: 19 May 2024

Accepted: 21 May 2024

Published: 26 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: Artificial intelligence (AI)-based chatbots have shown promise in providing counseling to patients with metabolic dysfunction-associated steatotic liver disease (MASLD). While ChatGPT3.5 has demonstrated the ability to comprehensively answer MASLD-related questions in English, its accuracy remains suboptimal. Whether language influences these results is unclear. This study aims to assess ChatGPT's performance as a counseling tool for Italian MASLD patients. Methods: Thirteen Italian experts rated the accuracy, completeness and comprehensibility of ChatGPT3.5 in answering 15 MASLD-related questions in Italian using a six-point accuracy, three-point completeness and three-point comprehensibility Likert's scale. Results: Mean scores for accuracy, completeness and comprehensibility were 4.57 ± 0.42 , 2.14 ± 0.31 and 2.91 ± 0.07 , respectively. The physical activity domain achieved the highest mean scores for accuracy and completeness, whereas the specialist referral domain achieved the lowest. Overall, Fleiss's coefficient of concordance for accuracy, completeness and comprehensibility across all 15 questions was 0.016, 0.075 and -0.010 , respectively. Age and academic role of the evaluators did not influence the scores. The results were not significantly different from our previous study focusing on English. Conclusion: Language does not appear to affect ChatGPT's ability to provide comprehensible and complete counseling to MASLD patients, but accuracy remains suboptimal in certain domains.

Keywords: MASLD; artificial intelligence; counselling; diet; physical activity; steatosis; chatbot

1. Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD) represents a significant global public health concern, anticipated to emerge as the foremost indication for liver transplantation in the coming years [1–4]. Its prevalence ranges between 20% and 40% in the general population, with a higher incidence observed among individuals struggling with obesity, type 2 diabetes (T2D), and/or dyslipidemia [1–3]. The spectrum of MASLD encompasses diverse pathological presentations ranging from simple hepatic steatosis (SLD) to metabolic dysfunction-associated steatohepatitis (MASH), characterized by inflammation, hepatocellular ballooning and fibrosis, predisposing to severe hepatic sequelae such as cirrhosis, end-stage liver disease and hepatocellular carcinoma (HCC) [4,5]. The increasing prevalence of T2D and obesity worldwide is expected to lead to a parallel increase in MASLD and MASH cases, further exacerbating the burden on healthcare systems worldwide [1].

Despite the recent Food and Drug Administration (FDA) approval of Resmetirom as the first prescribable drug for MASH and ongoing phase 3 trials investigating promising agents, lifestyle modification, particularly dietary changes and increased physical activity geared towards achieving weight reduction, remain pivotal in the management of MASLD [6–11]. Research has shown that weight loss, achievable through lifestyle modifications, significantly improves liver histology, diminishing inflammation, hepatocellular ballooning and fibrosis in patients with MASH [12,13].

Innovative strategies such as personalized counselling, behavioral interventions, and educational initiatives are essential to facilitate the adoption and perpetuation of lifestyle modifications among patients with MASLD. Given the inherent challenges in implementing lifestyle changes, comprehensive counselling assumes paramount importance to engendering patient motivation. Armed with comprehensive information about their condition, patients are empowered to cultivate an awareness that catalyzes habitual changes in diet and physical activity [14–17].

The growing interest in using artificial intelligence (AI) technologies in healthcare settings to deliver personalized support and patient education underscores the potential utility of AI-based chatbots. These conversational agents, which use natural language processing, have demonstrated efficacy in various healthcare settings, including mental health, chronic disease management and medication adherence [18–20]. By providing patients with round-the-clock support and information, AI-based chatbots complement existing

healthcare resources, addressing patient queries and concerns regarding their condition. The OpenAI Foundation’s development of ChatGPT, an AI-based chatbot predicated on the third-generation Generative Pretrained Transformer-3.5 (GPT-3.5) architecture, has received widespread acclaim for its ability to handle various question-answer tasks [21,22].

A recent study conducted by our research group evaluated the potential of ChatGPT as a counselling tool for English-speaking patients with MASLD. While the initial findings were promising, they also highlighted discrepancies in the accuracy of ChatGPT-generated responses [23]. Presently, there is a paucity of research investigating the efficacy of chatbot-mediated counselling in non-English speaking MASLD patients, especially in the field of hepatology [24].

Against this backdrop, our study aims to assess the accuracy, completeness and comprehensibility of ChatGPT responses to basic questions posed in Italian by MASLD patients seeking insight into their condition. Specifically, we aim to determine whether language influences the quality of ChatGPT’s counselling by comparing its performance in answering MASLD-related questions in Italian to previous results in English. The insights gained from this study could have a significant impact on the use of AI-based tools in multilingual healthcare settings, providing tailored support to a wider range of patients and potentially improving the management and outcomes of MASLD worldwide.

2. Methods

2.1. Study Design

This study was conducted in October 2023 and was exempt from Institutional Review Board approval as it did not include patient-level data.

The study used English-language questions from a previous investigation designed to evaluate the efficacy of ChatGPT as a counselling tool for patients with MASLD [23]. These questions were carefully translated into Italian to ensure linguistic accuracy and cultural relevance. The translation process involved a linguistic expert, fluent in both English and Italian, who meticulously adapted the questions to maintain their original intent and clarity in the target language (Table 1).

Table 1. Questions posed to ChatGPT.

	Questions in English [11]	Questions in Italian
1	I have been diagnosed with steatotic liver disease. Should I be referred to a liver specialist? Should I have additional blood tests or radiological examinations?	Mi è stata diagnosticata la steatosi epatica. Devo rivolgermi ad un epatologo? Devo sottopormi ad esami del sangue o ad esami radiologici?
2	How much weight should I lose if I have steatotic liver disease?	Quanto peso devo perdere se ho la steatosi epatica?
3	Which diet should I follow if I have steatotic liver disease?	Quale dieta devo seguire se ho la steatosi epatica?
4	Which foods should I avoid if I have steatotic liver disease?	Quali cibi devo evitare se ho la steatosi epatica?
5	Can I drink alcohol if I have steatotic liver disease?	Posso bere alcolici se ho la steatosi epatica?
6	Which foods have a positive effect on steatotic liver disease?	Quali cibi hanno un effetto positivo sulla steatosi epatica?
7	Can I use herbal remedies if I have steatotic liver disease?	Posso usare rimedi erboristici per la mia steatosi epatica?
8	Can I drink coffee if I have steatotic liver disease?	Posso bere caffè se ho la steatosi epatica?
9	Can I smoke if I have steatotic liver disease?	Posso fumare se ho la steatosi epatica?
10	How much exercise should I do if I have steatotic liver disease?	Quanta attività fisica devo svolgere se ho la steatosi epatica?

Table 1. Cont.

	Questions in English [11]	Questions in Italian
11	Which type of exercise is better for steatotic liver disease?	Qual è la migliore tipologia di attività fisica per la steatosi epatica?
12	Should I do cardio or lift weights for steatotic liver disease?	Dovrei fare cardiofitness o sollevare pesi se ho la steatosi epatica?
13	Which drugs should I take if I have steatotic liver disease?	Quali farmaci devo assumere se ho la steatosi epatica?
14	How do I understand if steatotic liver disease is worsening?	In che modo mi rendo conto che la mia steatosi epatica è in peggioramento?
15	How do I understand if steatotic liver disease is improving?	In che modo mi rendo conto che la mia steatosi epatica è in miglioramento?

In the original study, questions were formulated based on guideline-based preventive measures and interventional therapies, categorizing them into the following three domains: specialist referral, dietary composition and physical activity [9,10]. The specific questions included in this task are as follows:

- (1) Specialist referral: questions 1, 14 and 15 addressed the need for specialist referral and the ability to detect improvement or worsening of MASLD.
- (2) Dietary composition: Questions from 2 to 9 and 13 focused on dietary composition, weight loss, medication, coffee consumption, and alcohol consumption.
- (3) Physical activity: questions 10, 11 and 12 asked about the type of physical activity recommended for MASLD patients.

The selection of the questions and domains was informed by established guidelines and best practices in MASLD management, ensuring that the assessment covered a comprehensive range of topics relevant to patient counselling.

ChatGPT, an AI-based natural language processing model developed by OpenAI, was used as the counselling interface for this study. ChatGPT, a variant of the GPT-3.5 Large Language Model (LLM), has undergone extensive pre-training on a large dataset of online text up to 2021 [13]. This pre-training enables ChatGPT to generate responses that are linguistically coherent and contextually appropriate. In addition, ChatGPT has been continuously refined through human feedback, increasing its accuracy and reliability in generating relevant responses.

On 1 October 2023, two Milan-based researchers, Nicola Pugliese and Alessio Aghemo, entered the translated questions into ChatGPT using the March 14 version of the 'New Chat' feature. Each question was entered individually as an autonomous prompt to ensure independent and unbiased responses from the AI [25].

2.2. Evaluation of the AI-Generated Questions and Answers Section

The results of ChatGPT's output were emailed to the thirteen Italian-speaking experts in MASLD (Supplementary Materials), including four senior experts who had previously evaluated ChatGPT's performance in English (Luca Valenti, Luca Miele, Salvatore Petta, Giada Sebastiani), and nine junior experts specifically focused on metabolic liver diseases (Rosa Lombardi, Grazia Pennisi, Federico Ravaioli, Angelo Armandi, Elena Buzzetti, Rosanna Villani, Andrea Dalbeni, Alessandro Mantovani, Antonio Liguori) [23]. The experts involved in the evaluation come from different universities located in different regions of Italy (Piemonte, Lombardia, Veneto, Lazio, Emilia-Romagna, Puglia, Sicilia), thus ensuring a geographically diversified representation. The senior experts selected have an H-index of at least 40, while the junior experts have an H-index of at least 10. These H-index values were determined using Scopus at the time of the evaluation. This targeted selection aimed to include experts with a wide range of experience and expertise in the field of metabolic liver diseases, thus ensuring a comprehensive and accurate assessment of the responses generated by ChatGPT.

Additionally, the responses generated by ChatGPT were shared with an Italian-speaking non-physician (Ivan Gardini) with expertise in patient advocacy for liver disease.

Each expert was instructed to independently and blindly evaluate each AI-generated response using Likert scales to assess accuracy, completeness and comprehensiveness, drawing on their expertise and experience with MASLD [26]. The non-physician was specifically asked to rate the responses only for comprehensiveness using the same Likert scale.

For accuracy, a six-point Likert scale was employed, ranging from 1 (completely incorrect) to 6 (correct), allowing experts to gauge the degree of correctness in the responses.

Completeness was assessed using a three-point Likert scale, with rating ranging from 1 (incomplete, addresses some aspects of the question, but significant parts are missing or incomplete) to 3 (comprehensive, addresses all aspects of the question and provides additional information or context beyond what was expected), indicating the extent to which the responses addressed all aspects of the questions posed.

Moreover, the comprehensibility of the text was evaluated using a separate three-point Likert scale, with ratings reflecting the ease of understating, ranging from 1 (difficult to understand) to 3 (easy to understand).

For detailed Likert scale criteria, Supplementary Table S1 provides further clarification.

2.3. Data Collection and Statistical Analysis

The experts were instructed to rate each answer blindly.

The investigation findings underwent analysis utilizing descriptive statistics and concordance measures. Descriptive statistics were employed to compute the mean and standard deviation of the ratings for each question. Evaluations from the two groups of specialists (senior and young experts) underwent a t-test, with statistical significance determined at a threshold of $p < 0.05$. In addition, the senior experts' ratings of ChatGPT-generated responses in Italian were compared with their ratings of English responses from a previous study by the NAFLD Expert Chatbot Working Group [23]. Concordance among experts was assessed using Fleiss's Kappa, a non-parametric measure determining the degree of agreement among raters, considering both the magnitude and direction of differences. Interpretation of Kappa values followed the Landis and Koch criteria as follows: poor (<0.00), slight ($0.00-0.20$), fair ($0.21-0.40$), moderate ($0.41-0.60$), substantial ($0.61-0.80$), and almost perfect agreement ($0.81-1.00$).

3. Results

3.1. Accuracy

The survey results revealed that ChatGPT's responses to 15 questions regarding MASLD were accurate, with an average score of 4.57 ± 0.42 on a Likert scale ranging from 1 to 6. Variations in scores were observed across different questions, with questions 9 ("Can I smoke if I have steatotic liver disease?") and 10 ("How much exercise should I do if I have steatotic liver disease?") achieving the highest scores, with mean scores of 5.07 ± 0.95 and 5.07 ± 0.86 , respectively. Conversely, the lowest-scoring questions were questions 14 ("How do I understand if steatotic liver disease is worsening?") and 15 ("How do I understand if steatotic liver disease is improving?"), averaging 3.84 ± 1.21 and 3.84 ± 1.14 , respectively.

In terms of domain-specific accuracy, the physical activity domain received the highest mean score of 4.82 ± 0.22 , while the specialist referral domain received the lowest score with 4.1 ± 0.44 (see Figure 1).

No significant differences were observed in the evaluation of responses between senior and junior experts, with mean scores of 4.23 ± 0.6 and 4.72 ± 0.42 , respectively ($p = 0.11$).

Similarly, there were no significant difference in the evaluation of Italian and English responses among senior experts, with mean scores of 4.23 ± 0.6 and 4.76 ± 0.48 , respectively ($p = 0.21$).

Agreement among experts was slight, with an overall Fleiss's coefficient of 0.016.

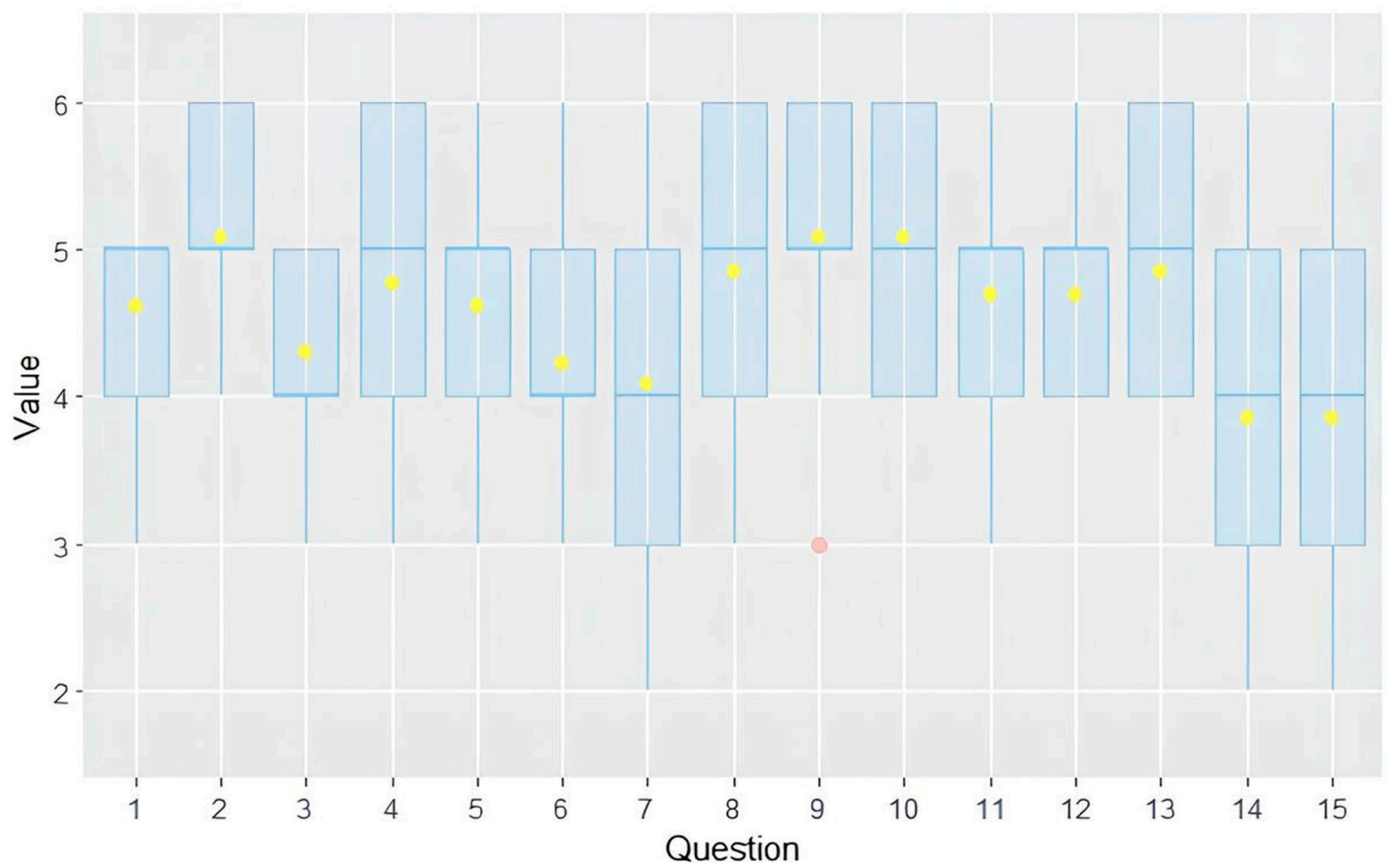


Figure 1. Box plot showing the distribution of accuracy scores. Graph shows interquartile range (box), median (horizontal line), mean (yellow dot) and outliers (whiskers).

3.2. Completeness

The mean score for completeness was 2.14 ± 0.31 on a Likert scale ranging from 1 to 3. Among the individual questions, question 13 (“Which drugs should I take if I have steatotic liver disease?”) received the highest score of 2.53 ± 0.51 , while question 15 (“How do I understand if steatotic liver disease is improving?”) received the lowest score of 1.46 ± 0.66 . Notably, only five responses (questions 1, 2, 11, 13, and 19) received unanimous evaluation scores of 2 and 3.

In terms of domain-specific completeness, the physical activity domain obtained the highest mean score of 2.35 ± 0.11 , while the specialist referral domain had the lowest score of 1.66 ± 0.29 (see Figure 2).

Similar to the evaluation of accuracy, no significant differences were found in how senior experts evaluated the responses compared to the junior ones, with mean scores of 1.98 ± 0.36 and 2.21 ± 0.35 , respectively ($p = 0.3$).

Additionally, there were no significant differences in the scoring of Italian and English responses among senior experts, with mean scores of 1.98 ± 0.36 and 1.79 ± 0.36 , respectively ($p = 0.48$).

Agreement among experts was slight, with an overall Fleiss’s coefficient of 0.075.

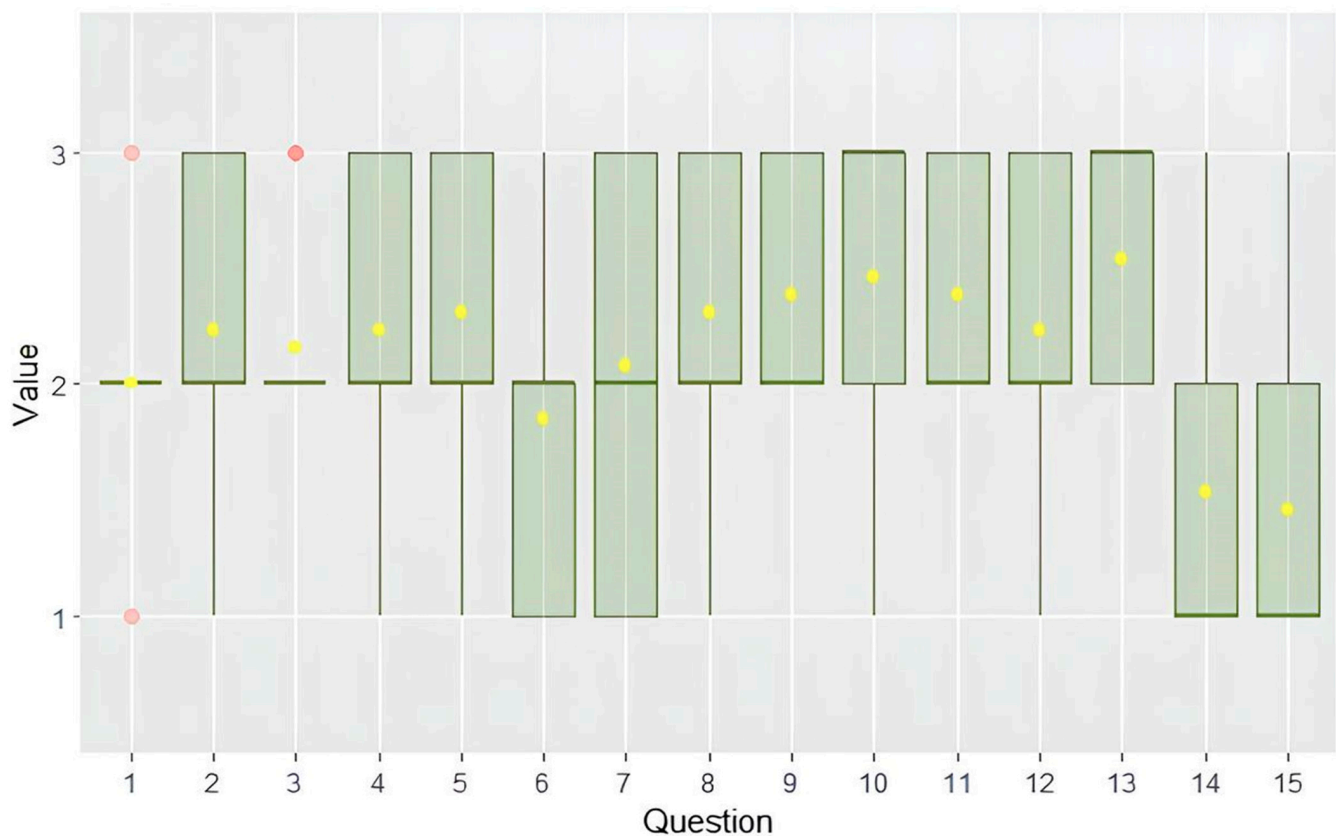


Figure 2. Box plot showing the distribution of completeness scores. Graph shows interquartile range (box), median (horizontal line), mean (yellow dot) and outliers (whiskers).

3.3. Comprehensibility

The survey results indicate that ChatGPT's responses to the 15 questions on MASLD received an average score of 2.91 ± 0.07 on the Likert scale ranging from 1 to 3. Notably, questions 1 ("I have been diagnosed with steatotic liver disease. Should I be referred to a liver specialist? Should I have additional blood tests or radiological examinations?"), 6 ("Which foods have a positive effect on steatotic liver disease?"), 7 ("Can I use herbal remedies if I have steatotic liver disease?"), and 8 ("Can I drink coffee if I have steatotic liver disease?") received unanimous scores of 3, suggesting a high level of comprehensibility in these responses. However, agreement among experts was poor, with an overall Fleiss's coefficient of -0.010 .

Comprehensibility was also assessed by a non-physician who rated all questions as 3, indicating a consistent perception of high comprehensibility across all responses (see Figure 3).

Similarly to the evaluation of accuracy and completeness, no significant differences were observed in the evaluation of responses by senior experts compared to junior experts, nor in the assessment of Italian and English responses by senior experts.

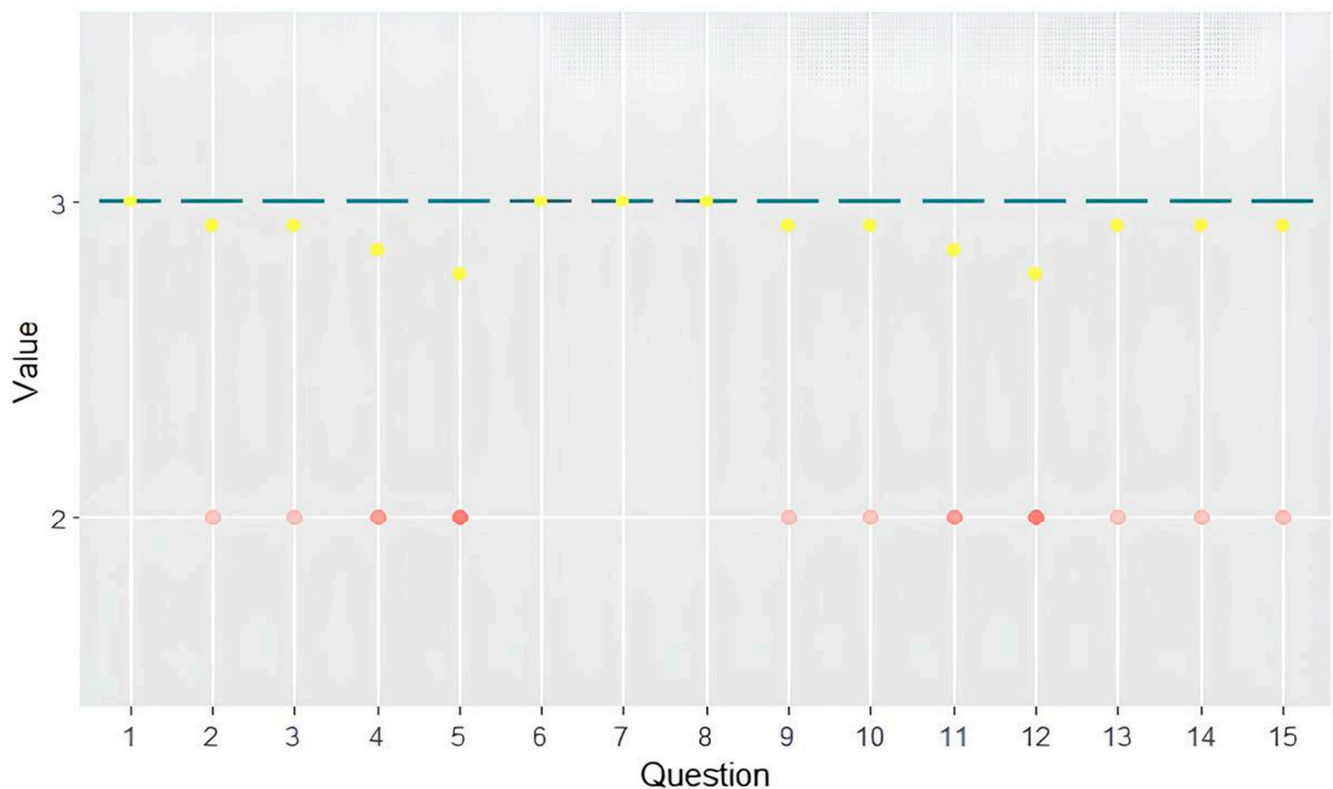


Figure 3. Box plot showing the distribution of comprehensibility scores. Graph shows interquartile range (box), median (horizontal line), mean (yellow dot) and outliers (whiskers).

4. Discussion and Conclusions

AI is anticipated to have a substantial impact on the medical field in the coming years and decades, affecting daily clinical practice [21,27,28]. Current research has primarily focused on developing AI-based tools to improve diagnostic accuracy and reduce inter- or intra-observer variability. However, the implementation of AI language model-based chatbots, like ChatGPT, has the potential to revolutionize the field by providing patients with immediate, free and on-demand counselling.

MASLD is a significant global public health issue, projected to become the primary indication for liver transplantation in the coming decades. This is due to the increasing prevalence of metabolic syndrome and obesity, key risk factors for MASLD [1–3]. Lifestyle changes have proven beneficial, making MASLD an ideal context to test the accuracy, completeness, and comprehensibility of ChatGPT-generated responses to patients' questions about managing their disease [9,10,12,13].

In a recent study, our group evaluated the performance of ChatGPT 3.5 in addressing MASLD-related questions. While the chatbot provided understandable responses, its accuracy could have been improved [11]. Notably, the study did not delve into the role and impact of the language used to interact with the chatbot.

The current study aimed to assess the accuracy, completeness, and comprehensibility of ChatGPT 3.5 in answering 15 pre-determined MASLD-related questions in Italian. Thirteen Italian-speaking MASLD experts participated, ensuring a diverse and geographically representative sample. The results showed that ChatGPT's ability to provide complete (mean score on a 3-point scale of 2.14) and understandable (mean score on a 3-point scale of 2.91) counseling to patients with MASLD remained consistent across different languages. However, the accuracy of ChatGPT responses remained suboptimal, with a mean score of 4.57 on a 6-point scale. Nonetheless, the responses were highly rated for comprehensibility, with five out of fifteen questions receiving the maximum grade from the experts. This was corroborated by the non-physician evaluator who also assigned the maximum grade to all

responses. The results pertaining to completeness were also promising, albeit with only light agreement among experts in evaluating ChatGPT responses.

Recent studies underscore accuracy as a critical area requiring improvement to prevent the dissemination of potentially harmful or erroneous information to patients [11,29]. For instance, in question 15, 'How do I understand if steatotic liver disease is improving?', six out of ten experts graded the answer ≤ 3 , indicating misinformation in the chatbot-generated response. Interestingly, five of these six experts also found the response incorrect but easy to comprehend. Hence, a notable limitation of ChatGPT as a counselling tool is its propensity to generate inaccurate or hallucinated information [30]. While our study demonstrates high completeness on average, it is essential to acknowledge ChatGPT's potential to yield incomplete or inaccurate responses, particularly in healthcare, where misinformation can have severe consequences. For example, the risk of 'hallucinations' raises concerns about the tool generating content that may not be grounded in evidence-based medical knowledge. Acting on health information without consulting a healthcare professional can have adverse health consequences.

Furthermore, the influence of socio-cultural factors on ChatGPT responses warrants consideration. The tool's ability to provide culturally sensitive advice depends on the patient's socio-cultural context, and the effectiveness of the advice may vary based on cultural nuances, health literacy levels, and patient preferences. Therefore, while ChatGPT represents a valuable resource, its utilization must be approached with an understanding of patients' cultural diversity [31].

This research is relevant as it is the first to evaluate the performance of ChatGPT as a counselling tool for Italian-speaking patients. While previous studies have explored the use of AI language models in English-speaking contexts, this study addresses the crucial need for such tools also in non-English speaking environments, particularly among Italian-speaking MASLD patients. Additionally, it benefits from the participation of experts from different age groups and academic positions in evaluating ChatGPT responses. The participation of a panel of highly qualified experts in the field of MASLD ensures the robustness of the evaluations and the applicability of the results to clinical practice. Moreover, the independent and blind evaluation of responses by experts helps to reduce the risk of bias and ensures the objectivity of the assessments. Lastly, the inclusion of a non-medical evaluator with specific expertise in patient advocacy for liver diseases offers a complementary perspective and enriches our understanding of the accessibility and usefulness of AI-based counselling resources for a wide range of users.

However, it is acknowledged that this study has several limitations. Initially, we enlisted a restricted group of experts to assess the answers based on subjective and self-reported ratings, using Likert scales. However, this method may not be optimal for evaluating medical texts, as it allows responses to be classified as 'nearly all correct'. In a medical context, an answer that is 'almost' correct can still be significantly incorrect, with potential adverse consequences. Another potential limitation to consider is the availability of new and potentially improved versions of ChatGPT (e.g., ChatGPT 4). Nonetheless, it should be noted that, not being freely accessible, it may be difficult for patients to access.

In conclusion, while our study presents promising results, the universal adoption of ChatGPT as a stand-alone counselling tool poses challenges. The identified limitations, particularly the potential for hallucinatory content and socio-cultural conditioning, underscore the need for continued refinement and validation of AI models in healthcare settings. For the future integration of ChatGPT or similar tools, collaboration between AI developers, healthcare professionals, and cultural experts is imperative in order to improve accuracy, cultural sensitivity, and overall reliability.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jpm14060568/s1>, Table S1: Likert scales.

Author Contributions: Conceptualization, N.P., R.L., C.H. and A.A. (Alessio Aghemo); Methodology, N.P., D.P., C.H. and A.A. (Alessio Aghemo); Investigation, R.L., G.P., F.R., A.A. (Angelo Armandi), E.B., A.D., A.L., A.M., R.V., I.G., L.V., L.M., S.P. and G.S.; Data curation, N.P., D.P., C.H. and A.A. (Alessio Aghemo); Writing—original draft, N.P. and D.P.; Writing—review and editing, R.L., G.P., F.R., A.A. (Angelo Armandi), E.B., A.D., A.L., A.M., R.V., I.G., C.H., L.V., L.M., S.P., G.S. and A.A. (Alessio Aghemo); Supervision, A.A. (Alessio Aghemo). All authors have read and agreed to the published version of the manuscript.

Funding: Giada Sebastiani is supported by a Senior Salary Award from Fonds de Recherche du Quebec—Santé (FRQS) (#296306). The research leading to these results was funded by MIUR under PNRR M4C2I1.3 Heal Italia project PE00000019 CUP B73C22001250006 to Prof. Salvatore Petta who is also supported by the Italian PNRR-MAD-2022-12375656 project, PRIN 2022 2022L273C9 project and RF-2021-12372399 project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Materials; further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to acknowledge Francesca Casamassima for her support in statistical analysis and Laurent Castera, Pinelopi Manousou, Raquel Peck, Zobair M Younossi, Vincent Wai-Sun Wong and Manuel Romero-Gomez for their help as part of the NAFLD Expert Chatbot Working Group.

Conflicts of Interest: The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Younossi, Z.M.; Golabi, P.; Paik, J.M.; Henry, A.; Van Dongen, C.; Henry, L. The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): A systematic review. *Hepatology* **2023**, *77*, 1335–1347. [[CrossRef](#)]
2. Lazarus, J.V.; Mark, H.E.; Villota-Rivas, M.; Palayew, A.; Carrieri, P.; Colombo, M.; Ekstedt, M.; Esmat, G.; George, J.; Marchesini, G.; et al. The global NAFLD policy review and preparedness index: Are countries ready to address this silent public health challenge? *J. Hepatol.* **2022**, *76*, 771–780. [[CrossRef](#)] [[PubMed](#)]
3. Quek, J.; Chan, K.E.; Wong, Z.Y.; Tan, C.; Tan, B.; Lim, W.H.; Tan, D.J.H.; Tang, A.S.P.; Tay, P.; Xiao, J.; et al. Global prevalence of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in the overweight and obese population: A systematic review and meta-analysis. *Lancet Gastroenterol. Hepatol.* **2023**, *8*, 20–30. [[CrossRef](#)]
4. Rinella, M.E.; Lazarus, J.V.; Ratzliff, V.; Francque, S.M.; Sanyal, A.J.; Kanwal, F.; Romero, D.; Abdelmalek, M.F.; Anstee, Q.M.; Arab, J.P.; et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *J. Hepatol.* **2023**, *79*, 1542–1556. [[CrossRef](#)]
5. Powell, E.E.; Wong, V.W.-S.; Rinella, M. Non-alcoholic fatty liver disease. *Lancet* **2021**, *397*, 2212–2224. [[CrossRef](#)]
6. Harrison, S.A.; Bedossa, P.; Guy, C.D.; Schattenberg, J.M.; Loomba, R.; Taub, R.; Labriola, D.; Moussa, S.E.; Neff, G.W.; Rinella, M.E.; et al. A Phase 3, Randomized, Controlled Trial of Resmetirom in NASH with Liver Fibrosis. *N. Engl. J. Med.* **2024**, *390*, 497–509. [[CrossRef](#)] [[PubMed](#)]
7. Petta, S.; Targher, G.; Romeo, S.; Pajvani, U.B.; Zheng, M.; Aghemo, A.; Valenti, L.V.C. The first MASH drug therapy on the horizon: Current perspectives of resmetirom. *Liver Int.* **2024**. [[CrossRef](#)] [[PubMed](#)]
8. Harrison, S.A.; Allen, A.M.; Dubourg, J.; Noureddin, M.; Alkhoury, N. Challenges and opportunities in NASH drug development. *Nat. Med.* **2023**, *29*, 562–573. [[CrossRef](#)]
9. European Association for the Study of the Liver (EASL); European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J. Hepatol.* **2016**, *64*, 1388–1402. [[CrossRef](#)]
10. Rinella, M.E.; Neuschwander-Tetri, B.A.; Siddiqui, M.S.; Abdelmalek, M.F.; Caldwell, S.; Barb, D.; Kleiner, D.E.; Loomba, R. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. *Hepatology* **2023**, *77*, 1797–1835. [[CrossRef](#)]
11. Pugliese, N.; Torres, M.C.P.; Petta, S.; Valenti, L.; Giannini, E.G.; Aghemo, A. Is there an ‘ideal’ diet for patients with NAFLD? *Eur. J. Clin. Investig.* **2022**, *52*, e13659. [[CrossRef](#)] [[PubMed](#)]
12. Romero-Gómez, M.; Zelber-Sagi, S.; Trenell, M. Treatment of NAFLD with diet, physical activity and exercise. *J. Hepatol.* **2017**, *67*, 829–846. [[CrossRef](#)] [[PubMed](#)]
13. Vilar-Gomez, E.; Martinez-Perez, Y.; Calzadilla-Bertot, L.; Torres-Gonzalez, A.; Gra-Oliveras, B.; Gonzalez-Fabian, L.; Friedman, S.L.; Diago, M.; Romero-Gomez, M. Weight Loss Through Lifestyle Modification Significantly Reduces Features of Nonalcoholic Steatohepatitis. *Gastroenterology* **2015**, *149*, 367–378.e5. [[CrossRef](#)] [[PubMed](#)]

14. Ding, R.; Kong, Q.; Sun, L.; Liu, X.; Xu, L.; Li, Y.; Zuo, X.; Li, Y. Health information in short videos about metabolic dysfunction-associated steatotic liver disease: Analysing quality and reliability. *Liver Int.* **2024**. [[CrossRef](#)] [[PubMed](#)]
15. Feng, S.; Mantymaki, M.; Dhir, A.; Salmela, H. How Self-tracking and the Quantified Self Promote Health and Well-being: Systematic Review. *J. Med. Internet Res.* **2021**, *23*, e25171. [[CrossRef](#)] [[PubMed](#)]
16. Lazarus, J.V.; Villota-Rivas, M.; Jimenez-Gonzalez, C.; Santos-Laso, A.; Iruzubieta, P.; Arias-Loste, M.T.; Rice-Duek, L.; Leigh, S.; Kopka, C.J.; Turnes, J.; et al. Physicians? Use of Digital Health Interventions in the Management of Nonalcoholic Fatty Liver Disease. *Clin. Liver Dis.* **2023**, *27*, 515–533. [[CrossRef](#)]
17. Carrieri, P.; Mourad, A.; Marcellin, F.; Trylesinski, A.; Calleja, J.L.; Protopopescu, C.; Lazarus, J.V. Knowledge of liver fibrosis stage among adults with NAFLD/NASH improves adherence to lifestyle changes. *Liver Int.* **2022**, *42*, 984–994. [[CrossRef](#)] [[PubMed](#)]
18. Sallam, M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare* **2023**, *11*, 887. [[CrossRef](#)] [[PubMed](#)]
19. Varghese, J.; Chapiro, J. ChatGPT: The transformative influence of generative AI on science and healthcare. *J. Hepatol.* **2024**, *80*, 977–980. [[CrossRef](#)]
20. Blease, C.; Torous, J. ChatGPT and mental healthcare: Balancing benefits with risks of harms. *BMJ Ment. Health* **2023**, *26*, e300884. [[CrossRef](#)]
21. Lee, P.; Bubeck, S.; Petro, J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *N. Engl. J. Med.* **2023**, *388*, 1233–1239. [[CrossRef](#)] [[PubMed](#)]
22. OpenAI. Introducing ChatGPT. Available online: <https://openai.com/blog/chatgpt> (accessed on 30 November 2022).
23. Pugliese, N.; Wong, V.W.-S.; Schattenberg, J.M.; Romero-Gomez, M.; Sebastiani, G.; Aghemo, A.; Castera, L.; Hassan, C.; Manousou, P.; Miele, L.; et al. Accuracy, Reliability, and Comprehensibility of ChatGPT-Generated Medical Responses for Patients With Nonalcoholic Fatty Liver Disease. *Clin. Gastroenterol. Hepatol.* **2024**, *22*, 886–889.e5. [[CrossRef](#)] [[PubMed](#)]
24. Samaan, J.S.; Yeo, Y.H.; Ng, W.H.; Ting, P.-S.; Trivedi, H.; Vipani, A.; Yang, J.D.; Liran, O.; Spiegel, B.; Kuo, A.; et al. ChatGPT's ability to comprehend and answer cirrhosis related questions in Arabic. *Arab. J. Gastroenterol.* **2023**, *24*, 145–148. [[CrossRef](#)] [[PubMed](#)]
25. ChatGPT (2023). Available online: <https://chat.openai.com> (accessed on 1 October 2023).
26. Likert, R. A technique for the measurement of attitudes. *Arch. Psychol.* **1932**, *22*, 55.
27. Moor, M.; Banerjee, O.; Abad, Z.S.H.; Krumholz, H.M.; Leskovec, J.; Topol, E.J.; Rajpurkar, P. Foundation models for generalist medical artificial intelligence. *Nature* **2023**, *616*, 259–265. [[CrossRef](#)] [[PubMed](#)]
28. Calderaro, J.; Seraphin, T.P.; Luedde, T.; Simon, T.G. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. *J. Hepatol.* **2022**, *76*, 1348–1361. [[CrossRef](#)]
29. Benary, M.; Wang, X.D.; Schmidt, M.; Soll, D.; Hilfenhaus, G.; Nassir, M.; Sigler, C.; Knoedler, M.; Keller, U.; Beule, D.; et al. Leveraging Large Language Models for Decision Support in Personalized Oncology. *JAMA Netw. Open* **2023**, *6*, e2343689. [[CrossRef](#)]
30. Giuffrè, M.; You, K.; Shung, D.L. Evaluating ChatGPT in Medical Contexts: The Imperative to Guard Against Hallucinations and Partial Accuracies. *Clin. Gastroenterol. Hepatol.* **2023**, *22*, 1145–1146. [[CrossRef](#)]
31. Tariq, R.; Khanna, S. AI-Powered Patient Education for Comprehensive and Individualized Understanding for Patients. *Clin. Gastroenterol. Hepatol.* **2023**. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.