information

MDPI

Article

A Survey on Using Linguistic Markers for Diagnosing Neuropsychiatric Disorders with Artificial Intelligence

Ioana-Raluca Zaman [1] and Stefan Trausan-Matu [1,2,*]

[1] Department of Computer Science and Engineering, Politehnica Bucharest, National University for Science and Technology, 060042 Bucharest, Romania; ioana_raluca.zaman@stud.acs.upb.ro
[2] Research Institute for Artificial Intelligence "Mihai Draganescu" of the Romanian Academy, 050711 Bucharest, Romania
* Correspondence: stefan.trausan@upb.ro

Abstract: Neuropsychiatric disorders affect the lives of individuals from cognitive, emotional, and behavioral aspects, impact the quality of their lives, and even lead to death. Outside the medical area, these diseases have also started to be the subject of investigation in the field of Artificial Intelligence: especially Natural Language Processing (NLP) and Computer Vision. The usage of NLP techniques to understand medical symptoms eases the process of identifying and learning more about language-related aspects of neuropsychiatric conditions, leading to better diagnosis and treatment options. This survey shows the evolution of the detection of linguistic markers specific to a series of neuropsychiatric disorders and symptoms. For each disease or symptom, the article presents a medical description, specific linguistic markers, the results obtained using markers, and datasets. Furthermore, this paper offers a critical analysis of the work undertaken to date and suggests potential directions for future research in the field.

Keywords: neuropsychiatric disorders; depression; dementia; hallucinations; linguistic markers; natural language processing; artificial intelligence

Citation: Zaman, I.-R.; Trausan-Matu, S. A Survey on Using Linguistic Markers for Diagnosing Neuropsychiatric Disorders with Artificial Intelligence. *Information* **2024**, *15*, 123. https://doi.org/10.3390/info15030123

Academic Editor: Arkaitz Zubiaga

Received: 30 January 2024
Revised: 10 February 2024
Accepted: 19 February 2024
Published: 22 February 2024

1. Introduction

In recent years, the advances of Artificial Intelligence (AI) have been seen in different areas of medicine, such as: oncology [1], cardiology [2], endocrinology [3], neurology, and psychiatry [4,5]. Neuropsychiatric disorders are becoming a challenge faced by more and more people nowadays. The conditions include both mental health problems (e.g., depression, anxiety, and schizophrenia) and neurological diseases (e.g., Alzheimer's disease, Parkinson's disease, and epilepsy) [6]. One challenge regarding the detection and the understanding of the disorders is the complexity of the symptoms, which vary from patient to patient but also overlap between certain diseases. Problems related to neuropsychiatric conditions are encountered more and more often, especially due to certain contexts (e.g., epidemics) or for categories exposed to certain factors (e.g., low income) [7]. In a meta-analysis [8], it was discovered that the emergence of the first mental disorder takes place before the age of 14 in over a third of cases (34.6%) and before the age of 25 in almost two-thirds of cases (62.5%). Therefore, particularly for this group of disorders, early detection has a significant impact; applying treatment in time ensures that worsening of the symptoms is slowed down and that the patients have the needed support.

One method for the discovery of new and less obvious symptoms of neuropsychiatric disorders implies studying the language of people, focusing on clues unnoticeable by humans (e.g., the presence or high frequency of specific words, syntactic density, grammar complexity, etc.) [9,10]. In order to find these differences between healthy people and those suffering from certain neuropsychiatric disorders, their speech may be analyzed using AI Natural Language Processing (NLP) methods. This paper presents some of such work and also analyses their approaches.

In recent years, NLP systems have used several Machine Learning (ML) techniques, especially Deep Artificial Neural Networks (DANNs), which perform very well and can include analyzing patients' utterances in neuropsychiatric dialogues [11]. However, people need to trust the decisions made by the ML models, particularly in the medical field. Currently, Transformer-based models [12] have the best performance; however, their results are based only on experience, which can cause the classifications to be based on superficial or invalid criteria [13]. Analyzing conversations from patients and finding patterns in data using AI tools should also allow the interpretability of the results provided by DANNs (which can be seen as black-box models), helping people to have more trust in the AI's contributions to medicine. An online study (N = 415) that measured people's trust in the involvement of AI in medicine based on questionnaires referring to medical scenarios demonstrated that people still have more trust in a doctor than in an AI model [14]. Linguistic markers are characteristics or traits of the text or speech that can provide information about the speaker. These markers can be divided into several categories, for example: grammar markers or lexical semantic markers [15]. If such markers (which can be understood by humans) would be provided for assisting the diagnosis of a patient, the interaction between AI systems based on ML models and doctors would face fewer challenges, and patients would be more open to considering the indications coming from AI.

For a clear and systematic picture aiming to aid the reader with understanding this paper, a concept map illustrating a summary of the main topics and their relations discussed in our work is shown in Figure 1.

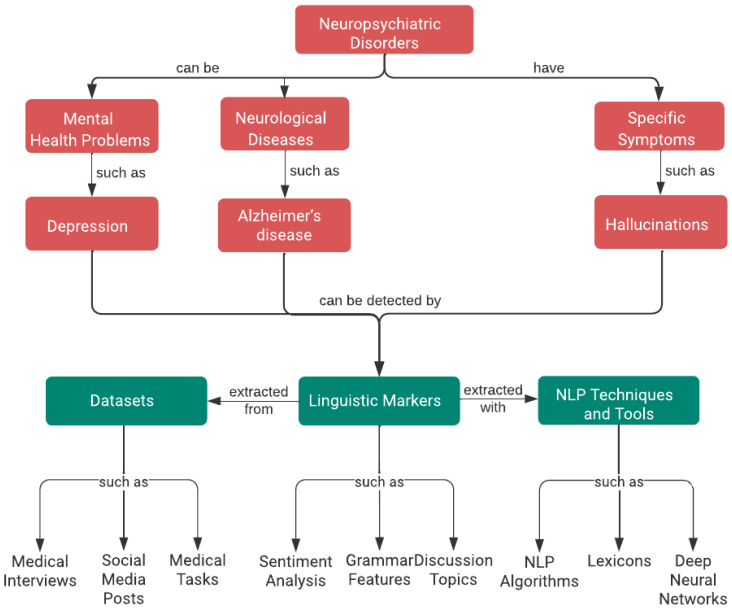

Figure 1. Concept map of the main topics and their relations discussed in the paper.

2. Materials and Methods

A formal literature search was conducted on Google Scholar from 29 August 2023 to 7 September 2023. The used search terms were the following: *("depression" OR "dementia" OR "Alzheimer's disease" OR "hallucinations") AND ("linguistic markers" OR "linguistic analysis" OR "linguistic style")*. There were several inclusion and exclusion criteria used in this study. Firstly, the year of publication was chosen to be at least 2015 in order to analyze only information from recent years when ML and especially DANN architectures dramatically increased the performance of the implemented systems. Another screening criterion involved the domain. This research exclusively incorporated papers related to Computer Science. Consequently, papers addressing neuropsychiatric linguistic analysis through an AI-related approach were taken into account. Research studies originating from diverse domains, such as Medicine, were not taken into account. The ultimate criterion

pertained to the language of the publication, wherein only documents available in English were included. This criterion aimed to reduce the complications linked to the process of translation. Subsequently, for the selected papers, their eligibility was tested firstly by the abstracts of the papers, and then full papers were subjected to a detailed review. This process was important to guarantee that only those papers having linguistic analysis of the mentioned disorders or symptoms as their focus were examined. Regarding the datasets, in addition to those utilized in the selected publications, datasets found using the following terms were also selected: *("depression" OR "dementia" OR "Alzheimer's disease" OR "hallucinations") AND ("dataset" OR "corpus")*.

3. Medical Descriptions

This section provides a description of the clinical characteristics, symptoms, and impacts of depression and the neurocognitive disorder (NCD) dementia. As regards NCDs, Alzheimer's disease (AD), the most common form of dementia, will be studied in particular. Moreover, hallucinations will be analyzed, these being a specific symptom of several mental diseases. In addition to that, a comparison between hallucinations produced by humans and those produced artificially by Large Language Models (LLMs—DANNs trained with a huge number of texts) [16] will be illustrated. A deep understanding of the medical symptoms of neuropsychiatric diseases is relevant for effective application of NLPs in studies. Knowing the medical symptoms of the disorders can help with finding associations between certain symptoms and patterns in speech or text. For instance, if two diseases have a common symptom, it could be useful to search for the same linguistic features associated with that symptom for both diseases.

3.1. Depression

Depression, medically known as major depressive disorder (MDD) [17], is the general condition for the class of depressive disorders. It can be seen in several forms, from medication-induced depressive disorder to dysthymia (persistent depressive disorder), and the disease is marked by distinct episodes lasting at least two weeks [17]. The criteria on which the diagnosis of this disease is based are the following: depressed mood (i.e., from feeling hopeless to even feeling irritated, especially for adolescents and children), deeply reduced enjoyment in activities, notable changes in appetite and weight, daily sleep problems (i.e., insomnia or hypersomnia), overwhelming fatigue, feelings of worthlessness or guilt, and in some cases even delusion thoughts, indecisiveness or trouble concentrating, and even suicidal thoughts [17]. Depression's evolution or appearance can be influenced by various risk factors such as: temperament (i.e., neurotic people have a tendency towards anxiety, anger, and emotional instability [18]), environment (i.e., shocking events, especially in childhood), and genetics.

3.2. Dementia and Alzheimer's Disease

Dementia involves conditions wherein the main problem affects the cognitive functions that were acquired rather than present from birth [19]. These conditions affect a category of people over a certain age; at the age of 65, the general prevalence of dementia is approximately 1–2%, and by the age of 85, it is 30% [17]. Dementia is a general term that refers to a series of diseases having various early symptoms depending on which area of the brain is affected [20]. Due to the fact that in the majority of cases of AD, the first part of the brain affected is the hippocampus, the patient initially has problems remembering facts from the recent past. After that, if the amygdala is affected, the person refers to memories more from an emotional point of view than a factual one. As AD progresses, its damage extends to various brain areas and lobes, resulting in a thinner cortex and overall brain shrinkage. The left hemisphere's impairment leads to issues with semantic memory and language, causing difficulty with word retrieval. Damage to the visual system in the temporal lobes hampers face and object recognition, although auditory recognition might still be possible. Right parietal lobe damage affects spatial judgment (e.g., tasks

like navigating stairs). Frontal lobe damage results in challenges with decision making, planning, and organizing complex tasks. Despite these losses, long-acquired abilities like procedural memories (e.g., dancing or playing the piano) tend to remain intact, even in advanced stages of AD [20]. Besides AD, there are other forms of dementia such as: vascular dementia, frontotemporal dementia, Lewy body dementia [21], and dementia caused by other diseases (e.g., Huntington's disease or Parkinson's disease) [22].

3.3. Hallucinations

Hallucinations are a symptom present in a variety of diseases, from mental health conditions such as psychotic depression to AD; however, most often they are found in conditions within the spectrum of schizophrenia [23]. This symptom manifests as vivid experiences resembling perceptions but arising without the external triggers [17]. Hallucinations can affect all the senses; however, a survey conducted on 10,448 participants showed that the most frequent are auditory hallucinations (29.5%) (e.g., voices, laughing, and crying), succeeded by visual hallucinations (21.5%) (e.g., shadows and people moving), tactile hallucinations (19.9%) (e.g., being touched and formication), and olfactory hallucinations (17.3%) (e.g., fire, food, and drinks) [24]. Besides these types, there are also gustatory hallucinations (e.g., metallic taste), presence hallucinations (i.e., the feeling that someone is present in the room or behind the subject), and proprioceptive hallucinations (i.e., the feeling that the body is moving) [23].

4. State of the Art

This section provides an overview of the current state of the art in utilizing AI for analyzing neuropsychiatric disorders and their symptoms. The section is structured as follows. The first subsection presents NLP techniques used to understand linguistic signs in conversations about the disorders. It highlights the use of sentiment analysis, topic modeling, and patterns concerning depression, dementia, and hallucinations. In the second subsection, we examine the distinctive linguistic markers associated with each of the diseases. Additionally, a comparison between the differences in linguistic markers between human- and LLM-generated hallucinations is illustrated. The last subsection examines datasets in order to offer insights into selecting suitable resources for NLP-based analysis of neuropsychiatric disorders.

4.1. NLP Techniques

In the field of NLP, a variety of techniques and tools have been employed to investigate linguistic markers associated with neuropsychiatric disorders and have provided valuable insights from the textual data from individuals with these conditions. In this section, the techniques and tools used in state-of-the-art works will be presented. All the studies and techniques mentioned in this section will be presented in more detail in Section 4.2.

Sentiment analysis is a fundamental method utilized to evaluate the emotional tone and sentiments from text or speech. One of the main approaches for sentiment analysis is the usage of lexicons particularly created for this task. Linguistic Inquiry and Word Count (LIWC) [25] is a lexicon-based tool used by researchers for extracting emotional and psychological dimensions. This tool was used to extract features for predicting depression and anxiety from therapy sessions [26] and for the detection of Reddit posts related to depression [27,28]. There are sentiment lexicons specialized for scores of positivity, negativity, and neutrality, such as: SentiWordNet [29] and VADER [30]. The former was utilized by Titla-Tlatelpa et al. [31] for extracting the polarity of posts in order to create a user profile. Moreover, lexicons designed for specific linguistic markers can be created: for instance, the Behaviour Activation lexicon [32].

Topics of discussion represent indicators of certain mental disorders, and they can be identified by selecting key words. One often utilized method for this task is to consider the Term Frequency–Inverse Document Frequency (TF-IDF) [33], which measures the importance of words within a corpus. A smoothed version of TF-IDF was used by

Wolohan et al. [28], who combined it with LIWC and n-grams (sequences of n words) in order to capture word sequences and patterns. Another topic modeling algorithm is Latent Dirichlet Allocation (LDA) [34]; an example of using this method is illustrated in the work of Tadesse et al. [27]. Furthermore, tools such as KHCoder [35] can be utilized for plotting co-occurrence networks or other statistics from texts. The results from part-of-speech (POS) tagging tasks [36] are also relevant markers for neuropsychiatric disorders. For certain disorders (e.g., depression), the tense of verbs is an important clue, and tools such as the Stanford University Time (SUTime) temporal tagging system [37] can be used for analyzing the tenses.

4.2. Linguistic Markers

4.2.1. Depression

Understanding how language can reveal insights about depression has become an area of growing interest that is marked by evolving findings and methodologies. There exists an established association between self-centeredness and depression [38], and this preoccupation with the self is also reflected in linguistic patterns [27]. In the meta-analysis conducted by Tølbøll [39], 26 papers published between the years 2004 and 2019 were examined to study the link between the existence and severity of depression and first-person singular pronouns (e.g., 'I' and 'me'), positive emotion words, and negative emotion words. The conclusions related to the usage of first-person singular pronouns and depression indicated a medium effect (Cohen's d of 0.44) and a positive correlation (Pearson's r of 0.19). One study analyzed Reddit posts from 12,106 users and reconfirmed the link between first-person singular pronouns and depression [28]. Furthermore, the authors found that individuals experiencing depression used more dialogue terms in their posts, specifically addressing the readers using second-person pronouns (e.g., "you") and writing the posts as if talking directly to them. In addition to the linguistic markers discovered, Wolohan et al. [28] created a depression classification system that performed best using LIWC features and n-grams and achieved an F1 score of 0.729. Burkhardt et al. [26] evaluated therapy sessions on Talkspace from over 6500 unique patients and stated correlations between both first-person singular and plural pronouns, which is a conclusion that has also been validated in other research [32] for singular but not plural forms.

Regarding POS tagging, we analyzed the Distress Analysis Interview Corpus/Wizard-of-Oz (DAIC-WOZ) dataset [40] from the University of Southern California and concluded that individuals suffering from depression utilized fewer prepositions, conjunctions, and singular proper nouns. Regarding verb tenses, depressives also have a tendency to use more verbs in gerund or past participle form [41]. Moreover, there are studies supporting future language as an indicator of depression [42]. Using SUTime, the authors discovered that depressed participants refer to future events more distally and think more about the past and future rather than the present. The researchers created an FTR (Future Time Reference) classifier that offers more information about the modality of verbs and achieved an F score over 0.99 on a Reddit post classification task.

Some emotions are more often found in people with certain mental disorders. Tølbøll [39] discovered a strong effect (Cohen's d of 0.72) between depression and negative emotions and a negative correlation (Pearson's r of -0.21) between the disease and the usage of positive words; they also confirmed the correlation between negative emotions and depression for the analyzed conversations. Burkhardt et al. [26] extracted 49 emotion-related features using both LIWC and the GoEmotion dataset [43]. The authors measured the explanatory impact of each feature by using the amount of variance explained by R2 (i.e., the variability of a dependent variable that is explained by an independent variable in a regression model), and the top LIWC features for depression had values in the interval [0.716, 0.729]. With the first tool, sadness, anger, anxiety, and both negative and positive emotions were identified as indicators of the mental disease. The most relevant emotions for depression are: grief, less pride, less excitement, relief, disgust [26], and fear [41]. These were confirmed as well in the work of Tadesse et al. [27], which additionally highlighted:

words of anger and hostility (e.g., hate), suicidal thoughts (e.g., stop-stop, want die), interpersonal processes (e.g., feel alone, lonely), and cues of meaninglessness (e.g., empty, pointless) and hopelessness (e.g., end, need help). Moreover, the usage of absolutist words (e.g., all, always, never) is a marker for depression and its symptoms of suicidal ideation and anxiety [44].

The topics addressed in discussions can be indicators for depression. One method to acquire the topics is by selecting the 100 most-used words with TF-IDF and dividing them into categories. Using this methodology, Wolohan et al. [28] concluded that people suffering from depression more often discuss: therapy (e.g., psychiatrist) and medications (e.g., Prozac) or Reddit, manga, and video games (e.g., Goku). By developing a new lexicon, [26] found that depressed individuals more frequently approach subjects from biology and health categories, and individuals having severe depression talk less about activities and relate them less with positive feelings (e.g., enjoyment, reward). Using LIWC, Tadesse et al. [27] detected correlations (with a Pearson's r coefficients in the interval [0.11, 0.19]) between depressed people and psychological processes such as social processes (e.g., mate, talk), affective processes (e.g., cry, hate), cognitive processes (e.g., know, think), as well as personal concerns such as work, money, and death using. By analyzing depression-related text with LDA, Tadesse et al. [27] selected the top 20 most frequent topics, combined the extracted features with LIWC, bigrams, and an MLP, and obtained an F1 score of 0.93 on a Reddit post classification task. The authors reconfirmed the job topic but also added keywords such as: tired, friends, and broke; they also added sleep and help [41]. In their study, they used the KHCoder tool to identify the topics of the interviews using co-occurrence networks and concluded that in terms of relationships, depressed people talked more about child–parent relationship, while the control group talked more about friends and family, and in terms of jobs, the first category referred more to finding a job, while the second category referred to a dream job. Another approach is to take into consideration the profile (i.e, gender and age) of the speaker when analyzing the text for depression and using age-based classifiers and gender-based classifiers [31]. With this methodology, the authors revealed differences between depressed and non-depressed users per category (e.g., the word calories used in negative contexts can be a marker for depression in young females, while drunk can be used as a marker for senior male users).

4.2.2. Dementia and Alzheimer's Disease

Although it is a field that is just at the beginning, lexical–semantic and acoustic metrics show promising results as digital voice biomarkers for AD [45]. Automating the analysis of vocal tasks like semantic verbal fluency and storytelling provides an undemanding method to support early patient characterization. Some of the speech features we extracted have unique patterns (e.g., the ones related to tone and rhythm). This method could be used as a clear way to tell if someone has depression or mild cognitive problems [46]. Patients with AD have shortfalls with using verbs and nouns [47]: especially verbs during arguments [48]. Using only information from POS tagging, some features (e.g., readability, propositional density, and content density) can be extracted and show promising results for AD classification tasks. For instance, Guerrero et al. [49] achieved an accuracy of 0.817 for Pitt corpus by using a Random Forest (RF) model and, as input, a fusion of features extracted from grammar characteristics, TF-IDF, and Word2Vec (W2V). Eyigoz et al. [50] predicted the beginning of AD by analyzing linguistic characteristics. One of the conclusions they reached was that participants who will be diagnosed with AD had telegraphic speech, writing mistakes, and more repetitions. Telegraphic speech is summarized and contains only the essential words (e.g., nouns and verbs), the connective POS (e.g., articles or adverbs) being omitted. Another characteristic of AD speech was referential specificity: a semantic feature by which unique nouns are differentiated from general nouns (e.g., proper names). More studies support the idea that one of the earliest signs in terms of linguistics is semantic impairment [51,52]. Karlekar et al. [53] identified clusters specific to this disease: namely, clarification questions (e.g., 'Did I say elephant?'), outbursts in speaking and brief answers

(e.g., 'oh!', 'yes'), and statements starting with interjections (e.g., 'Well ... ', 'Oh ... '). An accuracy of 0.911 was obtained by researchers [53] in an experiment using POS-tagged utterances and a CNN-RNN model. In the case of dementia and AD, the results can be improved by combining linguistic markers with features extracted using Computer Vision (CV) or biomarkers. For instance, Koyama et al. [54] highlighted the role of peripheral inflammatory markers in dementia and AD and found links between increased levels of C-reactive protein or interleukin-6 and dementia. By using CV, neuroimaging techniques can be utilized to detect changes in the brain that are signs of AD or mild cognitive impairment (MCI), such as increased grey matter brain atrophy or hyperactivation within the hippocampus memory network [55].

4.2.3. Hallucinations

Hallucinations from People with Neuropsychiatric Disorders

Hallucinations are a complex phenomenon that can manifest in a unique way from person to person. This symptom, especially an auditory one, is difficult to detect, particularly the moment of its appearance, but using mobile data [56], dictation devices [57], or auditory verbal recognition tasks [58], it is still possible. In accord with a review [59], hallucinations are influenced by cultural aspects such as: religion, race, or environment (e.g., magical delusions exhibited a high frequency in rural areas). Gender is not a factor for auditory hallucinations, but female patients reported experiencing olfactory, tactile, and gustatory hallucinations more frequently [60].

In a study of Dutch language [61], the researchers compared the auditory verbal hallucinations from clinical (i.e., diagnosed with schizophrenia, bipolar disorder, or psychotic disorder) and non-clinical participants and observed that the hallucinations from the first category of participants were more negative (i.e., 34.7% vs. 18.4%); this aspect was also confirmed by [9]. They identified the most frequently encountered semantic classes in the auditory hallucinations in Twitter posts, with the top three being abusive language (e.g., hell), relatives (e.g., son), and religious terms (e.g., prayer), followed by semantic classes related to the sense of hearing (e.g., audio recording, audio and visual media, audio device, or communication tools). Another observation is that tweets containing auditory hallucinations exhibited a greater proportional distribution during the hours of 11 p.m. to 5 a.m. compared to other tweets. By using a set of 100 semantic features, the authors of [9] classified if a Twitter post was related to auditory hallucination and with a Naive Bayes (NB) model reached an AUC of 0.889 and an F2 score of 0.831; the baseline value was 0.711. In this study, the leave-one-out technique showed that the best results were obtained when lexical distribution features were excluded (i.e., an AUC of 0.889 and F2 score of 0.833).

Artificial Hallucinations from ML Models

This subsection presents specific contexts (e.g., tasks or topics of discussion) in which hallucinations were not emitted by humans but were generated from AI systems that generate texts based on LLMs. The models from the studies presented in this section represent a range of representative DANN models, such as: Generative Pre-trained Transformer models (e.g., GPT-2, GPT-4, and ChatGPT) or Transformer-based multimodal models (e.g., VLP, and LXMERT-GDSE). Image captioning is a task in which models may hallucinate; for example, Testoni and Bernardi [62] used the GuessWhat?! game (the goal of the game is for one player to guess the target object by asking the other player binary questions about an image) to force the models to produce hallucinations. The majority of hallucinations manifested in consecutive turns, leading to hypotheses such as previous triggering words and the cascade effect (i.e., the amplification of hallucinations) [62–65]; these phenomena are not present in human hallucinations. However, the models can detect that they are wrong: ChatGPT [66] detects 67.37% of cases and GPT-4 detects 87.03% [65]. Another difference is that in these experiments, the hallucinations appeared more frequently after negative responses; in human dialogues, this is not the case.

Dziri et al. [63] tried to discover the origin of hallucinations in conversational models based on Verbal Response Modes (VRM) [67] and affirmed that the most effective strategies for creating hallucinations were disclosure (i.e., sharing subjective opinions) and edification (i.e., providing objective information). The researchers [63] also studied the level of amplification of hallucinations and concluded that, for instance, GPT2 amplifies full hallucinations by 19.2% in the Wizard of Wikipedia (WOW) dataset. Alkaissi and McFarlane [68] tested ChatGPT for scientific writing, and the model generated nonexistent paper titles with unrelated PubMed IDs and artificial hallucinations [69] regarding medical subjects. Self-contradiction is a form of hallucination that can appear in human hallucinations and LLM-generated hallucinations; for the second type of hallucinations, there are algorithms regarding their evaluation, detection, and mitigation [70]. The authors created a test covering 30 subjects (e.g., Angela Merkel and Mikhail Bulgakov) for the models, and for the detection task, they achieved F1 scores with values up to 0.865.

An overview of each research study presented in Sections 4.1 and 4.2 in chronological order and grouped by medical condition is shown in Tables 1–4 following.

Table 1. Overview of the linguistic markers for depression extracted in the selected papers. Source: Own work.

Dataset Source	Data Type	Linguistic Markers or Features	Tools and Techniques	Year	Ref.
Reddit	Reddit posts	N-grams, topics, psychological and personal concern process features	N-grams, LDA, LIWC	2019	[27]
Reddit	Reddit posts	N-grams, topics, grammatical features, emotions	N-grams, smoothed TF-IDF, LIWC	2019	[28]
Reddit and Twitter	Social media posts	Polarity, gender, age, Bow/BoP representations	Bag of Words (BoW), Bag of Polarities (BoP), SentiWordNet	2021	[31]
Talkspace	Messaging therapy sessions	Grammatical features, topics and emotions	LIWC, GoEmotions	2022	[26]
Reddit	Reddit posts	Temporal features, modal semantics	SUTime	2022	[42]
Public forums	Forum posts	Absolutist index, LIWC features	LIWC, absolutist dictionary	2022	[44]
DAIC-WOZ	Clinical interviews	POS tagging, grammatical features, topics and emotions	NLTK, NRCLex, TextBlob, pyConverse, KHCoder	2023	[41]

Table 2. Overview of the linguistic markers for dementia extracted in the selected papers. Source: Own work.

Dataset Source	Data Type	Linguistic Markers or Features	Tools and Techniques	Year	Ref.
Public blogs	Posts from public blogs	Context-free grammar features, POS tagging, syntactic complexity, psycholinguistic features, vocabulary richness, repetitiveness	Stanford Tagger, Stanford Parser, L2 Syntactic Complexity Analyzer	2017	[10]
Pitt Corpus—Dementia Bank	Cookie Theft picture description task	Grammatical features, POS tagging	Activation clustering, first-derivative saliency heat maps	2018	[53]
Pitt Corpus—Dementia Bank	Cookie Theft picture description task	Word embeddings, grammatical features, POS tagging	Word2Vec, TF-IDF	2020	[49]
FHS study	Cookie Theft picture description task	Word embeddings, grammatical features, POS tagging	GloVe, NLTK	2020	[50]

4.3. Relevant Datasets

This subsection presents an overview of the relevant datasets used in state-of-the-art works in which the mentioned neuropsychiatric disorders were studied. These datasets are utilized for both the detection of the disorder and the extraction of linguistic markers specific to the disease. The data can be obtained by web scraping (e.g., social media posts), artificially (e.g., content generated with an LLM following a pattern), or from medical sources (e.g., dialogues between a patient and a doctor). Another aspect of the data is that it should be gathered over a period of time (e.g., having interviews with a patient over

five years periodically), which allows early detection and the evolution of symptoms to be studied.

Table 3. Overview of the linguistic markers for hallucinations from people extracted in the selected papers. Source: Own work.

Dataset Source	Data Type	Linguistic Markers or Features	Tools and Techniques	Year	Ref.
Twitter	Twitter posts	Semantic classes, POS tagging, use of nonstandard language, polarity, key phrases, semantic and lexical features	TweetNLP tagger, MySpell	2016	[9]
Clinical study	Audio reports from sleep onset and REM and non-REM sleep	Grammatical features	Measure of Hallucinatory States (MHS)	2017	[57]
"Do I see ghosts?" Dutch study	Auditory verbal recognition task	Age, gender, education, and the presence of visual, tactile, and olfactory hallucinations	IBM SPSS Statistics	2017	[57]
Clinical study	Electronic health records (EHRs)	Age, gender, race, NLP symptoms	Clinical Record Interactive Search (CRIS)	2020	[60]
Clinical study	Recordings of participants' hallucinations	Grammatical features, emotions, POS tagging	CLAN software, Pattern Python package, Dutch lexicons	2022	[61]
Clinical study	Audio diary by mobile phone with periodic pop-ups asking about the hallucinations	Word embeddings	VGGish model, BERT, ROCKET	2023	[56]

Table 4. Overview of the linguistic markers for artificial hallucinations extracted in the selected papers. Source: Own work.

Dataset Source	Data Type	Linguistic Markers or Features	Tools and Techniques	Year	Ref.
500 randomly selected images	Image captioning task	CHAIR metrics—CHAIR-i and CHAIR-s, METEOR, CIDEr, SPICE	MSCOCO annotations, FC model, LRCN, Att2In, TopDown, TopDown-BB, Neural Baby Talk (NBT)	2018	[64]
GuessWhat?! game	Utterances from GuessWhat?! game	CHAIR metrics—CHAIR-i and CHAIR-s, analysis of hallucinations	MSCOCO annotations, BL, GDSE, LXMERT-GDSE, VLP	2021	[62]
Wizard of Wikipedia (WOW), CMUDOG, TOPICALCHAT	Dialogues between two speakers	Hallucination rate, entailment rate, Verbal Response Modes (VRMs)	GPT2, DoHA, CTRL	2022	[63]
3 new datasets consisting of yes/no questions	QA task answers	Snowballing of hallucinations, hallucination detection, LM (in)consistency	ChatGPT, GPT-4	2023	[65]
Dataset consisting of generated encyclopedic text descriptions for Wikipedia topics	Description task	Average no. of sentences, perplexity, self-contradictory features	ChatGPT, GPT-4, Llama2-70B-Chat, Vicuna-13B	2023	[70]

4.3.1. Depression

In our depression study [41], we used the DAIC-WOZ dataset, which is a corpus containing the conversations between an agent Ellie and 189 participants: 133 non-depressed and 56 depressed. The agent is human-controlled and operates based on a predefined set of questions for the conversations. In order to label the participants, the Patient Health Questionnaire-8 (PHQ-8) is utilized, and for each entry, the database contains: an audio recording and transcript of the conversation, the responses for the PHQ-8, the gender of the participant, and metadata (i.e., non-verbal and verbal features). To minimize the effects of dataset imbalance, we created an additional subset of similar conversations of depressed patients using ChatGPT. Depression-related challenges are another source for datasets; for instance, DepreSym is a corpus created from the eRisk 2023 Lab.

A methodology used to retrieve medical dialogues is through online platforms, such as those specialized for therapy sessions (e.g., Talkspace) [26], or forums [44]. Extracting social media posts is a popular method for constructing new corpora; for instance, Shen et al. [71] developed from Twitter posts a dataset with three subsets: depression, non-depression, and depression candidate. Several researchers [31,72] have also used Twitter as a source for their data. Another social platform from which data are collected is Reddit [27,28,73].

4.3.2. Dementia and Alzheimer's Disease

One method to create a dataset for dementia or AD is from tasks designed to emphasize the particular symptoms of the conditions, such as "Boston Cookie Theft" (a task in which the participants were asked to describe a given picture) or a recall test (a task in which the participants were asked to recall attributes of a previously shown story or picture). DementiaBank [74] is a database of corpora containing video, audio, and transcribed materials for AD, Mild Cognitive Impairment (MCI), Primary Progressive Aphasia (PPA), and dementia in multiple languages (e.g., English, German, and Mandarin).

The Framingham Heart Study (FHS) is a study started in 1948 and which continues to this day. Its aim is to discover factors that play a role in the development of cardiovascular disease (CVD). However, it also contains recordings of participants suffering from conditions such as AD, MCI, or dementia. Researchers have used the data from this study in order to detect linguistic markers that can be utilized for the early prediction of the previously mentioned diseases [50,75]. Dementia and AD can also be studied in an online environment, such as blog posts. For instance, Masrani et al. [10] created the Dementia Blog Corpus by scraping 2805 posts from 6 public blogs, and the authors of [76] studied dementia using data from Twitter.

4.3.3. Hallucinations

One of the signs of the presence of hallucinations in speech can be the unreliability of the facts presented in the conversation. To highlight this sign, Liu et al. [77] created HaDeS (HAllucination DEtection dataSet), a corpus built by perturbing raw texts from the WIKI-40B [78] dataset using BERT [79], and then checked the validity of the hallucinations with human annotators. The authors of [80] studied the correlations between hallucinations and psychological experiences using a dataset containing 10,933 narratives from patients diagnosed with mental illnesses (e.g., schizophrenia or obsessive compulsive disorder); the data had been previously collected by the authors [81].

Artificial hallucinations are usually generated from conversational agents by using certain games [62] or by addressing sensitive subjects such as religion, politics, or conspiracy ideas. The Medical Domain Hallucination Test (Med-HALT) [82] is a collection of seven datasets containing hallucination from LLMs in the medical field. The datasets are based on two tasks: more precisely, the Reasoning Hallucination Test (RHT) (i.e., a task in which the model has to choose an option from a set of options for questions) and the Memory Hallucination Test (MHT) (i.e., a task in which the model has to retrieve information from a given input). The data utilized as input for the models are questions from medical exams (e.g., the residency examination from Spain and the United States Medical Licensing Examination (USMILE)) and PubMed.

5. Discussion and Challenges

A key area for the improvement of the discussed approaches involves the expansion and refinement of existing datasets and the development of new corpora; for instance, more emphasis should be on collecting data periodically over a longer period of time to study the evolution of diseases and to find the most relevant linguistic symptoms. Additionally, the building of diverse datasets covering various demographic groups and different stages of these disorders could improve the results. Integrating multimodal approaches that combine linguistic markers with medical imaging or other biological signals could offer a

more comprehensive understanding of these disorders. Correlating linguistic patterns with physiological and visual data may amplify the accuracy of early diagnosis and prediction.

Considering that ethics is indispensable in a project using data from people, especially such sensitive data as those from patients suffering from neuropsychiatric disorders, various aspects such as algorithmic fairness, biases, data privacy, informed consent to use, safety, and transparency [83] have to be taken into account for a project to be ethically valid. Fulfilling all these conditions can create difficulties in a project, such as non-cooperation and lack of patient consent for the collection of new data or legal challenges that require the involvement of legal professionals. Another problem is represented by the limited access to such data; for example, a significant part of medical datasets are accessible only to researchers affiliated with certain universities or having certain citizenship.

Another aspect is the interpretability of the results. Especially in the medical field, each diagnostic offered by a model should be argued and explained; the Explainable Artificial Intelligence (XAI) [84] domain is at the beginning of development. DANN models perform better than classic ML models (e.g., SVM, RF, and NB), yet they have the disadvantage of a black-box nature; therefore, a trade-off between interpretability and performance is still necessary [85]. An application based on a DANN model, particularly in the medical field, should have the following characteristics: fairness (i.e., ensure that the model does not discriminate), accountability (i.e., decide who is responsible for the decision), transparency (i.e., interpretability and understandability of the model's decisions), privacy, and robustness [86]. Meeting these criteria can pose challenges in situations where data are scarce or originate exclusively from a specific category, such as being restricted to more-developed countries. Lastly, future research should concentrate on refining these linguistic markers and models to support real-time diagnostics, early intervention, and treatment monitoring for neuropsychiatric disorders. Validation studies in clinical settings are necessary to evaluate the reliability and generalizability of these linguistic models.

The generalizability of the presented research findings can represent a potential challenge to the use of AI in the medical field, especially in such subjective areas as mental or psychiatric illnesses. A wrong generalization can be generated from the beginning by using data limited only to certain categories of people. For example, a study [87] performed on 94 adults demonstrated the link between depression and demographics and clinical and personality traits. Larøi et al. [88] studied the influence that culture (i.e., multiple factors such as religion and political beliefs) has on hallucinations. Taking these into account, the existence of a heterogeneous dataset that includes as many different elements as possible would contribute to the discovery of linguistic symptoms that are as general as possible. Another perspective from which this problem can be viewed is that of the model. As mentioned, the models with the best performance are based on DANN; these types of models are prone to unreliable results based on incorrect criteria if the training data are biased.

A fundamental theoretical problem of DANNs, which are now considered the best approaches for NLP and were used in the research discussed herein, is that transformers and neural networks, in general, are based on an empiricist paradigm. It should be mentioned that to obtain the best results, there is a need to integrate empiricist with nativist perspectives, the latter being used in symbolic, knowledge-based AI. These two paradigms correspond, in fact, to the two main, opposing philosophical schools of thought that have Aristotle and Plato as parents, with the latter being also advocated by Chomsky [13].

6. Conclusions

This survey demonstrates the potential of NLP for identifying linguistic patterns related to neuropsychiatric disorders. Advanced methods have identified specific linguistic traits and offer promising results for the early recognition and treatment of these disorders. The identified markers (e.g., specific emotions and verb tenses) linked to conditions such as depression, dementia, or hallucinations represent cues that are sometimes undiscoverable by conventional diagnosis methods. This interdisciplinary field that combines linguistic

analysis, medical science, AI, and multimodal approaches offers a promising direction for future research and practical applications and will potentially revolutionize early detection, treatment, and care for neuropsychiatric disorders. However, despite these advancements, future efforts are needed to enhance AI model accuracy and interpretability. At last, but of course not at least, it should be mentioned that the very important ethical aspects need be permanently considered, and it should also be taken into account that AI ethics is now a major subject of discussion, research, and regulation [89–91].

Author Contributions: Conceptualization, I.-R.Z. and S.T.-M.; methodology, I.-R.Z. and S.T.-M.; validation, S.T.-M.; investigation, I.-R.Z. and S.T.-M.; resources, I.-R.Z. and S.T.-M.; data curation, I.-R.Z. and S.T.-M.; writing—original draft preparation, I.-R.Z.; writing—review and editing, S.T.-M.; supervision, S.T.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: We would like to thank the authors of all datasets described in this paper for making the data available to the community for research purposes.

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

References

1. Luchini, C.; Pea, A.; Scarpa, A. Artificial intelligence in oncology: Current applications and future perspectives. *Br. J. Cancer* **2021**, *126*, 4–9. [CrossRef]
2. Gupta, M.; Kunal, S.; Mp, G.; Gupta, A.; Yadav, R.K. Artificial intelligence in cardiology: The past, present and future. *Indian Heart J.* **2022**, *74*, 265–269. [CrossRef] [PubMed]
3. Giorgini, F.A.; Di Dalmazi, G.; Diciotti, S. Artificial intelligence in endocrinology: A comprehensive review. Journal of Endocrinological Investigation. *J. Endocrinol. Investig.* **2023** . [CrossRef]
4. Zhong, Y.; Chen, Y.; Zhang, Y.; Lyu, Y.; Yin, J.; Yujun, G. The Artificial intelligence large language models and neuropsychiatry practice and research ethic. *Asian J. Psychiatry* **2023**, *84*, 103577. [CrossRef] [PubMed]
5. Rainey, S.; Erden, Y.J. Correcting the brain? the convergence of neuroscience, neurotechnology, psychiatry, and artificial intelligence. *Sci. Eng. Ethics* **2020**, *26*, 2439–2454. [CrossRef] [PubMed]
6. World Health Organization (WHO). Available online: https://platform.who.int (accessed on 10 November 2023).
7. Leung, C.M.C.; Ho, M.K.; Bharwani, A.A.; Cogo-Moreira, H.; Wang, Y.; Chow, M.S.C.; Fan, X.; Galea, S.; Leung, G.M.; Ni, M.Y. Mental disorders following COVID-19 and other epidemics: A systematic review and meta-analysis. *Transl. Psychiatry* **2022**, *12*, 205. [CrossRef]
8. Solmi, M.; Radua, J.; Olivola, M.; Croce, E.; Soardo, L.; de Pablo, G.S.; Shin, J.I.; Kirkbride, J.B.; Jones, P.; Kim, J.H.; et al. Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies. *Mol. Psychiatry* **2021**, *27*, 281–295. [CrossRef]
9. Belousov, M.; Dinev, M.; Morris, R.;Berry, N.; Bucci, S.; Nenadic, G. Mining Auditory Hallucinations from Unsolicited Twitter Posts. 2016. Available online: https://ep.liu.se/en/conference-article.aspx?series=&issue=128&Article_No=5 (accessed on 29 January 2024).
10. Masrani, V.; Murray, G.; Field, T.; Carenini, G. Detecting dementia through retrospective analysis of routine blog posts by bloggers with dementia. *ACL Anthol.* **2017**. Available online: https://aclanthology.org/W17-2329/ (accessed on 29 January 2024).
11. Yoon, J.; Kang, C.; Kim, S.; Han, J. D-VLog: Multimodal vlog dataset for Depression Detection. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 12226–12234. [CrossRef]
12. Vaswani, A.; Shazeer, N.; Parmar, N. Attention Is All you Need. Part of Advances in Neural Information Processing Systems 30. 2017. Available online: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (accessed on 29 January 2024).
13. Ranaldi, L.; Pucci, G. Knowing Knowledge: Epistemological Study of knowledge in transformers. *Appl. Sci.* **2023**, *13*, 677. [CrossRef] [CrossRef]
14. Yokoi, R.; Eguchi, Y.; Fujita, T.; Nakayachi, K. Artificial Intelligence Is Trusted Less than a Doctor in Medical Treatment Decisions: Influence of Perceived Care and Value Similarity. *Int. J. Hum.–Comput. Interact.* **2020**, *37*, 981–990. [CrossRef]
15. Kozhemyakova, E.A.; Petukhova, M.E.; Simulina, S.; Ivanova, A.M.; Zakharova, A. Linguistic markers of native speakers. In Proceedings of the International Conference "Topical Problems of Philology and Didactics: Interdisciplinary Approach in Humanities and Social Sciences" (TPHD 2018), 2019 . [CrossRef]

16. Rawte, V.; Chakraborty, S.; Pathak, A.; Sarkar, A.; Tonmoy, S.M.T.I.; Chadha, A.; Sheth, A.P.; Das, A. The troubling emergence of hallucination in large language models—An extensive definition, quantification, and prescriptive remediations. *arXiv* **2023**, arXiv:abs/2310.04988.
17. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR) (5TH ED)*; British Library; American Psychiatric Association: Washington, DC, USA, 2013.
18. Widiger, T.A.; Oltmanns, J.R. Neuroticism is a fundamental domain of personality with enormous public health implications. *World Psychiatry Off. J. World Psychiatr. Assoc. (WPA)* **2017**, *16*, 144–145. [CrossRef]
19. Dementia Australia. Available online: https://www.dementia.org.au (accessed on 11 November 2023).
20. Alzheimer's Society. Available online: https://www.alzheimers.org.uk (accessed on 11 November 2023).
21. Dementia UK. Available online: https://www.dementiauk.org (accessed on 11 November 2023).
22. Alzheimer's Association. Available online: https://www.alz.org (accessed on 11 November 2023).
23. Cleveland Clinic. Available online: https://my.clevelandclinic.org (accessed on 11 November 2023).
24. Linszen, M.M.J.; de Boer, J.N.; Schutte, M.J.L.; Begemann, M.J.H.; de Vries, J.; Koops, S.; Blom, R.E.; Bohlken, M.M.; Heringa, S.M.; Blom, J.D.; et al. Occurrence and phenomenology of hallucinations in the general population: A large online survey. *Schizophrenia* **2022**, *8*, 41. [CrossRef]
25. Tausczik, Y.R.; Pennebaker, J.W. The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.* **2019**, *29*, 24–54. [CrossRef]
26. Burkhardt, H.; Pullmann, M.; Hull, T.; Aren, P.; Cohen, T. Comparing emotion feature extraction approaches for predicting depression and anxiety. In Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology, Seattle, USA, July 2022 . [CrossRef]
27. Tadesse, M.M.; Lin, H.; Xu, B.; Yang, L. Detection of depression-related posts in Reddit social media forum. *IEEE Access* **2019**, *7*, 44883–44893. [CrossRef]
28. Wolohan, J.; Hiraga, M.; Mukherjee, A.; Sayyed, Z.A.; Millard, M. Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with NLP. In Proceedings of the First International Workshop on Language Cognition and Computational Models, Santa Fe, NM, USA, August 2018 ; pp. 11–21. Available online: https://aclanthology.org/W18-4102/ (accessed on 29 January 2024).
29. Baccianella, S.; Esuli, A.; Sebastiani, F. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *ACL Anthol.* **2010**, *10*, 2200–2204. https://aclanthology.org/L10-1531/.
30. Hutto, C.; Gilbert, E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proc. Int. AAAI Conf. Web Soc. Media* **2014**, *8*, 216–225. [CrossRef]
31. Titla-Tlatelpa, J.d.J.; Ortega-Mendoza, R.M.; Montes-y-Gómez, M.; Villaseñor-Pineda, L. A profile-based sentiment-aware approach for depression detection in social media. *EPJ Data Sci.* **2021**, *10*, 54. [CrossRef]
32. Burkhardt, H.A.; Alexopoulos, G.S.; Pullmann, M.D.; Hull, T.D.; Areán, P.A.; Cohen, T. *Behavioral Activation and Depression Symptomatology: Longitudinal Assessment of Linguistic Indicators in Text-Based Therapy Sessions (Preprint)*; JMIR Publications Inc.: Toronto, ON, Canada, 2021. [CrossRef]
33. Jones, K.S. A statistical interpretation of term specificity and its application in retrieval. *J. Doc.* **1972**, *28*, 11–21. [CrossRef]
34. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
35. Higuch, K. A Two-Step Approach to Quantitative Content Analysis: KH Coder Tutorial Using Anne of Green Gables (Part II). *Ritsumeikan Soc. Sci. Rev.* **2017**, *52*, 77–91.
36. Toutanova, K.; Manning, C.D. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In Proceedings of the 2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics, Hong Kong, China, 7–8 October 2000. [CrossRef]
37. Chang, A.X.; Manning, C.D. SUTIME: A library for recognizing and normalizing time expressions. In Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey, 21–27 May 2012; pp. 3735–3740.
38. Wegemer, C.M. Selflessness, depression, and neuroticism: An interactionist perspective on the effects of self-transcendence, perspective-taking, and materialism. *Front. Psychol.* **2020**, *11*, 523950. [CrossRef] [PubMed]
39. Tølbøll, K.B. Linguistic features in depression: A meta-analysis. *J. Lang.-Work.–Sprogvidenskabeligt Stud.* **2019**, *4*, 39–59.
40. Gratch, J.; Artstein, R.; Lucas, G.; Stratou, G.; Scherer, S.; Nazarian, A.; Wood, R.; Boberg, J.; DeVault, D.; Marsella, S.; et al. The Distress Analysis Interview Corpus of human and computer interviews. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, May 2014; pp. 3123–3128. Available online: https://aclanthology.org/L14-1421/ (accessed on 29 January 2024).
41. Zaman, I.-R.; Trausan-Matu, S.; Rebedea, T. Analysis of medical conversations for the detection of depression. In Proceedings of the 19th International Conference on Human-Computer Interaction—RoCHI 2023, 2023; pp. 15–22. Available online: http://rochi.utcluj.ro/articole/11/RoCHI2023-Zaman.pdf (accessed on 29 January 2024).
42. Robertson, C.; Carney, J.; Trudell, S. Language about the future on social media as a novel marker of anxiety and depression: A big-data and experimental analysis. *Curr. Res. Behav. Sci.* **2023**, *4*, 100104. [CrossRef] [PubMed]
43. Demszky, D.; Movshovitz-Attias, D.; Ko, J.; Cowen, A.; Nemade, G.; Ravi, S. GoEmotions: A dataset of fine-grained emotions. *arXiv* **2020**, arXiv:2005.00547.

44. Al-Mosaiwi, M.; Johnstone, T. In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clin. Psychol. Sci.* **2018**, *6*, 529–542. [CrossRef] [PubMed]
45. Lanzi, A.M.; Saylor, A.K.; Fromm, D.; Liu, H.;MacWhinney, B.; Cohen, M.L. DementiaBank: Theoretical rationale, protocol, and illustrative analyses. *Am. J. Speech-Lang. Pathol.* **2023**, *32*, 426–438. [CrossRef] [PubMed]
46. Labbé, C.; König, A.; Lindsay, H.; Linz, N.; Tröger, J.; Robert, P. Dementia vs. Depression: New methods for differential diagnosis using automatic speech analysis. *Alzheimer's Dement. J. Alzheimer's Assoc.* **2022**, *18*, e064486. [CrossRef]
47. Almor, A.; Aronoff, J.M.; MacDonald, M.C.; Gonnerman, L.M.; Kempler, D.; Hintiryan, H.; Hayes, U.L.; Arunachalam, S.; Andersen, E.S. A common mechanism in verb and noun naming deficits in Alzheimer's patients. *Brain Lang.* **2009**, *111*, 8–19. [CrossRef]
48. Kim, M.; Thompson, C.K. Verb deficits in Alzheimer's disease and agrammatism: Implications for lexical organization. *Brain Lang.* **2004**, *88*, 1–20. [CrossRef]
49. Guerrero-Cristancho, J.; Vasquez, J.C.; Orozco, J.R. Word-Embeddings and grammar features to detect language disorders in alzheimer's disease patients. *Inst. Tecnol. Metrop.* **2020**, *23*, 63–75. Available online: https://www.redalyc.org/journal/3442/344262603030/html/ (accessed on 29 January 2024). [CrossRef]
50. Eyigoz, E.; Mathur, S.; Santamaria, M.; Cecchi, G.; Naylor, M. Linguistic markers predict onset of Alzheimer's disease. *EClinicalMedicine* **2020**, *28*, 100583. [CrossRef]
51. Martin, A.; Fedio, P. Word production and comprehension in Alzheimer's diseáse: The breakdown of semantic knowledge. *Brain Lang.* **1983**, *19*, 124–141. [CrossRef] [PubMed]
52. Appell, J.; Kertesz, A.; Fisman, M. A study of language functioning in Alzheimer patients. *Brain Lang.* **1982**, *17*, 73–91. [CrossRef] [PubMed]
53. Karlekar, S.; Niu, T.; Bansal, M. Detecting linguistic characteristics of Alzheimer's dementia by interpreting neural models. *ACL Anthol.* **2018**. Available online: https://aclanthology.org/N18-2110/ (accessed on 29 January 2024).
54. Koyama, A.; O'Brien, J.; Weuve, J.; Blacker, D.; Metti, A.L.; Yaffe, K. The role of peripheral inflammatory markers in dementia and Alzheimer's disease: A meta-analysis. *J. Gerontol. Ser. Biol. Sci. Med. Sci.* **2012**, *68*, 433–440. [CrossRef]
55. Ewers, M.; Sperling, R.A.; Klunk, W.E.; Weiner, M.W.; Hampel, H. Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia. *Trends Neurosci.* **2011**, *34*, 430–442. [CrossRef]
56. Mirjafari, S.; Nepal, S.; Wang, W.; Campbell, A.T. Using mobile data and deep models to assess auditory verbal hallucinations. *arXiv* **2023**, arXiv:2304.11049.
57. Speth, C.; Speth, J. A new measure of hallucinatory states and a discussion of REM sleep dreaming as a virtual laboratory for the rehearsal of embodied cognition. *Cogn. Sci.* **2017**, *42*, 311–333. [CrossRef]
58. de Boer, J.N.; Linszen, M.M.J.; de Vries, J.; Schutte, M.J.L.; Begemann, M.J.H.; Heringa, S.M.; Bohlken, M.M.; Hugdahl, K.; Aleman, A.;Wijnen, F.N.K.; et al. Auditory hallucinations, top-down processing and language perception: A general population study. *Psychol. Med.* **2019**, *49*, 2772–2780. [CrossRef]
59. Viswanath, B.; Chaturvedi, S.K. Cultural aspects of major mental disorders: A critical review from an Indian perspective. *Indian J. Psychol. Med.* **2012**, *34*, 306–312. [CrossRef]
60. Irving, J.; Colling, C.; Shetty, H.;Pritchard, M.; Stewart, R.; Fusar-Poli, P.; McGuire, P.; Patel, R. Gender differences in clinical presentation and illicit substance use during first episode psychosis: A natural language processing, electronic case register study. *BMJ Open* **2021**, *11*, e042949. [CrossRef] [PubMed]
61. de Boer, J.N.; Corona Hernández, H.; Gerritse, F.; Brederoo, S.G.; Wijnen, F.N.K.; Sommer, I.E. Negative content in auditory verbal hallucinations: A natural language processing approach. *Cogn. Neuropsychiatry* **2021**, *27*, 139–149. [CrossRef]
62. Testoni, A.; Bernardi, R. "I've seen things you people wouldn't believe": Hallucinating entities in guesswhat?! In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Student Research Workshop, Online, August 2021. [CrossRef]
63. Dziri, N.; Milton, S.; Yu, M.; Zaiane, O.; Reddy, S. On the Origin of Hallucinations in Conversational Models: Is it the Datasets or the Models? In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, DC, USA, 10–15 July 2022. [CrossRef]
64. Rohrbach, A.; Hendricks, L.A.; Burns, K.; Darrell, T.; Saenko, K. Object hallucination in image captioning. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018. [CrossRef]
65. Zhang, M.; Press, O.; Merrill, W.; Liu, A.; Smith, N.A. How language model hallucinations can snowball. *arXiv* **2023**, arXiv:2305.13534. https://arxiv.org/abs/2305.13534.
66. OpenAI. ChatGPT 2023. Available online: https://chat.openai.com/chat (accessed on 29 January 2024).
67. Stiles, W.B. Verbal response modes taxonomy. In *The Cambridge Handbook of Group Interaction Analysis*; Cambridge University Press: Cambridge, UK, 2019; pp. 630–638. [CrossRef]
68. Alkaissi, H.; McFarlane, S.I. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. 2023. Available online: https://pubmed.ncbi.nlm.nih.gov/36811129/ (accessed on 29 January 2024).
69. Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y.J.; Madotto, A.; Fung, P. Survey of hallucination in natural language generation. *ACM Comput. Surv.* **2023**, *55*, 1–38. [CrossRef]

70. Mündler, N.; He, J.; Jenko, S.; Vechev, M. Self-contradictory hallucinations of large language models: Evaluation, detection and mitigation. *arXiv* **2023**, arXiv:2305.15852.
71. Shen, G.; Jia, J.; Nie, L.; Feng, F.; Zhang, C.; Hu, T.; Chua, T.-S.; Zhu, W. Depression detection via harvesting social media: A multimodal dictionary learning solution. In Proceedings of the IJCAI, Melbourne, Canada, 2017; pp. 3838–3844. Available online: https://www.ijcai.org/proceedings/2017/536 (accessed on 29 January 2024).
72. Megan, E.; Wittenborn, A.; Bogen, K.; McCauley, H. #MyDepressionLooksLike: Examining public discourse about depression on twitter. *JMIR Ment. Health* **2017,** *4*, e8141. [CrossRef]
73. Cohan, A.; Desmet, B.; Yates, A.; Soldaini, L.; MacAvaney, S.; Goharian, N. SMHD: A large-scale resource for exploring online language usage for multiple mental health conditions. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, August 2018; pp. 1485–1497. Available online: https://aclanthology.org/C18-1126/ (accessed on 29 January 2024).
74. Becker, J.T.; Boiler, F.; Lopez, O.; Saxton, J.; McGonigle, K.L. The Natural History of Alzheimer's Disease: Description of study cohort and accuracy of diagnosis. *Arch. Neurol.* **1994**, *51*, 585–594. [CrossRef] [PubMed]
75. Xue, C.; Karjadi, C.; Paschalidis, I.C.; Au, R.; Kolachalama, V.B. Detection of dementia on voice recordings using deep learning: A Framingham Heart Study. *Alzheimer's Res. Ther.* **2021**, *13*, 1–15. [CrossRef]
76. Azizi, M.; Jamali, A.A.; Spiteri, R. Identifying Tweets Relevant to Dementia and COVID-19: A Machine Learning Approach (Preprint). *PREPRINT-SSRN* **2023**. Available online: https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/pt/ppzbmed-10.2139.ssrn.4458777 (accessed on 29 January 2024).
77. Liu, T.; Zhang, Y.; Brockett, C.; Mao, Y.; Sui, Z.; Chen, W.; Dolan, W.B. A token-level reference-free hallucination detection benchmark for free-form text generation. *ACL Anthol.* **2022**. Available online: hhttps://aclanthology.org/2022.acl-long.464/ (accessed on 29 January 2024).
78. Guo, M.; Dai, Z.; Vrandečić, D.; Al-Rfou', R. Wiki-40B: Multilingual language model dataset. In Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France, May 2020; pp. 2440–2452. Available online: https://aclanthology.org/2020.lrec-1.297/ (accessed on 29 January 2024).
79. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
80. Ghosh, C.C.; McVicar, D.; Davidson, G.; Shannon, C.; Armour, C. Exploring the associations between auditory hallucinations and psychopathological experiences in 10,933 patient narratives: Moving beyond diagnostic categories and surveys. *BMC Psychiatry* **2023**, *23*, 1–10. [CrossRef]
81. Ghosh, C.C.; McVicar, D.; Davidson, G.; Shannon, C. Measuring diagnostic heterogeneity using text-mining of the lived experiences of patients. *BMC Psychiatry* **2021**, *21*, 1–12. [CrossRef] [PubMed]
82. Pal, A.; Umapathi, L.K.; Sankarasubbu, M. Med-HALT: Medical domain hallucination test for large language models. *arXiv* **2023**, arXiv:2307.15343.
83. Gerke, S.; Minssen, T.; Cohen, G. Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare. 2020. Available online: https://www.sciencedirect.com/science/article/pii/B9780128184387000125 (accessed on 29 January 2024).
84. Gohel, P.; Singh, P.; Mohanty, M. Explainable AI: Current status and future directions. *arXiv* **2021**, arXiv:2107.07045.
85. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *arXiv* **2019**, arXiv:1910.10045.
86. Wei, W.; Landay, J. CS 335: Fair, Accountable, and Transparent (FAccT) Deep Learning, Stanford University, April 2020. Available online: https://hci.stanford.edu/courses/cs335/2020/sp/lec1.pdf (accessed on 29 January 2024).
87. Enns, M.W.; Larsen, D.K.; Cox, B.J. Discrepancies between self and observer ratings of depression. *J. Affect. Disord.* **2000**, *60*, 33–41. [CrossRef]
88. Larøi, F.; Luhrmann, T.M.; Bell, V.; Christian, W.A.; Deshpande, S.N.; Fernyhough, C.; Jenkins, J.H.; Woods, A. Culture and Hallucinations: Overview and Future Directions. *Schizophr. Bull.* **2014**, *40*, S213–S220. [CrossRef]
89. Council of the EU Artificial Intelligence Act: Council and Parliament Strike a Deal on the First Rules for AI in the World. Available online: https://www.consilium.europa.eu (accessed on 28 January 2024).
90. AI HLEG. Ethics Guidelines for Trustworthy AI. Available online: https://ec.europa.eu (accessed on 28 January 2024).
91. European Parliament. EU Guidelines on Ethics in Artificial Intelligence: Context and Implementation. Available online: https://www.europarl.europa.eu (accessed on 28 January 2024).