Turkish Journal of Psychiatry 2024;XX(X):XX
https://doi.org/10.5080/u27604

Improving Psychiatry Services with Artificial Intelligence: Opportunities and Challenges

Muhammed BALLI[1], Aslı ERCAN DOĞAN[2], Hale YAPICI ESER[3]

ABSTRACT

Mental disorders are a critical global public health problem due to their increasing prevalence, rising costs, and significant economic burden. Despite efforts to increase the mental health workforce in Türkiye, there is a significant shortage of psychiatrists, limiting the quality and accessibility of mental health services. This review examines the potential of artificial intelligence (AI), especially large language models, to transform psychiatric care in the world and in Türkiye. AI technologies, including machine learning and deep learning, offer innovative solutions for the diagnosis, personalization of treatment, and monitoring of mental disorders using a variety of data sources, such as speech patterns, neuroimaging, and behavioral measures. Although AI has shown promising capabilities in improving diagnostic accuracy and access to mental health services, challenges such as algorithmic biases, data privacy concerns, ethical implications, and the confabulation phenomenon of large language models prevent the full implementation of AI in practice. The review highlights the need for interdisciplinary collaboration to develop culturally and linguistically adapted AI tools, particularly in the Turkish context, and suggests strategies such as fine-tuning, retrieval-augmented generation, and reinforcement learning from human feedback to increase AI reliability. Advances suggest that AI can improve mental health care by increasing diagnostic accuracy and accessibility while preserving the essential human elements of medical care. Current limitations need to be addressed through rigorous research and ethical frameworks for effective and equitable integration of AI into mental health care.

Keywords: artificial intelligence, health, large language model, machine learning, psychiatry

ECONOMIC BURDEN AND WORKFORCE LIMITATIONS IN PSYCHIATRY: INSIGHTS FROM GLOBAL AND TURKISH PERSPECTIVES

Mental health disorders represent a critical global public health burden due to their prevalence and economic costs. A study published in 2022, analysing data from 204 countries on 12 mental disorders between 1990 and 2019, found that global disability-adjusted life years (DALYs) due to mental disorders increased from 80.8 million to 125.3 million, with the share of DALYs attributed to mental disorders increasing from 3.1% to 4.9% (GBD 2019 Mental Disorders Collaborators 2022). This finding shows that the burden of mental disorders is increasing on a global scale. The economic disadvantages associated with mental disorders, an increase in risk factors such as social exclusion, and limited access to protective factors like education contribute to the further increase in the prevalence of mental disorders (Niemeyer et al. 2019, Heinz et al. 2020). According to a 2021 study on the economic burden of major depressive disorder (MDD) in the United States, the cost associated with MDD increased by 37.9% from 2010 to 2018, from $236.6 billion to $326.2 billion. During this period, the number of adults diagnosed with MDD increased from 15.5 million to 17.5 million; however, the proportion of patients receiving treatment decreased in 2018 compared to 2010 (Greenberg et al. 2021). This shows that although the frequency of diagnoses such as MDD has increased, access to effective treatment has not increased at the same rate. As mental health disorders are associated with serious outcomes such as disability and death, preventive measures and accessible services are crucial (Martin-Carrasco et al. 2016).

Received: 07.10.2024, **Accepted:** 27.11.2024, **Available Online Date:** 06.12.2024

[1]PhD Candidate, Koç University, Graduate School of Health Sciences, Istanbul; [2]Psychiatrist, [3]Assoc. Prof., Koç University School of Medicine, Department of Psychiatry, Istanbul, Turkey.

e-mail: hyapici@ku.edu.tr

According to the National Mental Health Action Plan published in 2021, while there were 2.2 psychiatry specialists per 100,000 people in Türkiye in 2011, this number increased to 3.43 in 2020; the number of child and adolescent mental health and diseases specialists was 1.63 per 100,000 people (Ulusal Ruh Sağlığı Eylem Planı 2021). Despite efforts to increase the number of mental health professionals, the insufficient number of mental health professionals continues and this situation negatively affects the quality and accessibility of mental health services. According to the 2016 report of the Psychiatric Association of Türkiye, mental health centres lack up-to-date technological facilities and psychiatry specialists in public hospitals examine up to 80 patients a day. Appointment durations are limited to 1-2 minutes, which falls significantly below the World Health Organization's recommended minimum of 20 minutes. Consequently, it is not always possible to prioritize acute and high-risk patients (Türkiye Psikiyatri Derneği 2016). In addition, increasing stigmatisation at the societal and individual level is a significant barrier to seeking mental health support (Zweifel 2021) and, according to the same report of the Psychiatric Association of Türkiye, it pushes patients to seek help from non-specialists in the field of psychiatry, causing their current health conditions to worsen and thus further aggravating the burden on the mental health system (Türkiye Psikiyatri Derneği 2016). Patients' access to psychiatry services is already limited due to the limited number of specialists and time constraints in public hospitals. Additionally, the exclusion of psychiatric services from many private insurance plans further restricts access to care. Although efforts are being made to increase the number of mental health professionals, the training and supervision of a psychiatrist usually takes 4 to 5 years. This process involves the transfer of knowledge and skills from supervisors/lecturers to trainees and involves direct observation, assessment and follow-up of patient interviews and case files. Currently this training is highly dependent on human input, making it both time consuming and difficult to standardise training across institutions. In addition to clinical practice and training, there is also a need for more precise standardisation, classification and documentation of psychopathological symptoms and findings in order to create larger datasets for research. For these standardised assessments, common consultants and evaluators across centres are needed. Conducting and standardising all these assessments requires time and effort.

EVOLUTION OF ARTIFICIAL INTELLIGENCE AND INTEGRATION WITH MENTAL HEALTH SERVICES

Artificial intelligence (AI) is a multidisciplinary field of computer science that focuses on performing tasks where human intelligence is required. Endowing machines with the ability to perform tasks such as perception, reasoning, learning, and decision-making—tasks that humans can easily accomplish—falls within the scope of AI and its sub-disciplines. The beginning of AI studies in the modern sense dates back to 1950, when Alan Turing published his article Computing Machinery and Intelligence (Turing 2009). The term AI was coined in 1956 at the Dartmouth conference, which brought together researchers working in this field. Despite the optimistic atmosphere at the beginning, AI did not reach the expected development in the 1970s and 1980s due to the lack of sufficient computing power of computers (Delipetrev 2020). Exponentially increasing data and processing resources, together with machine learning techniques such as artificial neural networks and deep learning, revitalised the field of AI in the late 1990s (Pastur-Romay et al. 2016). IBM-developed Deep Blue's defeat of chess master Garry Kasparov in 1996 is one of the symbolic moments of this re-emergence (Delipetrev 2020). In the years after 2020, the exponential growth of AI has led to increased debate on AI safety and usefulness. It has been suggested that AI systems using machine learning algorithms and artificial neural networks can detect diseases such as cancer, heart disease and neurological disorders with high sensitivity by analysing complex data such as medical imaging and genetic information (Jiang et al. 2017). It has also been suggested that AI can accelerate drug discovery by predicting molecular interactions and identifying potential drug candidates (Gupta et al. 2021). AI can assist in adjusting treatment plans in real-time on an individual basis by using algorithms developed from information about treatment responses throughout the course of the disease.This dynamic method aims to increase the likelihood of a more effective treatment compared to conventional treatment approaches that still involve trial and error (Espejo et al. 2023).

The application of AI in mental health services started in the mid-20th century with preliminary studies on robots designed to mimic human thought processes (Olawade et al. 2024). The potential of this technology in mental health was first demonstrated by Joseph Weizenbaum in the 1960s when he developed a chatbot that mimics therapeutic conversations (Basset 2019). As stated in Ayhan's (2023) article, despite concerns about ethical use and reliability, AI in mental health is expected to integrate large amounts of data such as genetic data, magnetic resonance imaging results, daily experiences, data from wearable devices that track physical changes, and behavioural phenotypes in order to provide better diagnosis, follow-up, risk assessment and prevention. (Ayhan 2023). These AI-enabled tools can provide stigma-free, accessible and measurable mental health services and provide a solution that allows patients to receive support outside of traditional therapy and psychiatric examination

sessions. This change may not only increase the effectiveness of interviews, but also enable mental health resources to reach those in need more widely (Omarov et al. 2023). The impact of AI can also extend to the post-treatment phase, where continuous monitoring of fluctuations in mental health dimensions, whether they improve, worsen or remain stable, can play a critical role in monitoring patient behaviour and outcomes. Regular monitoring and processing of mental health dimensions (functioning, sleep, appetite, psychomotor movement, etc.) can enable early detection of relapses and allow for timely interventions that can significantly improve treatment effectiveness and outcomes through comprehensive assessment of patients' mental health status (Zlatintsi et al. 2022, Hickey et al. 2021, Krysta et al. 2024).

UNDERSTANDING AI PROCESSES: DATA MANAGEMENT, MACHINE LEARNING AND DEEP LEARNING

AI works through a series of processes that enable machines to perform tasks. This process starts with 'data collection', where a large amount of information is gathered from various sources such as images, texts or sensor inputs. This data is then cleaned and organised so that the AI system can effectively extract meaningful patterns and relationships (Roh et al. 2019). AI then uses algorithms, a set of mathematical instructions, to analyse the data and identify patterns or relationships (Basu et al. 2010). For example, in image recognition, an AI model can gain the ability to distinguish cats from dogs by examining thousands of labelled images (Elgendy 2020). This learning phase, often referred to as training, involves adjusting the algorithm's variables to improve accuracy. Once trained, the AI system can make predictions or decisions about new data that it has not encountered before, based on the model it has built from previous data (Baduge et al. 2022).

At this point, it would be useful to explain the concepts of Machine Learning (ML) and Deep Learning (DL), which are branches of AI. ML focuses on the development of algorithms and statistical models that enable computer systems to improve their performance on a given task through experience, without explicit programming. It is necessary to use large amounts of data to train models that can make predictions, recognise patterns or make decisions with minimal human intervention (Jordan and Mitchell 2015). An example of such decision-making is when autonomous driving vehicles stop at a red light. A common approach to prediction in ML is to use linear regression, which attempts to model the relationship between input features (such as house size, number of bedrooms, location) and the target variable (house price). The algorithm learns by minimising the difference between its predictions and actual house prices in a training dataset. It adjusts the weights assigned to each attribute to find the line of best fit. Once trained, it can predict the model price considering the features of a new house (Ghosalkar et al. 2018). In the field of machine learning, both linear and non-linear algorithms play crucial roles in addressing various prediction tasks. While linear regression offers a simple approach for problems with linear relationships, many real-world scenarios require more complex, non-linear models. Random forests and neural networks are nonlinear ML algorithms that can capture complex relationships in data. Apart from these, another commonly used nonlinear technique is support vector machines (SVM). SVM can find discriminative hyperplanes by moving the data to a higher dimensional space for classification or regression (Greener et al. 2022). The choice between linear and non-linear algorithms usually depends on the data structure and the problem to be solved (Suthaharan et al. 2016). DL (DL) is a sub-branch of ML that utilises multilayer artificial neural networks. While traditional ML algorithms usually require manually determined features, DL models can automatically extract these representations from data. This feature enables DL to work with unstructured data such as images, audio and text. The most important difference of DL is that it can process raw input data through a multi-layered structure. These layers extract increasingly more abstract features at each step. For example, in an image recognition system, while the first layers detect edges, deeper layers can recognise objects and more complex shapes (LeCun et al. 2015).

This hierarchical learning enables the DL to handle more complex patterns and achieve superior performance in tasks such as speech recognition and computer vision (Mikolov et al. 2011, Krizhevsky et al. 2012). However, compared to simpler ML algorithms, it usually requires larger datasets and more computational resources, and its decision-making process is less interpretable due to its complex network structure (Hestness et al. 2019, Lipton 2018). Despite these challenges, the capacity of DL to learn complex patterns has led to advances in AI applications in various fields. The features related to ML and DL are summarised in Figure 1.

MACHINE LEARNING IN PSYCHIATRY: DIAGNOSTIC POTENTIAL, LIMITATIONS AND TRANSDIAGNOSTIC FRAMEWORKS

Psychiatry, which focuses on understanding the natural processes of emotion, behaviour and thought in humans and their disorders, can benefit from AI's ability to analyse large and multifaceted data sets such as speech content (Joshi et al. 2022), scales, brain imaging, blood and digital biomarkers, audio data (Huang et al. 2021), mood (Barzilay et al. 2019), social media data, oculo-metric system dynamics, eye movements for attention assessment and peripheral physiological signals (Saxena et al. 2021, Morar et al. 2020, Kalmady 2019, Giorgi 2021) to reveal complex patterns.

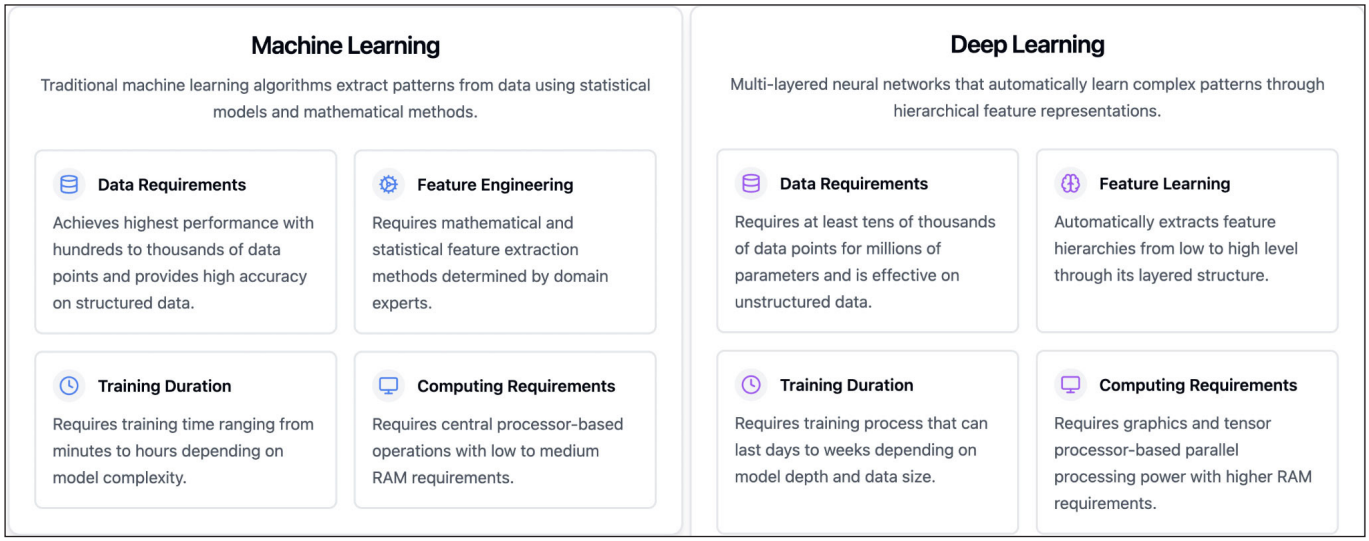

Figure 1. Comparison of machine learning and deep learning.
Note: RAM (Random Access Memory) is a type of fast-access memory that a computer uses to store temporary data. In data-intensive operations such as machine learning and deep learning, RAM plays a critical role, especially during training and data processing procedures. CPU (Central Processing Unit) refers to the processor unit designed for general-purpose computing operations; GPU (Graphics Processing Unit) refers to the processor unit capable of parallel processing, optimized especially for deep learning operations such as matrix multiplications; TPU (Tensor Processing Unit) refers to the processor unit specifically developed by Google for deep learning applications, providing high performance in matrix operations.

Collectively, these methods contribute to the assessment and understanding of both brain and body activities and their interactions. For example, AI-enabled tools can assess speech patterns, facial expressions and social media activity to detect early signs of mental disorders such as depression, anxiety and bipolar disorder (Smrke et al. 2021, Su et al. 2020). A systematic review compiling ML models used to predict psychiatric disorder diagnoses from genetic data revealed that model performances varied across disorders. Schizophrenia (SchZ) models showed the widest range of Area Under Curve (AUC) accuracy (0.541-0.95), bipolar disorder predictions were less accurate (0.482-0.65), and anorexia nervosa models showed moderate accuracy (0.623-0.693). The diagnostic predictive accuracy of ML for Autism Spectrum Disorder (ASD) was found to vary (0.516-0.806). The algorithms used included SVMs, neural networks and random forests. The authors warn that the results carry a high risk of bias and potential overfit due to sample overlap and population structure issues (Bracher-Smith et al. 2021).

The review by Zhang et al. in 2020 examined DL applications for four brain disorders: Alzheimer's Disease (AD), Parkinson's Disease (PD), Autism Spectrum Disorder (ASD) and Schizophrenia (SchZ). The study reported impressive results. DL models were able to predict diseases with accuracy ranging from 90-99.3% for AD, 85-98.8% for PD, 70-91% for ASD and 70-98.09% for Schizophrenia compared to healthy controls. These diagnostic algorithms were developed using various DL architectures such as 3D-Convolutional Neural Networks (CNNs) and stacked autoencoders applied to neuroimaging data, including magnetic resonance imaging and functional magnetic resonance imaging (Zhange et al. 2020).

According to Squarcina et al. (2021), DL methods show promise in predicting response to depression treatment. A review of eight studies revealed that when DL models analysed various types of data (clinical, genetic, neuroimaging and electroencephalography), the power to predict treatment response with DL was better than traditional methods. The AUC values above 0.80 found in some studies indicate the promising performance of the models in integrating various data and identifying complex patterns that predict treatment outcomes.

There are also studies on suicide risk assessment with machine learning. However, according to a recent systematic review, its effectiveness in this area currently remains around 80% accuracy on average with survey-based data (Some et al. 2024). Most existing studies have achieved similar levels of accuracy for different psychiatric disorders using different measurement and diagnostic tools. Our research group's work in the peer-review phase with a sample of undergraduate and graduate students at Koç University did not exceed 80% accuracy in detecting suicidal/self-harm ideation using various ML algorithms with routine psychometric scales and inventories answered by people applying to the university's psychological counselling and therapy centre (Ballı et al. 2024, work in progress).

As mentioned above, research currently indicates that the capacity of AI technologies to reliably diagnose mental disorders and predict complex behaviours such as suicide remains limited. This shortcoming can be attributed to the combined effect of several factors, including the inherent complexity of human cognition and behaviour, and potential

human-induced biases that may arise during the design and editing of training data sets. The ML models used may have some methodological limitations. These include low sample size, lack of leave-one-out or nested cross-validation, problems in the design of training, testing and validation data sets, inappropriateness in the selection of the ML method and the variables used, etc. (Cearns et al. 2019). Future studies should improve their methods accordingly to achieve higher accuracy rates.

On the other hand, psychiatric diagnoses are currently classified according to diagnostic criteria determined in terms of symptom clusters, functionality and duration, as specified in systems such as the International Classification of Diseases-11 (ICD-11) or the Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5) (World Health Organization 2018, American Psychiatric Association 2013, Di Vincenzo 2023). However, these diagnoses are actually structures developed by humans and psychiatric disorders are highly co-occurring (Yapici Eser et al. 2018). This co-occurrence poses difficulties for ML methods and makes it difficult to create pure diagnostic groups and to distinguish between them. Furthermore, there are many common environmental, genetic, cellular and neuroanatomical pathways between psychiatric disorders (Scangos et al. 2023). Moreover, an individual may be diagnosed with more than one psychiatric diagnosis during his/her lifetime, which may impose limitations on ML approaches. The transdiagnostic approach in psychiatry focuses on the common processes underlying different mental disorders, rather than addressing each diagnosis separately. This approach recognises common mechanisms, views symptoms as a continuum, and allows for flexible interventions targeting core psychological processes (Fusar-Poli et al. 2019). The Research Domain Criteria (RDoC) introduced by the National Institute of Mental Health (NIMH) focuses on this transdiagnostic approach. The RDoC aims to establish a new framework for investigating mental disorders based on dimensions of observable behaviour and neurobiological measures (Cuthbert 2014). The ability of AI to combine different types of data, such as imaging, biochemistry, self-report and behavioural measures provided by clinician observation, and to capture non-linear relationships between these data, may make a cross-diagnostic approach possible and feasible and thus enable more accurate classification of psychiatric disorders. According to the review study by Bzdok et al. in 2018, ML has significant potential in advancing personalised medicine and high-precision psychiatry, particularly in treatment optimisation and dimensional approaches to psychopathology, rather than traditional classification systems. The authors suggest that predictive models can be used for individualised drug regulation, going beyond traditional trial-and-error methods. Representation learning algorithms are ML methods that automatically learn meaningful and useful representations (features) from raw data. Multimodal data analysis refers to the simultaneous processing of multiple data types, for example, text, image, audio, genetic or neuroimaging. Representation learning-based ML algorithms adapt to the RDoC framework by working with multimodal data. For example, by evaluating an individual's genetic data, brain images and behavioural measures together, it can discover hidden dimensions of psychopathology. This approach can help us understand biologically-based and clinically predictable disease processes, going beyond traditional diagnostic boundaries. At the same time, it may make it possible to reach a more biologically accurate and holistic psychiatric classification system (nosology), taking into account the heterogeneity in existing diagnostic categories and common features among disorders.

LEVERAGING NATURAL LANGUAGE PROCESSING AND LARGE LANGUAGE MODELS TO IMPROVE PSYCHIATRY SERVICES

Natural Language Processing (NLP) is a branch of AI that processes and analyses human language. In the medical field, it is used for many critical applications such as diagnosing, supporting clinical research, detecting drug interactions and making sense of patient data (Juhn and Liu 2020, Savova et al. 2019, Luo et al. 2017, Bhatnagar et al. 2022). NLP can extract rich information, including semantic and emotional content, from speech and texts, overcome the limitations of standard scales used in psychiatry, and make a more in-depth and personalised assessment by analysing patterns in individuals' language use (Kjell et al. 2023), and thus can make significant contributions to the development of medical decision support systems (Yang et al. 2022).

Large language models (LMMs), which are advanced NLP based AI systems designed to understand, produce and interact with a language in a human-like manner, entered the lives of the masses in November 2022, when OpenAI released ChatGPT. LMMs developed with DL techniques, especially transformer architectures, are trained on large and diverse textual datasets and can recognise complex linguistic patterns and contextual relationships. The transformer architecture is a DL architecture that effectively models contexts and complex relationships in long texts using the self-attention mechanism and is fundamental to the success of LLMs, with features that allow parallel processing (Vaswani et al. 2017). The process of developing LLMs involves training the model with large amounts of text data to predict the next word in a sequence, allowing the model to learn grammar, general information and various writing styles. This training allows LLMs to perform a wide range of tasks such as text completion, translation, summarisation, and speech interactions (Radford et al. 2019).

Multimodal LLMs extend the capabilities of traditional LLMs by integrating and processing multiple data types such as text, images, and audio within a single framework (Wu et al. 2023). This development enables these models to understand and generate content that combines different modalities, increasing their ability to perform tasks such as image captioning, video analysis, and interactive conversations involving visual information. For example, a multimodal LLM can analyse a photo, generate a descriptive paragraph or respond to questions about the image, thus providing a more comprehensive and intuitive user experience.

Large Language Models (LLMs) can be used in psychiatry and can enrich the descriptive nature of the field by providing advanced tools for analysing complex human emotions, thoughts and behaviours expressed through language. LLMs can help clinicians to recognise clues to mental disorders such as depression, anxiety or schizophrenia in patients' speech or writing. These models can provide information about an individual's mental health by analysing salient language patterns such as frequently repeated phrases, emotional tones or particular themes in texts (Kjell et al. 2023). For example, by analysing therapy session transcripts, an LLM can detect linguistic cues and thematic content that may be missed during traditional assessments (Volkmer et al. 2024). Multimodal LLMs can also perform analyses based on image data of psychiatric patients for whom mood assessment is critical (AlSaad et al. 2024). In addition, these models can synthesise the extensive psychiatry literature, helping researchers and practitioners to stay abreast of the latest developments.

Another advantage of language models is that, due to their training methods and design features, they can produce correct responses despite spelling or grammatical errors in user commands. Training on large and diverse text corpora, including examples of incorrect and informal language use, enables models to develop a nuanced and context-sensitive understanding of language that goes beyond strict grammatical rules (Gao et al. 2020). These models prioritise semantic integrity over superficial features and assess the importance of different input elements using attention mechanisms, so that they can often correctly infer intended meanings even when errors are found (Zhao et al. 2023). Furthermore, some models are specifically trained with 'noisy' inputs to increase fault tolerance (Karpukhin et al. 2019). This multifaceted approach can also be useful in understanding the reports of people who present to psychiatry and have some difficulties in expressing their problems. However, this may have some negative consequences, especially for people who are in the process of language learning; getting correct answers despite their mistakes may make the learning process difficult for people, and the understanding shown by the language models may cause distortion of everyday language because it is not the kind that can be found in daily life (Adeshola et al. 2023). The limitations of these models are that although the input is the same, they can produce different responses each time and these responses can be contradictory in some cases (de Leon 2023).

LIMITATIONS OF ARTIFICIAL INTELLIGENCE IN PSYCHIATRY: BIAS, ACCURACY AND ETHICAL CHALLENGES

Despite their impressive capabilities, LLMs have a number of limitations. First, building a LLM requires access to big data and the resources to analyse it (Hoffman et al. 2022). Second, various biases may be present in the training data and these biases may be 'inherited' and magnified during LLM training, potentially leading to inequitable outputs (Acerbi et al. 2023). Furthermore, LLMs have a fixed information cut-off date, so they do not have access to information or events that occur after their training period, which may result in outdated or inaccurate responses (Wang et al. 2023). Humans rely on sensory information derived from the physical environment to understand and navigate the world. In addition, they use 'embodied cognition', which enhances cognitive processes by integrating mind-body interactions, leading to a more comprehensive understanding of their environment (Dove 2023). LLM models are also being worked on to be embodied agents for better goal-oriented decision making. However, at the current level of progress, as discussed in detail by Özer (2024), a critical problem for LLMs is the output of LLMs, also called hallucinations, which appear coherent and plausible but are in reality false, misleading or completely fabricated (Banerjee et al. 2024, Maynez et al. 2020). This is because LLMs are trained to predict possible word sequences based on patterns in the training data, rather than having the ability to understand or reason about the real world in the sense that the human mind works. While this can yield creative results in some contexts, it can pose significant challenges in applications that require 'accuracy' in the sense of matching reality (Woodland 2023). In fact, this fabricated output is incompatible with the clinical use of hallucination as a term, as hallucinations are perceptual disturbances related to the senses of hearing, sight, touch and smell (Sterzer et al. 2018, Russo et al. 2019). Although the term hallucination evokes the process of making up (creating) something that does not exist in many people's minds, confabulation may be a more appropriate term in this context. Because confabulation is a type of memory error characterised by the unconscious production of fabricated, distorted or misinterpreted memories. As Özer (2024) argues, a number of recent scientific papers prefer the term confabulation to hallucination (Smith et al. 2023). Contrary to assumptions, confabulations are not perceived experiences, but inaccurate reconstructions of information due to the influence of past experiences, expectations and context. When answering questions, LLMs produce responses based

on patterns learnt from very large data sets. The phenomenon of confabulation in both humans and LLMs involves internal processes that produce outputs that are not fully grounded in external reality, emphasising a parallel between the brain's predictive processing and statistical prediction mechanisms in LLMs (Wang et al. 2023).

Bias is another critical issue in AI models; for example, if certain demographic groups are underrepresented in training data, the results produced by AI may not be fair to the community to which it will be applied and may increase health inequalities (Rajkomar et al. 2018). Moreover, the complexity and subjectivity of mental health pathologies pose a challenge for AI systems, which struggle to interpret nuanced human emotions, cultural contexts and individual experiences that are crucial for accurate assessments. Cultural representations of phenomena must be adequately captured in the context of language-based differences between countries and cultures in the datasets. Furthermore, the lack of transparency and explainability in some AI algorithms may make it difficult for clinicians to understand the rationale behind AI-generated recommendations, hinder their ability to make informed decisions, and undermine trust in the technology (Antoniadi et al. 2021).

The entrusting of the job of psychiatrists to AI raises ethical concerns; AI may not provide empathy and therapeutic alliance, which are essential elements in the patient-physician relationship, as expected (Morrow et al. 2023). Furthermore, as AI is increasingly relied upon in fields ranging from education to healthcare, the risk of widespread use of any misinformation that may be associated with AI will become a serious threat. Developing systems that alert users to the accuracy of these models in real time, for example by integrating feedback loops into use where user interactions help to improve model responses, can reduce the risks associated with confabulations (Nahar et al. 2024). Furthermore, when using LLMs in sensitive areas, transparency about how the content of these models is generated will enable users to critically evaluate the information presented to them (Reddy et al. 2024).

Addressing the shortcomings of LLMs in psychiatric applications requires a versatile approach. For example, fine-tuning is a process whereby a pre-trained language model is further trained on a smaller, customised dataset to adapt its knowledge and performance to a specific domain or task. This technique allows the model to learn domain-specific vocabulary, contexts and nuances, potentially leading to more accurate and reliable outputs (Gu et al. 2021). Fine-tuning models on high-quality, domain-specific psychiatric data can significantly improve accuracy and reduce confabulations. Another approach, Retrieval-Augmented Generation (RAG), increases factual accuracy by grounding LLM outputs with validated psychiatric information. This method combines the generative capabilities of LLMs with information retrieval from a regulated knowledge base (Lewis et al. 2020). When generating responses, the model first retrieves relevant information from a trusted psychiatric database, then uses this information to validate and constrain its output. This approach not only increases factual accuracy, but also increases transparency as the information sources are traceable and citable. Furthermore, Reinforcement Learning from Human Feedback (RLHF), proposed by Bai et al. in 2022, is a method that can be used to align AI models with best clinical practice in mental health. Mental health professionals evaluate the model's responses, rating them for accuracy, empathy and ethics. The model then fine-tunes using these expert ratings as rewards. This iterative approach involving human feedback increases the reliability of AI outputs, ensuring that responses are clinically appropriate and consistent with ethical principles. By incorporating expert knowledge, RLHF creates more reliable AI assistants for mental health practice. Through this secure and controlled structure, clinical standards can be improved by utilising the extraordinary capabilities of AI in pattern recognition. The continuous learning method allows new knowledge to be gradually incorporated into the model, overcoming the knowledge cut-off date problem and keeping AI models up-to-date with the latest psychiatric findings, treatment approaches, and ethical guidelines (Parisi et al. 2019).

ENSURING CONFIDENTIALITY AND ETHICAL INTEGRITY IN AI-SUPPORTED PSYCHIATRY SERVICES

Ensuring patient privacy during the training and use of LLMs poses a significant challenge and is a topic of ongoing debate. The performance of LLMs is often measured using metrics on reasoning, benchmarking, accuracy in data sets, and efficiency in specific tasks such as coding or language comprehension. LLMs owned by private companies require data to be temporarily stored on company servers in order to function, which can lead to breaches of patient privacy due to the risk of unauthorised access to data and problems with company compliance with data protection regulations (Nazi et al. 2024). Applications intended for use in the United States must comply with regulations specified by the Health Insurance Portability and Accountability Act (HIPAA) (U.S. Department of Health & Human Services 2023). According to the recommendations of AI language model server companies that are actively used today, a Business Partner Agreement with the AI company is required to ensure compliance with data protection and security standards (OpenAI 2023). Furthermore, according to the recommendations of experts working in the field of ethics, product design should take into account UNESCO's ethical guidelines on neurotechnology and the recommendations of the Personal Data Protection Authority for Türkiye regarding AI (International Bioethics Committee 2022, Kişisel Verileri

Koruma Kurumu 2023). This approach can ensure that personal data privacy is in line with both national and international regulations. A precautionary stance focussing on risk prevention and mitigation should be adopted to protect fundamental rights and freedoms. At all stages of data collection and processing, necessary measures should be taken to prevent discrimination and avoid adverse consequences for individuals, and the protection of fundamental rights should be consistently prioritised. The data used should be evaluated according to its quality, source and category, and only the minimum necessary data should be collected. In psychiatry, it is not always clear what the minimum data required should be and there is limited guidance on this. Determining what information is necessary is an important area of ongoing research.

Some open-access LLMs offer the possibility to run on local servers without the need to send data to their own servers, giving organisations the ability to integrate and manage AI systems independently (Jiyang et al. 2023, Touvron et al. 2023). This approach provides greater control over data privacy and security, complies with stringent healthcare requirements and allows for transparency and customisation. Healthcare providers can audit the code, implement additional security measures and fine-tune the models to meet specific needs in psychiatric care, thereby reducing bias, increasing accuracy and maintaining the high ethical standards required in mental health care. Models operating in speech-to-text localised environments are another important component of data and patient privacy. By processing and analysing spoken language in real-time, these applications can facilitate tasks such as transcription or the detection of speech patterns associated with specific mental health conditions. This approach, exemplified by the systems proposed by Collobert and colleagues, complies with strict regulations on healthcare confidentiality and can ensure that confidential information remains in a controlled environment (Collobert et al. 2016). These practices can improve the effectiveness of psychiatric care by allowing clinicians to gain valuable insights during therapy sessions without compromising patient confidentiality.

Another challenge to the widespread application of AI in psychiatry is its low acceptability (McCradden 2023). Patients and mental health professionals may be concerned about relying on AI for mental health assessments and interventions, fearing that it may undermine the therapeutic relationship or lack the empathy necessary in psychiatric care. To overcome this barrier, it is important to present AI as a tool that helps human professionals, not replaces them. Acceptance of AI tools with clinicians can be achieved by emphasising the support they provide, such as providing them with valuable insights about their patients, improving diagnostic accuracy and allowing them more time for patient interaction. This approach ensures that the human connection remains at the forefront of mental health care while utilising AI to improve outcomes and efficiency.

In addition to the limitations listed above, it should be emphasised that evidence-based data on psychiatric practices involving AI are still insufficient. Suggestions for solutions are listed in the following sections.

THE NEED FOR TURKISH LLMS AND PSYCHIATRIC CORPUS TO IMPROVE AI SERVICES IN TÜRKIYE

The use of AI in psychiatry in Türkiye is still in its infancy. As in other branches, apart from the use of large data sets such as genetics, biochemistry and imaging, studies on the detection of psychiatric disorders by speech analysis have started to be carried out. For example, a research group at Bilkent University is working on an AI-based software that detects depression using speech content, tone of voice, facial expression and posture data, and the first data of the study have been published (Kaynak and Dibeklioğlu 2024). In addition, the study conducted by Çabuk et al. (2024) on NLP in schizophrenia patients and the study conducted by Arslan et al. (2024) in patients diagnosed with bipolar disorder and psychosis make important contributions in this field. However, both studies have low sample sizes and their methods have not yet been replicated on different samples. Although research is conducted within various universities and health institutions, the integration and widespread use of these studies in clinical practice is still limited. Particularly in the field of digital health, steps need to be taken to establish public and private sector collaborations, and to establish technological infrastructure for collecting and analyzing clinical data. In Türkiye, collaboration between the Ministry of Health, the Psychiatric Association of Türkiye, universities, and other private institutions is crucial in developing applications that could be reflected in clinical use.

The development of Turkish-specific Large Language Models (LLMs) and psychiatric corpora is crucial for the advancement of AI services in Türkiye. Although global LLMs have made significant progress, they often lack the nuanced understanding of other languages and cultures necessary for optimal performance in local contexts (Tenzer et al. 2024). In particular, a large language model trained in a single language is likely to better understand that language's idioms, nuances, proverbs, and other cultural and local contexts compared to a model created by fine-tuning an existing multilingual model (Pires et al. 2023). In support of this, when various languages, including Turkish, were used with ChatGPT for different tasks, English consistently outperformed other languages in all tasks (Lai et al. 2023). This is particularly important in psychiatric applications where cultural sensitivity and semantic accuracy of language are crucial (Ratana et al. 2019).

However, developing LLMs for agglutinative languages like Turkish comes with its own challenges. Agglutinative languages form words by adding suffixes to a root, creating complex, informationally dense words (Chimalamarri et al. 2021). This morphological richness creates challenges for traditional tokenization methods used in LLMs (Petrov et al. 2024). Tokenization is a preprocessing method that divides text into smaller units such as words, roots, or affixes, enabling language models to process and make sense of these units. The multiplicity of possible word forms exponentially increases vocabulary size, making it difficult for models to learn and generalize effectively (Li et al. 2020). Additionally, the context-sensitive nature of affixes in these languages requires models to understand distantly related expressions and nuanced interactions between morphemes. To overcome these challenges and effectively capture the linguistic subtleties of Turkish in LLMs, specialized approaches in data preprocessing, model architecture, and training strategies are needed.

Today, the number of open-source language models that understand Turkish and respond to questions is increasing day by day. These models are typically developed by fine-tuning multilingual original versions with Turkish corpora. An alternative to this method is creating a new language model from scratch using only Turkish corpora.

FUTURE TRENDS FOR ARTIFICIAL INTELLIGENCE IN PSYCHIATRIC SERVICES

Artificial intelligence, beyond virtual therapists and psychiatrists, can assist mental health professionals in diagnosis, treatment, and follow-up stages, improve the quality and accessibility of mental health services, focus on individual needs, and can also be used as part of the psychoeducation process. Systems such as having a personal virtual psychotherapist available 24/7 for each individual with real therapists only providing supervision to virtual therapists, or assessing the need for psychiatric examination and even referral to a specific psychiatrist based on a person's daily conversations and communications, may not be too far away. Advanced AI systems can analyze an individual's conversations and all communications, provide instant feedback, and help provide customized effective therapy sessions according to their emotional and mental needs (Olawade et al. 2024). Smart systems that continuously collect data through wearable technologies and mobile applications can help better understand the connections between mental and physical health and create integrated treatment plans by analyzing physical health data. AI, combined with virtual reality and augmented reality technologies, can offer innovative and effective therapy methods in areas such as mindfulness applications, trauma therapy, and treatment of phobias.

In current conditions, making AI applications directly available to non-professional individuals would pose risks and might also hinder individuals' autonomy. For more reliable and effective use of AI in clinical settings, it might initially be appropriate to design these applications to assist experienced professionals in the field and offer them to patients under the guidance of experienced healthcare professionals within clinical processes. Considering the possibility that AI-based tests and applications may contain margins of error, it should be understood that the decision-making mechanism should ultimately belong to the clinician. Additionally, platforms reaching large audiences, such as social media, have the potential to contain biased or incomplete information-based content. Therefore, for the development and reliability of AI, more controlled progress should be ensured primarily using evidence-based clinical data and highly valid measurement tools. For now, awareness should be raised that the data provided by AI algorithms should not be considered directly as "reality" but rather as clinical "indicators" (for example, elevated troponin is an indicator of heart attack but not the heart attack itself) and consists of strong statistical predictions. Therefore, it will still be important today that AI is supported by clinical experience in every case where it is used in clinical processes.

CONCLUSION

Artificial intelligence applications integrated with mental health services offer significant potential to address psychiatry's global economic burden and workforce loss. AI techniques such as machine learning and natural language processing can enhance diagnosis, treatment, and patient follow-up. However, there are challenges such as diagnostic accuracy, ethical concerns, data privacy issues, and maintaining therapeutic collaboration. For countries like Türkiye, developing AI models that reflect linguistic and cultural nuances is crucial. Overcoming these barriers requires transparency, ethical strategies, and collaboration with mental health professionals. AI's potential to support professionals rather than replace them can improve the accessibility and outcomes of mental health services.

Turkish LLMs are still in the development phase, and an AI tool that can be used by psychiatrists in Türkiye would be very beneficial for psychiatry. There is a need for collaboration between patient representatives, psychiatrists, linguists, AI developers and researchers, ethics and legal experts in the development phase of AI tools to be used in psychiatry. This interdisciplinary collaboration can ensure that AI systems are culturally and linguistically appropriate and ultimately improve psychiatric care in our society. The pioneering role of experts in the field of psychiatry in the development of AI tools will play a role in increasing the confidence of both the mental health community and patient groups in this technology.

REFERENCES

Acerbi A, Stubbersfield JM (2023) Large language models show human-like content biases in transmission chain experiments. Proc Natl Acad Sci U S A 120: e2313790120.

Adeshola I, Adepoju AP (2023) The opportunities and challenges of ChatGPT in education. Interact Learn Environ: 1-14.

AlSaad R, Abd-Alrazaq A, Boughorbel S et al. (2024) Multimodal large language models in health care: applications, challenges, and future outlook. J Med Internet Res 26: e59505.

American Psychiatric Association (2013) Diagnostic and Statistical Manual of Mental Disorders: DSM-5. 5th ed. Washington DC, American Psychiatric Association.

Anderson SW, Rizzo M (1994) Hallucinations following occipital lobe damage: the pathological activation of visual representations. J Clin Exp Neuropsychol 16: 651-63.

Antoniadi AM, Du Y, Guendouz Y et al. (2021) Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. Appl Sci 11: 5088.

Arslan B, Kizilay E, Verim B et al. (2024) Computational analysis of linguistic features in speech samples of first-episode bipolar disorder and psychosis. J Affect Disord 15;363:340-7.

Ayhan Y (2023) The Impact of Artificial Intelligence on Psychiatry: Benefits and Concerns—An assay from a disputed 'author'. Turk J Psychiatry 34: 65.

Baduge SK, Thilakarathna S, Perera JS et al. (2022) Artificial intelligence and smart vision for building and construction 4.0: machine and deep learning methods and applications. Autom Constr 141: 104440.

Bai Y, Jones A, Ndousse K et al. (2022) Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv:2204.05862.

Ballı M, Ercan Doğan A, Hun Şenol Ş et al. (2024) Uncovering psychiatric predictors of suicidal ideation using non-suicidal predictors: machine learning insights from a university mental health clinic. (In preparation).

Banerjee S, Agarwal A, Singla S (2024) LLMs will always hallucinate, and we need to live with this. arXiv:2409.05746.

Barzilay R, Israel N, Krivoy A et al. (2019) Predicting affect classification in mental status examination using machine learning face action recognition system: a pilot study in schizophrenia patients. Front Psychiatry 10: 446117.

Basu JK, Bhattacharyya D, Kim TH (2010) Use of artificial neural network in pattern recognition. Int J Softw Eng Its Appl 4: 23-34.

Bhatnagar R, Sardar S, Beheshti M et al. (2022) How can natural language processing help model informed drug development?: a review. JAMIA Open 5: ooac043.

Bracher-Smith M, Crawford K, Escott-Price V (2021) Machine learning for genetic prediction of psychiatric disorders: a systematic review. Mol Psychiatry 26: 70-9.

Bzdok D, Meyer-Lindenberg A (2018) Machine learning for precision psychiatry: opportunities and challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 3: 223-30.

Cearns M, Hahn T, Baune B T (2019) Recommendations and future directions for supervised machine learning in psychiatry. Transl Psychiatry 9: 271.

Chimalamarri S, Sitaram D (2021) Linguistically enhanced word segmentation for better neural machine translation of low-resource agglutinative languages. Int J Speech Technol 24: 1047-53.

Collobert R, Puhrsch C, Synnaeve G (2016) Wav2letter: an end-to-end convnet-based speech recognition system. arXiv:1609.03193.

Cuthbert BN (2014) The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World Psychiatry 13: 28-35.

Çabuk T, Sevim N, Mutlu E et al.(2024) Natural language processing for defining linguistic features in schizophrenia: A sample from Turkish speakers. Schizophr Res 266:183-9.

Delipetrev B, Tsinaraki C, Kostic U (2020) Historical evolution of artificial intelligence.

Di Vincenzo M (2023) New research on validity and clinical utility of ICD-11 vs. ICD-10 and DSM-5 diagnostic categories. World Psychiatry 22: 171.

Dove GO (2023) Rethinking the role of language in embodied cognition. Philos Trans R Soc Lond B Biol Sci 378: 20210375.

de Leon J, De Las Cuevas C (2023) Will ChatGPT3 substitute for us as clozapine experts? J Clin Psychopharmacol 43: 400-2.

Ebden P, Sproat R (2015) The Kestrel TTS text normalization system. Nat Lang Eng 21: 333-53.

Elgendy M (2020) Deep learning for vision systems. New York, Simon and Schuster, p. 1-350.

Espejo G, Reiner W, Wenzinger M (2023) Exploring the role of artificial intelligence in mental healthcare: progress, pitfalls, and promises. Cureus 15: e37176.

Fusar-Poli P, Solmi M, Brondino N et al. (2019) Transdiagnostic psychiatry: a systematic review. World Psychiatry 18: 192-207.

GBD 2019 Mental Disorders Collaborators (2022) Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis. Lancet Psychiatry 9: 137-50.

Gao L, Biderman S, Black S et al. (2020) The pile: an 800GB dataset of diverse text for language modeling. arXiv:2101.00027.

Ghosalkar NN, Dhage SN (2018) Real estate value prediction using linear regression. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE, pp. 1-5.

Giorgi A, Ronca V, Vozzi A et al. (2021) Wearable technologies for mental workload, stress, and emotional state assessment during working-like tasks: a comparison with laboratory technologies. Sensors 21: 2332.

Greener JG, Kandathil SM, Moffat L et al. (2022) A guide to machine learning for biologists. Nat Rev Mol Cell Biol 23: 40-55.

Greenberg PE, Fournier AA, Sisitsky T et al. (2021) The economic burden of adults with major depressive disorder in the United States (2010 and 2018). Pharmacoeconomics 39: 653-65.

Gu Y, Tinn R, Cheng H et al. (2021) Domain-specific language model pretraining for biomedical natural language processing. ACM Trans Comput Healthc 3: 1-23.

Gupta R, Srivastava D, Sahu M et al. (2021) Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Mol Divers 25: 1315-60.

Hestness J, Ardalani N, Diamos G (2019) Beyond human-level accuracy: computational challenges in deep learning. In: Proceedings of the 24th Symposium on Principles and Practice of Parallel Programming. ACM, pp. 1-14.

Heinz A, Zhao X, Liu S (2020) Implications of the association of social exclusion with mental health. JAMA Psychiatry 77: 113-4.

Hickey BA, Chalmers T, Newton P et al. (2021) Smart devices and wearable technologies to detect and monitor mental health conditions and stress: a systematic review. Sensors 21: 3461.

Hoffmann J, Borgeaud S, Mensch A et al. (2022) Training compute-optimal large language models. arXiv:2203.15556.

Huang KL, Duan SF (2021) Affective voice interaction and artificial intelligence: acoustic features of gender and emotional states. Front Psychol 12: 664925.

International Bioethics Committee (2022) Ethical Issues of Neurotechnology: Report, Adopted in December 2021. UNESCO Publishing. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000383559

Jiang AQ, Sablayrolles A, Mensch A et al. (2023) Mistral 7B. arXiv:2310.06825.

Jiang F, Jiang Y, Zhi H et al. (2017) Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2: 230-43.

Johnson KB, Wei WQ, Weeraratne D et al. (2021) Precision medicine, AI, and the future of personalized health care. Clin Transl Sci 14: 86-93.

Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. Science 349: 255-60.

Joshi ML, Kanoongo N (2022) Depression detection using emotional artificial intelligence and machine learning: a closer review. Mater Today Proc 58: 217-26.

Juhn Y, Liu H (2020) Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. J Allergy Clin Immunol 145: 463-9.

Kalmady SV, Greiner R, Agrawal R et al. (2019) Towards artificial intelligence in mental health by improving schizophrenia prediction. NPJ Schizophr 5: 2.

Karpukhin V, Levy O, Eisenstein J, Ghazvininejad M (2019) Training on synthetic noise improves robustness to natural noise in machine translation. In: Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019). ACL, pp. 42-7.

Kaynak AB, Dibeklioğlu H (2024) Systematic analysis of speech transcription modeling for reliable assessment of depression severity. Sakarya University Journal of Computer and Information Sciences, 7(1), 77-91.

Kişisel Verileri Koruma Kurumu (2023). Yapay Zeka Alanında Kişisel Verilerin Korunmasına Dair Tavsiyeler. Ankara. Available from: https://www.kvkk.gov.tr/Icerik/7048/Yapay-Zeka-Alaninda-Kisisel-Verilerin-Korunmasina-Dair-Tavsiyeler

Kjell ON, Kjell K, Schwartz HA (2023) Beyond rating scales: with targeted evaluation, language models are poised for psychological assessment. Psychiatry Res 115667.

Knapp M, Wong G (2020) Economics and mental health: the current scenario. World Psychiatry 19: 3-14.

Krysta K, Cullivan R, Brittlebank A et al. (2024) Artificial intelligence in healthcare and psychiatry. Acad Psychiatry: 1-3.

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25: 1097-105.

Lai VD, Ngo NT, Veyseh APB et al. (2023) ChatGPT beyond English: towards a comprehensive evaluation of large language models in multilingual learning. arXiv:2304.05613.

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521: 436-44.

Lewis P, Perez E, Piktus A et al. (2020) Retrieval-augmented generation for knowledge-intensive NLP tasks. Adv Neural Inf Process Syst 33: 9459-74.

Li Z, Li X, Sheng J et al. (2020) AgglutiFiT: efficient low-resource agglutinative language model fine-tuning. IEEE Access 8: 148489-99.

Lipton ZC (2018) The mythos of model interpretability. Queue 16: 31-57.

Luo Y, Thompson WK, Herr TM et al. (2017) Natural language processing for EHR-based pharmacovigilance: a structured review. Drug Saf 40: 1075-89.

Martin-Carrasco M, Evans-Lacko S, Dom G et al. (2016) EPA guidance on mental health and economic crises in Europe. Eur Arch Psychiatry Clin Neurosci 266: 89-124.

Maynez J, Narayan S, Bohnet B, McDonald R (2020) On faithfulness and factuality in abstractive summarization. arXiv:2005.00661.

McCradden M, Hui K, Buchman DZ (2023) Evidence, ethics and the promise of artificial intelligence in psychiatry. J Med Ethics 49: 573-9.

Meltzer HY, Lowy MT (1986) Neuroendocrine function in psychiatric disorders. In: American Handbook of Psychiatry, 2nd ed. vol. 8, Berger PA, Brodie HKH (Eds), New York. Basic Books Inc, pp. 110-7.

Mikolov T, Deoras A, Povey D et al. (2011) Strategies for training large scale neural network language models. In: 2011 IEEE Workshop on Automatic Speech Recognition & Understanding. IEEE, pp. 196-201.

Minerva F, Giubilini A (2023) Is AI the future of mental healthcare? Topoi 42: 809-17.

Morar U, Martin H, Izquierdo W et al. (2020) A deep-learning approach for the prediction of mini-mental state examination scores. In: 2020 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, pp. 761-6.

Morrow E, Zidaru T, Ross F et al. (2023) Artificial intelligence technologies and compassion in healthcare: a systematic scoping review. Front Psychol 13: 971044.

Nahar M, Seo H, Lee EJ et al. (2024) Fakes of varying shades: how warning affects human perception regarding LLM hallucinations. arXiv:2404.03745.

Nazi ZA, Peng W (2024) Large language models in healthcare and medical domain: a review. Informatics 11: 57.

Nazimek JM, Hunter MD, Woodruff PW (2012) Auditory hallucinations: expectation–perception model. Med Hypotheses 78: 802-10.

Niemeyer H, Bieda A, Michalak J et al. (2019) Education and mental health: do psychosocial resources matter? SSM Popul Health 7: 100392.

Olawade DB, Wada OZ, Odetayo A et al. (2024) Enhancing mental health with artificial intelligence: current trends and future prospects. J Med Surg Public Health 100099.

Omarov B, Zhumanov Z, Gumar A et al. (2023) Artificial intelligence enabled mobile chatbot psychologist using AIML and cognitive behavioral therapy. Int J Adv Comput Sci Appl 14: 87-94.

OpenAI (2023) How can I get a Business Associate Agreement (BAA) with OpenAI? Available from: https://help.openai.com/en/articles/8660679-how-can-i-get-a-business-associate-agreement-baa-with-openai

Özer M (2024, Eylül). Yapay Zekanın Varsanıları mı Oluyor? Türk Psikiyatri Dergisi, Makale 27587. 31 Ekim 2024'de https://www.turkpsikiyatri.com/yayinlanmamis?type=c adresinden indirildi.

Parisi GI, Kemker R, Part JL et al. (2019) Continual lifelong learning with neural networks: a review. Neural Netw 113: 54-71.

Pastur-Romay LA, Cedrón F, Pazos A et al. (2016) Deep artificial neural networks and neuromorphic chips for big data analysis. Int J Mol Sci 17: 1313.

Pires R, Abonizio H, Almeida TS et al. (2023) Sabiá: Portuguese large language models. In: Brazilian Conference on Intelligent Systems. Cham, Springer Nature Switzerland, pp. 226-40.

Radford A, Wu J, Child R et al. (2019) Language models are unsupervised multitask learners. OpenAI Blog 1: 9.

Rajkomar A, Hardt M, Howell MD et al. (2018) Ensuring fairness in machine learning to advance health equity. Ann Intern Med 169: 866-72.

Ratana R, Sharifzadeh H, Krishnan J et al. (2019) A comprehensive review of computational methods for automatic prediction of schizophrenia with insight into indigenous populations. Front Psychiatry 10: 659.

Petrov A, La Malfa E, Torr P et al. (2024) Language model tokenizers introduce unfairness between languages. Adv Neural Inf Process Syst 36.

Reddy GP, Kumar YP, Prakash KP (2024) Hallucinations in large language models (LLMs). In: 2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream). IEEE, pp. 1-6.

Roh Y, Heo G, Whang SE (2019) A survey on data collection for machine learning. IEEE Trans Knowl Data Eng 33: 1328-47.

Russo M, Carrarini C, Dono F et al. (2019) The pharmacology of visual hallucinations in synucleinopathies. Front Pharmacol 10: 1379.

Saxena A (2021) Artificial intelligence and machine learning in healthcare. Singapore, Springer, pp. 1-200.

Savova GK, Danciu I, Alamudun F et al. (2019) Use of natural language processing to extract clinical cancer phenotypes from electronic medical records. Cancer Res 79: 5463-70.

Scangos KW, State MW, Miller AH et al. (2023) New and emerging approaches to treat psychiatric disorders. Nat Med 29: 317-33.

Squarcina L, Villa FM, Nobile M et al. (2021) Deep learning for the prediction of treatment response in depression. J Affect Disord 281: 618-22.

Simons JS, Garrison JR, Johnson MK (2017) Brain mechanisms of reality monitoring. Trends Cogn Sci 21: 462-73.

Smith AL, Greaves F, Panch T (2023) Hallucination or confabulation? Neuroanatomy as metaphor in large language models. PLoS Digit Health 2: e0000388.

Smrke U, Mlakar I, Lin S et al. (2021) Language, speech, and facial expression features for AI-based detection of depression. JMIR Ment Health 8: e30439.

Somé NH, Noormohammadpour P, Lange S (2024) The use of machine learning on administrative and survey data to predict suicidal thoughts and behaviors: a systematic review. Front Psychiatry 15: 1291362.

Sterzer P, Adams RA, Fletcher P et al. (2018) The predictive coding account of psychosis. Biol Psychiatry 84: 634-43.

Subramaniam K, Luks TL, Fisher M et al. (2012) Computerized cognitive training restores neural activity within the reality monitoring network in schizophrenia. Neuron 73: 842-53.

Su C, Xu Z, Pathak J et al. (2020) Deep learning in mental health outcome research: a scoping review. Transl Psychiatry 10: 116.

Sun H, Lin Z, Zheng C et al. (2021) PsyQA: a Chinese dataset for generating long counseling text for mental health support. arXiv:2106.01702.

Suthaharan S (2016) Support vector machine. In: Machine Learning Models and Algorithms for Big Data Classification. Springer, pp. 207-35.

Tenzer H, Feuerriegel S, Piekkari R (2024) AI machine translation tools must be taught cultural differences too. Nature 630: 820.

T.C. Sağlık Bakanlığı (2021) Ulusal Ruh Sağlığı Eylem Planı 2021-2023. Available from: https://hsgm.saglik.gov.tr/depo/Yayinlarimiz/Eylem_Planlari/Ulusal_Ruh_Sagligi_Eylem_Plani_2021-2023.pdf.

Türkiye Psikiyatri Derneği (2016) Devlet Hastanelerinde Psikiyatri Uygulamaları: Sorunlar ve Çözüm Önerileri. Available from: https://psikiyatri.org.tr/uploadFiles/2872016201037-DEVLET-HASTANELERINDE-PSIKIYATRI-UYGULAMALARI-SORUNLAR-VE-COZUM-ONERILERI-ozet-.pdf.

Touvron H, Lavril T, Izacard G et al. (2023) Llama: open and efficient foundation language models. arXiv:2302.13971.

Turing AM (2009) Computing machinery and intelligence. In: Epstein R, Roberts G, Beber G (Eds), Parsing the Turing Test. Springer, pp. 23-65.

U.S. Department of Health & Human Services (2023) Health Insurance Portability and Accountability Act (HIPAA). Available from: https://www.hhs.gov/hipaa/index.html

Vaswani A et al. (2017) Attention is all you need. Adv Neural Inf Process Syst 30: 5998-6008.

Volkmer S, Meyer-Lindenberg A, Schwarz E (2024) Large language models in psychiatry: opportunities and challenges. Psychiatry Res 116: 026.

Wang P, Zhang N, Tian B et al. (2023) EasyEdit: an easy-to-use knowledge editing framework for large language models. arXiv:2308.07269.

Wang S, Ding N, Lin N et al. (2023) Language cognition and language computation—human and machine language understanding. arXiv:2301.04788.

Woodland T (2023) ChatGPT for improving medical education: proceed with caution. Mayo Clin Proc Digit Health 1: 294-5.

World Health Organization (2018) The ICD-11 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research. Geneva, World Health Organization.

Wu J, Gan W, Chen Z et al. (2023) Multimodal large language models: a survey. In: 2023 IEEE International Conference on Big Data (BigData). IEEE, pp. 2247-56.

Yang X, Joukova A, Ayanso A et al. (2022) Social influence-based contrast language analysis framework for clinical decision support systems. Decis Support Syst 159: 113813.

Yapici Eser H, Kacar AS, Kilciksiz CM et al. (2018) Prevalence and associated features of anxiety disorder comorbidity in bipolar disorder: a meta-analysis and meta-regression study. Front Psychiatry 9: 229.

Zhang L, Wang M, Liu M et al. (2020) A survey on deep learning for neuroimaging-based brain disorder analysis. Front Neurosci 14: 779.

Zhao WX, Zhou K, Li J et al. (2023) A survey of large language models. arXiv:2303.18223.

Zlatintsi A, Filntisis PP, Garoufis C et al. (2022) E-prevention: advanced support system for monitoring and relapse prevention in patients with psychotic disorders. Sensors 22: 7544.

Zweifel P (2021) Mental health: the burden of social stigma. Int J Health Plann Manage 36: 813-25.