The Use of Artificial Intelligence in the Mental Health Space

Anya Garg

Abstract

The integration of artificial intelligence into mental healthcare has significantly changed service providers' ability to diagnose, treat, and monitor their patients. This review paper examines the various applications of AI in mental healthcare, as well as the challenges and opportunities presented by AI technology. We begin with a look at the cause and creation of AI technologies in the mental health field and an overview of the technologies that already exist and how they evolved over time. Next, we detail specific ways that AI helps both the patient and therapist and ways that technology makes up for the current shortcomings of mental healthcare providers. Finally, we take a look at concerns for using AI technology in the mental health field and different ways that these concerns are being handled. This review paper intends to elucidate the scope of AI technologies in the past and present, as well as its potential for the future.

Section 1: Introduction

AI technology is not a new idea, but it has been developed a lot over the years. The integration of artificial intelligence into the mental healthcare field has given therapists the additional help that they need in diagnosing patients, monitoring their wellbeing, and assisting therapists with other minor tasks. Autonomous from therapists, many AI programs are also able to provide people with conversational therapy. Though this feature has its limitations, it helps compensate for a shortage of therapists, who are also under-resourced. As development of AI technology continues in the field of mental health, AI systems will start to become much more accurate, reliable, and advanced. AI has the potential to do much more in this field, considering AI's usage was once limited to conversational agents and has now grown to diagnosing patients, prescribing medications, and predicting likelihoods of developing mental illnesses.

Although much work has been done in the field of AI so far, figuring out how to properly use it in mental health remains a challenge. This technology is still fairly new, and applying it in a field where mistakes have larger consequences requires developers to think more carefully when creating AI programs for mental health purposes. There is fear of bias in the system, which could result in an inaccurate diagnosis. This is especially dangerous to the patient because it could undermine the severity of the patient's condition and potentially mess up any medications that may be required for the patient to take. Many also worry about the AI's access to personal information and whether or not the AI system can maintain its users' privacy. Additionally, a majority of the users are part of the underprivileged demographic who cannot afford to pay for a therapist. This brings into question how well the AI system would work in scenarios where assistance for the underprivileged is crucial and there is less room for error. For instance, if a user is experiencing a panic attack and is dependent on an AI chatbot to walk them through the panic attack, it is critical that the chatbot be able to do so due to the urgency of the

situation. In more severe cases, if someone is planning to hurt themselves, the AI system must be able to talk the person out of it, similar to how a therapist would.

This review paper looks at the efficiency of using artificial intelligence in the mental healthcare field, exploring its background, its applications, a case study, and concerns regarding its use. Specifically, we begin by looking at why there is a need for AI in mental healthcare. We explore the development of AI technologies that have been used in this field and how the AI algorithms are developed. We then delve into the various ways that AI can independently provide assistance to both the therapist and the patient, also considering ways that an AI system can aid a therapist in helping a patient. Furthermore, we address one particular case in which an AI chatbot was used to help train future psychotherapists, assessing its efficiency and capability. Finally, we conclude with an overview of the different risks of using AI to provide mental healthcare and various ways that we can minimize those risks. By looking at current research regarding the different aspects of applying AI in the mental health field, this review aims to assess the capabilities and possibilities of AI technologies in providing mental health services.

Section 2: Background

Why AI is needed. Artificial intelligence has aided in providing mental healthcare for some time now, but demand significantly increased during the Covid-19 pandemic. Social outings and contact with others had become strictly regulated, causing an increase in mental health struggles for many people (Zhang et al., 2023). However, this was not the sole cause of the integration of AI into providing mental health care as there have been many issues necessitating outside help. For instance, there has long been a shortage of psychiatrists, in part due to a lack of funding to employ more of them. Additionally, many people who require assistance are unable to receive it because of a lack of providers in their area or even because they are unable to afford it due to socioeconomic issues. A large number of people also are simply uncomfortable talking to others about their problems, which has created the demand for AI to help reduce the fear of judgment and talking to another person face to face (Castañeda-Garza et al., 2023).

Timeline of AI use in mental health. The first instance of artificial intelligence in mental health services was in the 1950's, with a computer program called ELIZA that was developed at MIT's AI lab by Joseph Weizenbaum. This natural language processing chatbot became the first to simulate conversations between a human and computer. Though it was not advanced enough to truly understand much of what it was told, many people believed ELIZA to be fairly intelligent with some even growing attached to it (Vaidyam et al., 2019). Some time later, the DENDRAL (dendritic algorithm) system was created in the 1960's to help scientists identify molecular structures in the biomedical field. The DENDRAL system was quickly recognized in broader subjects of AI, including mental healthcare, with its ability to use its own reasoning and judgment supplementary to numerical calculations in order to produce an output ("Computers, Artificial Intelligence," 2019). As a result, psychiatrists were able to diagnose and treat mental illnesses using data from a patient's symptoms. The DENDRAL system was enhanced even further with the development of expert systems in the 1970's and the 1980's. Expert systems contain a database of predefined rules that can imitate the decision-making abilities of humans. They are able to take inputted data from the user and calculate the probability of a patient having a certain condition (Lubofsky, n.d.). For example, DIAGNO was the first expert system to play the role of a

human psychiatrist, with the ability to diagnose mental illnesses over a wide range (Luxton, 2016). Over time these AI technologies kept improving in accuracy and ability to analyze patient data and even determine the likelihood of someone developing a mental health disorder (Lubofsky, n.d.).

Overview of how AI works. There are a variety of ways that the algorithm of the AI model is trained in order to function properly, the most common ones being supervised machine learning (ML) and unsupervised machine learning. Through supervised ML, the algorithm is already given labeled data to learn from so the algorithm can find patterns to associate certain factors with specific labels. This algorithm is eventually tested by predicting the labels of data from datasets, comparing its predictions with the correct labels to determine its accuracy. These labels are essentially the "answer key" for the model, as they tell the algorithm what the correct diagnosis would be if given a specific set of patient data. Unsupervised ML teaches the algorithm how to group data by finding patterns in unlabeled data sets that it is given. The benefit of this method in comparison to supervised ML is that unsupervised ML prevents a bias in the AI, meaning that predictions are not based on prior knowledge that it has about a patient's data and their diagnosis from the training data set. Supervised and unsupervised ML can be used to train deep learning algorithms, which pass raw data through many different layers of artificial neural networks. Artificial neural networks function similar to how the brain does in order to learn and memorize the importance of each factor of the inputted data on the output. Algorithms trained by this method most closely resemble the human thinking process and allow the algorithm to create predictions based on more complex data. Another application of ML is natural language processing (NLP,) a field of AI that uses ML to take in different types of input (text, speech, and writing). This makes NLP much more useful to therapists who use it to process notes or conversation recordings in order to help give diagnoses (Graham, 2019).

Section 3: Methods / Applications of AI in the Mental Health Field

Introduction. Artificial intelligence can cater to a variety of tasks in mental healthcare. Though it is quite effective on its own, AI is even more beneficial when aiding a therapist. These technologies can help a therapist collect data about their patient. Such methods include taking and elaborating on notes from a session with the patient, monitoring their online presence, and even receiving live data to observe the patient in real time. It can also provide a second opinion when giving diagnoses and deciding which treatment to administer. The use of AI to aid therapists results in an overall increased efficiency as the technology is able to take care of smaller tasks and allow therapists to focus on more pressing matters (Rebelo et al., 2023). Thus, AI can be used to help therapists, patients, and improve the extent of support from therapists to patients.

Cognitive Behavioral Therapy. The majority of therapists use a method known as cognitive behavioral therapy (CBT) to treat patients with depression by eliminating the negative thought cycle. This method allows for a therapist and a patient to gain awareness of their feelings and thoughts, examine those thoughts and feelings (where they come from, what they mean), and strategize alternatives that promote a more beneficial state of mind. CBT is the preferred method for many patients not only because it is effective, but it is structured and allows the patient to create a structured plan on how to help themselves with guidance from their therapist instead of freely talking without much focus ("How it

Works," 2022). It is this effective method that AI is programmed to use when helping others in place of a therapist.

Methods that support the work of a therapist. There are many ways for AI to help a psychiatrist gather and process information about their patient. A patient's health records are typically quite extensive and can give the AI the necessary insight into the patient's existing medical background. This could help the provider in diagnosing and treating the patient. During sessions between the provider and the patient, AI tools are often used to take notes and recordings of the conversation. These can be processed and analyzed for concerning symptoms using natural language processing (NLP). AI also can be used to collect data from a patient's smartphone and social media in order to gauge their mental well-being. For example, NLP can be used to analyze a person's chats and emails to look for anything that may indicate deteriorating mental health. Data can also be derived from observing recent social media posts that the patient has been liking and google searches they have been making for the same purpose (Minerva & Giubilini, 2023). Many AI tools also process information that is given from the patient's responses to assessments and questionnaires. After processing information from the many available sources, an AI system can assist in finding important details that the therapist may have missed about a patient and give more accurate diagnoses (Rebelo et al., 2023). Information is also collected by wearable devices, such as smartwatches and health monitors to observe the patient's condition at any given moment and assess what their needs may be. These technologies can also record the patient's condition over time and in order to help determine if their conditions are getting better or if they need to use different treatments that the AI can better personalize to match each person's exact needs and predict its effectiveness (Minerva & Giubilini, 2023).

Support for patients. Though the benefits of AI for psychiatrists are numerous, they also provide a different level of support for patients. Technologies like chatbots give patients the assistance that they need at any time of the day and are especially helpful during emergencies if a provider is not immediately available. They can provide necessary intervention in extreme cases of mental health deterioration while also alerting a provider who can give more extensive assistance. AI chatbots also help people who simply are not able to share personal information with other people or struggle to interact with people (such as those with autism or other social disorders). It eliminates the need to meet with another person face to face and gives them the option of comfortably receiving assistance from behind a screen. Therapy has also gotten quite expensive with the surge in demand, so AI services give people in the middle and low income families a way of accessing the same resources online using technology and resources that they already have (Alowais et al., 2023).

Extending the support of a therapist to patients. Over time with constant testing and improvement, AI technology in mental health has become more accurate. It can even predict the likelihood of a patient attempting self-harm and provide treatments early on in order to prevent that level of severity in a high-risk patient. Other mental health conditions are a little more difficult for psychiatrists to properly treat, so AI can take the genetic and emotional information of one such patient and determine the best course of action for treating them. A major difference between a person and technology is that technology is unable to look at a person and make personal judgements about the person. AI can help eliminate biases that may alter a psychiatrist's judgment in diagnosing the patient and look at all the facts in order to give a more accurate diagnosis. Some of these eliminated biases may include gender, age,

social status, ethnicity, and past medical history. Though it is still possible that such biases may be present in the AI model, the model can be trained without data that may create bias, therefore eliminating it (Minerva & Giubilini, 2023).

Section 4: A case study

The intro and purpose. Now that we have looked into the foundations of AI-based support systems for therapists, we take a look at a specific system in action. In a study done by professors at the University of Utah, a conversational AI tool called ClientBot was evaluated on its ability to provide feedback to and assist in training aspiring psychotherapists. Traditional methods of training psychotherapists are extremely inefficient and time consuming. Typically there are workshops in which a conversation between a trainee and a patient is recorded, and then these conversations are evaluated by the supervisors. However, there are many problems with this approach. It takes an incredibly long time for the feedback to get back to the trainees, so the suggestions are usually no longer applicable to the patient. The feedback is also very vague and not very helpful. There is also an issue with selecting the patients that will help train the therapists and provide practice, as choosing higher risk patients could lead to making their condition worse if the help that they are provided is not from a professional. The NLP model tested in this study uses AI to eliminate these problems by providing a simulation of the patient and providing immediate feedback to the trainee (Tanana et al., 2019).

What it does. The AI model that is used to provide feedback to these trainees evaluates their capabilities in motivational interviewing (MI). This is a widely-used counseling method that involves a psychotherapist's interviewing and active listening skills. During the conversation simulation, the ClientBot is doing two things at once. The first thing it is doing is inputting the trainee's questions and outputting the best response. The second thing that it is doing is actually giving the trainee feedback and guiding them on how to proceed. This is done following the basis of motivational interviewing. As the ClientBot responds to the trainee, it is also helping them determine whether the next question they ask the simulation should be open, closed, or a reflection of the simulation's response. These suggestions help the trainee work on certain skills and teaches them how to respond to patients in a wide range of situations (Tanana et al., 2019).

How the LSTMs work. The ClientBot was designed to look like a texting platform and was developed using a combination of two LSTM (long short-term memory) models. Such models are advantageous because they have the ability to store information for an extended period of time, so they work better for translating information. First, these models filter through the information they are given and store only what is important. Then, they go through the information it has stored and make the decision to "forget" it once it is no longer needed by the model. When the model is giving an output to the user, it remembers what that output was and is able to use that information as needed when outputting the next value. Using this process, LSTM models are able to better find the relationships between words in the sentences of its input to properly construct an appropriate sentence for its output. They are typically used to translate words to different languages, perform speech-to-text functions, and recognize different emotions in words (Banoula, 2023).

How the ClientBot works. The first LSTM model of the ClientBot is a sequence to sequence (seq2seq) model, which uses beam searching to take the input of the trainee and decode it word by word to figure out and formulate what the best response would be. A seq2seq model uses the encoding and decoding method. The encoder takes the important information from the input and uses neural networks to find patterns in the input and figure out its overall meaning. The decoder then uses content from past responses and takes the meaning of the input to deliver the best possible output ("seq2seq model," 2024). The decoding function of the seq2seq model is assisted by beam searching, which uses an algorithm to look at all the possible output options and choose the most appropriate one. The seq2seq model was trained on a collection of English movie transcripts and 2,354 psychotherapy transcripts. In this study, this model was used for the first five responses of the stimulation with the trainee to make small talk and get the conversation started, since it was unable to formulate very long or complex responses. After a certain point, these responses are no longer useful and the ClientBot would switch to its second model, which is a simple LSTM model. In comparison to the seq2seq model, this one does not decode and encode the input. Instead, it makes a prediction regarding the best response. This model was trained only on a psychotherapy dataset, and was able to come up with longer responses that could go up to 50 words. The difference in this model was that instead of simply taking apart the trainee's question to turn it into a response, it had more freedom to predict the best response based on the information from the training dataset (Tanana et al., 2019). Figure 1 shows how both LSTM models work together to properly operate the ClientBot.

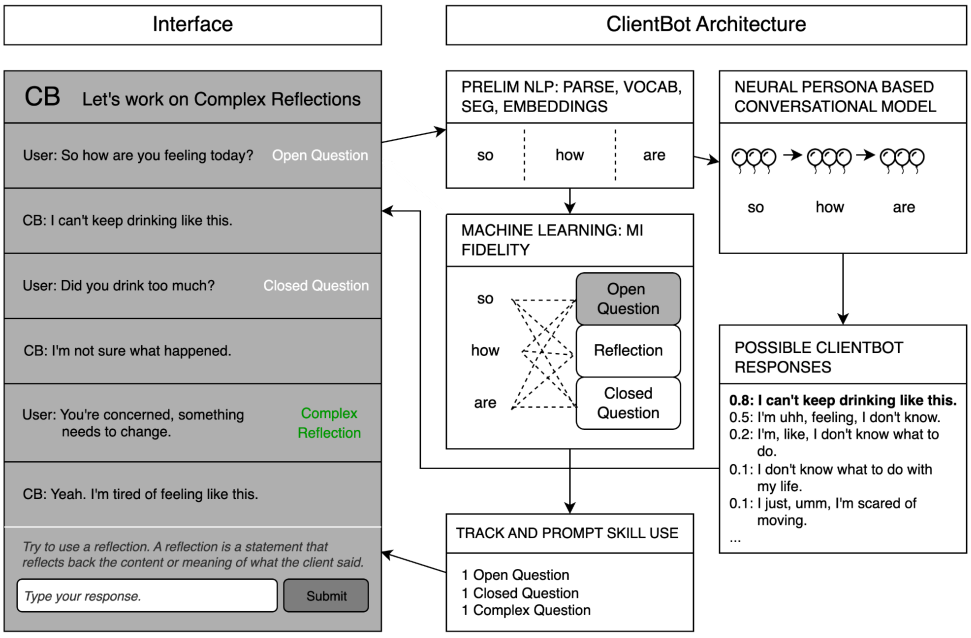

Figure 1: Model of ClientBot architecture (Tanana et al., 2019).

Takeaways about performance. In this study, researchers found that the ClientBot was effective in teaching the participants, who had no formal training, how to lead a conversation with a patient. The participants had become much better at using open-ended questions with the patient, and had especially shown great improvement in using reflections to get the simulation to think about its responses and what they meant. In comparison to the control group (the group that had not undergone training by the

ClientBot), the group that had gone through the ClientBot's training did much better in the evaluation at the end of the study, especially in the category of reflections (21.4% of the treatment group used them vs 11.2% of the control group) and open ended questions (30.4% of the treatment group used them vs 22.4% of the control group). The participants had also reported that the suggestions given to them by the ClientBot had been focused and helpful, as well as delivered in a timely manner. The model itself had also performed very well. The seq2seq model had achieved a perplexity of 9.06, which represents how well the model was able to take in some input and give an appropriate output (lower is better). The simple LSTM model was found to have a complexity of 38.01, which is not as effective as the seq2seq model but has varying results due to different data sets and model types. Overall, it was found to be a helpful tool in training beginner psychotherapists (Tanana et al., 2019).

Limitations. There were some aspects that were found lacking in the ClientBot. In the sequence to sequence model, the simulation very often responded with "I don't know," or "I love you." This was due to also being trained on movie transcripts and often thought that these were the best responses to the trainee's questions. However, this was easily fixed after retraining the model with less instances in the data that had those particular responses. Another result of being trained on movie transcripts was that the model often gave responses that were irrelevant or had no point to them. Relatedly, this entire simulation is much different from an actual conversation between two people, since ClientBot was based on written correspondence instead of verbal. Effectiveness could have been improved if the training data was taken from messages between real patients and therapists, or crisis interventions that occurred through text messages (Tanana et al., 2019).

Section 5: Risks and Concerns regarding Ethics and Potential Biases

Bias in AI. There are many concerns about using AI to treat mental health. Such concerns include users' privacy, incorrect predictions, delayed emergency response, and bias in the system. Biases, especially implicit (the ones that we are unaware of), are incredibly harmful to those affected, making them a cause for some of the biggest dangers of using AI in the mental health field. They are built off of negative stereotypes of different groups of people and often result in a higher level of inequality in school, politics, healthcare, and general social life. Bias is a learned behavior, and often comes from the environment in which one grows up and the people with whom one spends the most time (Timmons et al., 2022).

Types of Bias. There are many different ways that bias could alter the performance of the AI model, including sociocultural foundations, data collection, model building, model evaluation, and human deployment. Figure 2 gives an overview of how each method of introducing the different types of bias can affect each step of building the model. Sociocultural foundations consider structural inequities (biases in institutions, governments, and organizations that cause certain groups to be favored), historical bias (bias in data in regards to its historical context), and homogenous teams (AI development teams lacking diversity). Further, during the data collection process, data used to train the algorithm could be lacking in diversity, more accurate for some groups of people than others, or treated as data for one group when the data accounts for many different groups. While developers are building the model, the system may encounter confirmation bias (developers favor data that supports pre-existing beliefs),

label bias (during the machine learning process, labels given to the training data may not represent all of the possibilities for each variable), or feature-selection bias (certain variables of an input are selected to be the best predictors for the output, which may decrease more diversely representative variables). During the evaluation of the model, the system may encounter class-imbalance bias (certain groups have more training samples in the training data set than others, resulting in inaccuracies for the groups with less training data), covariate shifts (occurs when the distribution of the testing data does not correlate to the distribution of the training data due to subtle shifts in the population over time), or evaluation bias (when both the training and the testing data are equally unrepresentative, giving the model false accuracy). Lastly, during the deployment of the model, the model may experience deployment bias (the model is deployed in a setting that it was not designed for), user-interaction bias (a user's interactions with the model introduces new biases into the system), or feedback-loop bias (a user's interactions with the model reinforce pre-existing biases in the model) (Timmons et al., 2022).

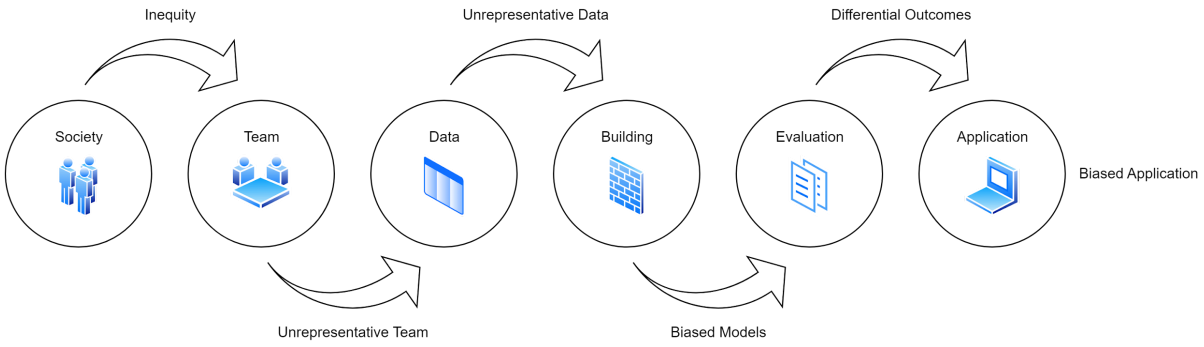

Figure 2: Stages of model-building where bias can enter the model (Timmons et al., 2022).

Risks of Bias in AI. The biggest threat of biased AI is that it may give an output that is not correct, such as giving a patient a false diagnosis. It may take that person's background and demographics into consideration, and could make incorrect assumptions that affect its decision. The AI may also misjudge the severity of the patient's condition, resulting in a lapse of intervention before the condition worsens. Tools that are capable of forming and administering treatments also must be unbiased. Such technology could choose a treatment that works for the majority of a large group of people, but may not work for everybody or for the underrepresented group of the patient being treated. These potential biases could result in a pattern of learned behavior for the AI that causes many of its future decisions to be based on biases as well (Timmons et al., 2022).

Privacy. Another concern of using AI would be patients' privacy. Sometimes the model may make predictions that require a breach in patient confidentiality (Timmons et al., 2022). According to set guidelines, mental health providers are obligated to report cases in which a patient has harmed themselves or others, or is planning to (Pederson, 2022). Patients fear the AI forcing a provider to break confidentiality in cases where patients are planning to harm themselves, but have not yet done so. Additionally, many users disagree with or dislike the idea of the AI going through the users' personal information and the data in their phone in order to give a diagnosis or monitor their wellbeing. Certain

technologies used to monitor a patients' well being can be considered invasive since that technology is constantly looking at patients' chats, search history, and overall device activity (Diaz-Asper et al., 2024).

Transparency and Explainability. To minimize the numerous risks of using AI to help patients, transparency is highly encouraged. In EU guidelines, transparency of an AI system can be defined by its "traceability" (documentation of data sets and how they are processed), "explainability" (ability to explain technical processes and decisions), and "communication" (providing information about the use, capabilities, and limitations of the system). Similarly, WHO guidelines state that any such program should be easily explainable by the developers, meaning that those using the program should know what exactly the program does and how it works. Users should be aware of which type of data the model is being trained on, any shortcomings in its performance, and biases that may arise in its predictions. To take this one step further, mental health providers should also be provided transparency with the technologies that they are using. When an AI system is helping a therapist diagnose a patient, the therapist should be able to understand the logic and reasoning behind that diagnosis. This also ensures the continued accuracy of the model if there is a person double checking the model's outcomes (Diaz-Asper et al., 2024).

How to Prevent Bias. The best way to minimize the risk of biases in AI is to be transparent with its users, as previously explained. It would help to let people know exactly what the potential risks are and how exactly the system works so that users can decide for themselves whether or not AI is worth using. Since the proper amount of data is not available to train the algorithm to properly serve underrepresented groups, developers should test the models on the underrepresented groups in order to determine its weak spots and ensure that they are fixed as best as possible by either removing some data from the datasets that they already have or tweaking it. Developers should be able to get as close as possible to having the same level of accuracy in diagnosing and otherwise assisting both the represented and underrepresented groups (Diaz-Asper et al., 2024).

Solution: human in the loop. In especially risky situations, it may even help for there to be an actual human monitoring the technology and filling in gaps for the AI as it produces an appropriate response for the user. Over time, human supervision can decrease and the AI can become autonomous once it has learned to look past any biases. At the end of the day, such technology must be safe to use and must serve its purpose in helping people rather than harming them (Diaz-Asper et al., 2024).

Regulations. Although there are many considerations to be made when creating guidelines for the use of AI, some rules have already been put in place. The AI Act in Europe, which is set to be put in motion sometime in the next two years, is the first legal framework that addresses the risks of AI. The rules in the AI Act will require the assessment of new AI systems, identify and prohibit unacceptably risky AI applications, limit high-risk AI systems, and encourage other nations to also consider addressing such topics. The framework outlined in the AI Act identifies four levels of risk that an AI system may pose: minimal, limited, high, and unacceptable risk. Anything that is considered as having unacceptable risk will be banned from use. In the case of a high-risk AI model (anything with the ability to cause harm or access private information), the model must be thoroughly assessed in its capabilities, trained with quality datasets, and heavily monitored. AI models with limited risk are usually considered so because of a lack of transparency, which the AI Act requires for any such system. Transparency is also required for

any general-purpose AI models, which are combined–in more ways than are possible to keep track of–to create more complex AI systems. Any AI model that is classified as low risk is allowed free use, which describes a large majority of the AI systems currently used in Europe. Through this act, Europe hopes to foster the development of ethical AI technologies and promote cooperation around the globe in regulating AI systems (“Regulatory Framework,” 2024).

Conclusion

With the application of AI technologies into mental healthcare, the ability to diagnose, treat, and monitor patients has been revolutionized and become much more accessible. This review has covered the history of AI technologies available in the mental healthcare field, ML/DL techniques used to train the AI, the use of current AI systems in assisting both patients and therapists, and current challenges that AI developers face with the increasing use of automated technology for mental healthcare. As we look to the future of artificial intelligence, these challenges, including, but not limited to, privacy, bias, and ethics, must be properly addressed in order for AI to reach its full potential in providing mental health services.

References

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A., Almohareb, S. N., Aldairem, A., Alrashed, M., Saleh, K. B., Badreldin, H. A., Yami, A., Harbi, S. A., & Albekairy, A. M. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Medical Education, 23(1). https://doi.org/10.1186/s12909-023-04698-z

Banoula, M. (2023, April 27). Introduction to Long Short-Term Memory(LSTM) | Simplilearn. Simplilearn.com. https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/lstm#:~:text=LSTMs%20are%20able%20to%20process

Castañeda-Garza, G., Ceballos, H. G., & Mejía-Almada, P. G. (2023). Artificial Intelligence for Mental Health: A Review of AI Solutions and Their Future. What AI Can Do, 373–399. https://doi.org/10.1201/b23345-22

Computers, Artificial Intelligence, and Expert Systems in Biomedical Research. (2019, March 12). Joshua Lederberg - Profiles in Science. https://profiles.nlm.nih.gov/spotlight/bb/feature/ai

Diaz-Asper, C., K. Hauglid, M., Chandler, C., S. Cohen, A., W. Foltz, P., & Elvevåg, B. (2024). A Framework for Language Technologies in Behavioral Research and Clinical Applications: Ethical Challenges, Implications, and Solutions. Psycnet.apa.org. https://psycnet.apa.org/fulltext/2024-44313-007.html

Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H.-C., & Jeste, D. V. (2019). Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. Current Psychiatry Reports, 21(11), 116. https://doi.org/10.1007/s11920-019-1094-0

How It Works - Cognitive Behavioural Therapy (CBT). (2022, November 10). Nhs.uk. https://www.nhs.uk/mental-health/talking-therapies-medicine-treatments/talking-therapies-and-counselling/cognitive-behavioural-therapy-cbt/how-it-works/

Lubofsky, M. (n.d.). History of AI in Behavioral Health. PsychotherapAI. https://psychotherapai.com/history-of-ai-in-behavioral-health/

Luxton, D. D. (2016). Artificial intelligence in behavioral and mental health care (p. 29). Elsevier/Academic Press. (Original work published 2015)

Minerva, F., & Giubilini, A. (2023). Is AI the Future of Mental Healthcare? Is AI the Future of Mental Healthcare?, 42(3). https://doi.org/10.1007/s11245-023-09932-3

Pedersen, T. (2022, July 14). Therapist Confidentiality: What Therapists Have to Report. Psych Central. https://psychcentral.com/health/what-do-therapists-have-to-report

Rebelo, A. D., Verboom, D. E., dos Santos, N. R., & de Graaf, J. W. (2023). The impact of artificial intelligence on the tasks of mental healthcare workers: A scoping review. Computers in Human Behavior: Artificial Humans, 1(2), 100008–100008. https://doi.org/10.1016/j.chbah.2023.100008

Regulatory framework on AI | Shaping Europe's digital future. (2024, March 6). European Commission. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

seq2seq model in Machine Learning. (2024, January 3). GeeksforGeeks. https://www.geeksforgeeks.org/seq2seq-model-in-machine-learning/

Tanana, M. J., Soma, C. S., Srikumar, V., Atkins, D. C., & Imel, Z. E. (2019). Development and Evaluation of ClientBot: Patient-Like Conversational Agent to Train Basic Counseling Skills. Journal of Medical Internet Research, 21(7), e12529. https://doi.org/10.2196/12529

Timmons, A. C., Duong, J. B., Simo Fiallo, N., Lee, T., Vo, H. P. Q., Ahle, M. W., Comer, J. S., Brewer, L. C., Frazier, S. L., & Chaspari, T. (2022). A Call to Action on Assessing and Mitigating Bias in Artificial Intelligence Applications for Mental Health. Perspectives on Psychological Science, 18(5), 1062–1096. https://doi.org/10.1177/17456916221134490

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie, 64(7), 456–464. https://doi.org/10.1177/0706743719828977

Zhang, M., Scandiffio, J., Younus, S., Jeyakumar, T., Karsan, I., Charow, R., Salhia, M., & Wiljer, D. (2023). The Adoption of AI in Mental Health Care–Perspectives From Mental Health Professionals: Qualitative Descriptive Study. JMIR Formative Research, 7(1), e47847. https://doi.org/10.2196/47847