HIGH SCHOOL EDITION
Journal of Student Research

Volume 13 Issue 4 (2024)

An Exploration of the Implications and Possible Applications of Generative AI in Mental Healthcare

Hans Manish[1] and Nicholas Traugott[#]

[1]Independence High School, USA
[#]Advisor

ABSTRACT

The past few years have seen significant growth in the capabilities of Generative Artificial Intelligence (GAI), and it has shown incredible potential to impact the field of mental healthcare. GAI encompasses a variety of AI architectures, all of which can process information and create output at tremendous speeds, showing promise in making mental healthcare more efficient and revolutionizing psychological research and patient care. In this work, we analyze the recent advancements made within the GAI space, consider the possible applications that GAI could have in the research and clinical environments of mental healthcare, and acknowledge the potential limitations of GAI when considering pathways to integration. Ultimately, we recognize the many benefits that GAI could create in the field of mental healthcare in improving patient outcomes, reducing workload for medical professionals and researchers, and reducing financial and resource consumption while also asserting that the implementation of GAI must be done with care to prevent harm and detriment to patients.

Introduction

Over the past decade, one of the most prominent forms of artificial intelligence, Generative Artificial Intelligence, has experienced exponential growth in its capabilities and general use. Generative AI, classified for its ability to create novel media (text, images, sound, etc.) through analysis of training data that is the same form of media, has become one of the most significant aspects of recent technological advancements in multiple distinct industries.

In the technology industry, many companies have worked to develop their own generative AI systems or implement existing GAI frameworks (Feuerriegel et al., 2024). These GAI systems include those pioneered by OpenAI (GPT-4, DALL-E 3, Sora) ("OpenAI API," 2024; "Sora | OpenAI," 2024), Google (Google Gemini) ("Gemini," 2024), and Microsoft ("Microsoft copilot," 2023), each of which is primarily implemented in everyday, commercial contexts. Chat GAIs, for example, have been implemented as chatbots that can provide 24/7 online technical support for many companies (Marr, 2024). Image GAIs have similarly been utilized to create relatively realistic image media for website design and slideshows (Yadav, 2024), and Video GAIs, though comparatively more recent than the other two types, have been discussed as a significant tool that could be utilized in film production (Rothman, 2024). Ultimately, it is clear that the capabilities of GAI make it a powerful tool that many can take advantage of in a multitude of industries.

With these advances, however, many medical professionals have debated the possible applications of GAI in clinical and research contexts in the field of healthcare. The most recent explorations into medical uses of GAI have centered mainly around fields like Oncology, Cardiology, and Medical Education, among others. More specifically, GAI has been implemented in cancer research to develop a possible T-cell engager immunotherapy (Teng, 2023), in Cardiovascular imaging to analyze echocardiograms more effectively (Christensen et al., 2024), and in medical training as patient simulations (Sardesai et al., 2024). In contrast, the field of mental healthcare, despite having many niches in which GAI can refine the quality of research, improve patient care outcomes, and better assist the already strained capacity of professionals and care providers in the field, has yet to adopt or fully explore the possible benefits of GAI.

In the past, novel technologies have been implemented into mental healthcare in various ways, especially with the spread of digitized medical technologies over the past twenty years. Virtual medical interventions and online methods of receiving care have especially permeated the field due to how they have bolstered access to mental healthcare, increased the availability of emergency support, and helped reduce the effect of pervasive mental health stigma. Similarly, asynchronous digital tools have been utilized to offer support outside healthcare provider-led interventions, accompanying these interventions with digitized symptom management and therapy accompaniment. Overall, the success of digitized mental health interventions is well-documented, and many studies, though not comprehensively evaluating the quality of digitized mental healthcare interventions in comparison to in-person interventions, often depicted that digitized mental healthcare frequently had equally as much benefit on patient care and that, in some ways, digital methods improved upon the in-person methods that were compared (Kalman et al., 2023). Ultimately, the recent success of digitized mental healthcare depicts the integration of growing technologies as an opportunity for medicine to become more effective and fruitful, making the development of medical-use GAI frameworks fundamental to the evolution of mental healthcare. As such, within the past four years, literature has been published about the inclusion of GAI in specific parts of mental healthcare, from developing various pharmacotherapies to creating supplemental therapeutic tools to formulating diagnoses in clinical practice.

This work aims to synthesize the work of various professionals in the field of mental health and explain the possible use cases of Generative Artificial Intelligence frameworks in mental healthcare, including from as early as the research stage to as late as post-intervention care. We will first explain some of the basic structures and functions of GAI relevant to mental healthcare. We will then explore the many possible applications of GAI in mental healthcare, including potential applications in clinical research, direct care provisions, and post-intervention continuance of care. Finally, we will discuss some of the limitations that may hinder the large-scale implementation of GAI in mental healthcare to highlight the pathways by which GAI can be safely integrated.

Literature Review

Specific Functions of Generative AI

GAI architecture construction centers around several vital processes, the most significant of which is the process of training the GAI model with existing data that fits its preassigned purpose. During the training phase, collected data is examined for quality, accuracy, and relevance. In this sense, the development of a GAI through the training of a model can be finely tuned through the use of specific, discrete training data (Bandi, 2024), which, therefore, can be utilized in medical contexts to ensure that medical GAIs are only trained on accurate, medically-affirmed data. Ultimately, the scalability of a GAI tool created for any medical concept, and therefore the training it receives, can be specialized to the purpose of the tool itself.

A paper by Bengesi and others elucidate how this variance in scalability can simultaneously be paired with the specific frameworks of artificial intelligence that fall under the umbrella term of GAI, each of which can be utilized to offer medically-constructed GAI tools applicable functionalities. Variational Autoencoders (VAEs) utilize the autoencoder infrastructure of encoding and decoding (with bottlenecks) for incredibly realistic image, audio, and video creation. These forms of outputs are similarly created using Generative Adversarial Networks (GANs), which utilize a generator-discriminator pair of neural networks to generate synthetic data of many formats, and Diffusion models, which are probability-based that use a pair of forward and reverse diffusion processes to create images. A slightly different form of GAI related to image processing is Neural Radiance Fields (NeRFs), which utilize neural network structures to analyze 2D imagery to create 3D image outputs (Zhang et al., 2020). In comparison, Bengasi also notes how text GAIs can be classified in a multitude of ways, including as Transformer-based models for text generation through encoder-decoder frameworks paired with self-attention, Large Language Models (LLMs) that utilize large sets of training data to generate text at scale. Finally, GAIs can be classified as being Unimodal (receiving one data format) or Multimodal (receiving data from multiple formats and analyzing input cohesively [ex. GPT-4]) (Bengesi et al., 2024). These variances in GAI models offer the process of constructing GAI tools in healthcare large amounts of specificity regarding function, input type, modality, and more, overall making GAI an extremely flexible tool that can fulfill a variety of niches in medical care. These general models can mainly be applied in mental healthcare as data synthesis and generation, image analysis, and multimodal data review can be utilized in various niches. In the next section, we will explore how these GAI frameworks can be applied specifically and as a whole throughout research, patient care, and post-intervention care.

Applications of Generative AI in Mental Healthcare

Clinical Drug Development and Research

GAI frameworks can be applied in clinical research settings in mental healthcare to improve the efficiency and quality of research. One of the most significant forms of clinical care in mental healthcare is Pharmacotherapy (Huhn et al., 2014), an area with multiple potential entry points for GAI tools. In many other fields of clinical research, GAI has already been utilized successfully in various capacities. For example, DrugGPT is a novel tool that utilizes GPT-2 models to design ligands that match specific proteins in its over 1.8 million protein training database, creating a possible pharmacological tool for creating medications that take advantage of ligand-protein pathways (Li et al., 2023). These functions are also achievable with Generative Adversarial Networks rather than Transformer-based models, in which case analysis of the ligand-binding site of various receptor proteins can result in novel drug design (Skalic et al. 2019). Simultaneously, the use of GAI for creating synthesis instructions for various molecules, as Stanford scientists did with SyntheMol to create de novo antibiotic medication, allows for designs created by GAI to be laboratory synthesized (Swanson et al., 2024). The combination of these two techniques in mental healthcare can design receptor-specific ligands to treat a particular mental health condition while creating the specific directions to create the molecule, excising a time-consuming and costly portion of the drug development process.

These methods have the capability of completely transforming the field of Pharmacopsychiatry, in which medication changes are critical to improving patient intervention outcomes through specializing prescriptions to a patient's needs and responses to other medications (Angell & Bolden, 2015), and therefore, the utilization of GAI to design and build new drugs will only increase the variety from which Psychiatrists can utilize to offer more effective, specialized care. Additionally, these tools can reduce the immense costs of drug development, especially considering updated estimates that include failed drug developments, only increasing such expenditures (Sertkaya et al., 2024). As GAI makes more efficient drug development systems through this generative ligand-receptor modeling, these costs are systematically reduced while final products become increasingly influential as GAI strengthens through training.

Already, AI has been successfully utilized to create an anti-fibrotic medication called INS018_055 to treat Idiopathic pulmonary fibrosis, and it has reached Phase II double-blind clinical trials (Ren et al., 2024). The nascent nature of these drugs, however, indicates a lack of further drug development reaching such late stages of testing. Still, the promise which these drugs show significant promise for the development of drugs on GAI frameworks in a multitude of fields, and it is only a matter of time before these frameworks are utilized for successful medication discovery in the field of mental health.

Clinical Trials with GAI Organizational Elements

Another significant sector in which Generative AI can improve the course of research is through facilitating the course of mental health drug clinical trials. Clinical drug trials themselves constitute a significant sector of drug development where costly production and testing contribute to the exorbitant costs of various medications, which thereby limit the accessibility of the developed drug while simultaneously creating a barrier to the initiation of clinical trials for crucial medication development (Stewart et al., 2015). Through GAI improvements to the organization and conduct of clinical drug trials, these costs can be mitigated, and the development of target medications, especially those in mental healthcare, can be much faster.

GAI has already shown considerable promise through its ability to improve the efficacy of clinical trials. One method by which this has been accomplished is through the integration of Generative AI in participant and patient recruitment, one of the more time-consuming tasks required in clinical trial organization. Trial Pathfinder, an AI system designed to analyze the risk of clinical trial cohorts based on a multitude of eligibility criteria, has been championed as a method to overcome the barrier of recruitment by reducing the strictness of eligibility criteria to allow a broader range of participants without exceeding a significant chance of risk (Liu et al., 2021). Methods augmenting this aspect of patient recruitment can not only avoid the time cost of searching for participants that fit narrow eligibility for clinical trials but also reduce the material costs accompanying it.

Similarly, the optimization of clinical drug trials through analysis of its various characteristics is a multifaceted approach to ensure success during trials. GAIs, like the Sequential Predictive Modeling of Clinical Trial Outcome (SPOT), utilize information about the medication being tested, eligibility data for patient recruitment, and the disease of interest, comparing these characteristics to prior clinical trial designs (from the AI's training data) to analyze the statistical risk and success likelihood of the trial being designed (Wang et al., 2023). These methods of comprehensive clinical trial analysis allow trial design to become increasingly streamlined and less risky, again improving the cost-effectiveness of clinical trials in ways that can benefit both researchers and patients in the long term.

Most recently, GAIs have also found a possible niche in trials themselves, improving patient retention and overall patient experience and monitoring. ClinicalGAN is a generative model that uses GAI to create a digital twin for a patient involved in a clinical trial, creating timelines that represent patient chronological progression within a clinical trial (Chandra et al., 2024). A lack of patient retention is one primary reason that clinical trials can fail, ultimately inefficiently using the resources invested in some trials and further reducing overall trial efficacy (Kearny et al., 2018). Through the development of patient retention strategies that are inspired by new GAI frameworks within the construction of Clinical trials, this possibility of failure can be mitigated, and within the field of mental health, can result in the development of novel care becoming much more common, therefore facilitating innovation.

Medical Documentation Improvements

Independently of research contexts, GAI has prominent potential in the clinical aspect of mental health interventions to make the work of medical professionals more efficient. For example, GAI has already shown promise in medical documentation processes. When analyzed in professional, though nonmedical, writing contexts, GAI increased productivity by 40% and the quality of writing by 18% (Noy, Zhang, 2023). When applied to medical documentation during medical interventions, these bumps in productivity and reduction to workload are relatively replicated, with emergency pediatric physicians reporting a similar 40% reduction in time required to create clinical summaries, and these summaries were rated fairly well regarding quality by these physicians (Barak-Corren et al., 2024). As mentioned in the study by Barak-Corren et al., physicians often spend more time during clinical documentation than during actual patient interventions, limiting the time available for these patient interactions and reducing overall efficiency while also contributing to physician burnout (which itself has compounding detriments to healthcare delivery). In the past, these downfalls have been combatted with medical scribes, but when analyzing their efficacy in a similar context to the Barak-Corren study, their improvements regarding efficiency (considering improvements in patients seen) were marginal (Ullman et al., 2021)l. As such, the inclusion of Generative AI in the creation of medical documentation can, for various purposes, improve overall medical efficiency.

Another concern in relation to clinical documentation is the transition of medical records from physical records (limited mainly to readers who are medical professionals themselves) to online records that are also viewed by patients, creating an additional necessity for medical records to be comprehensible to the average layperson (Blease et al., 2024). As noted by Blease and others, the use of GAI to address this issue can improve clarity issues when creating summaries for patients, reduce the time required for this purpose, and promote better care. Overall, as records continue to become available online and therefore, require a level of comprehensibility for patients, the use of GAI can further remove the burden of care providers to consistently create medical documentation instead of spending time on patient interaction. As the shortage of mental healthcare professionals grows, their time becomes increasingly valuable, and through these methods of prioritizing their time and reducing burnout, they can become better equipped to center patients within the care they provide (Hoffman et al., 2023).

GAI-Powered Diagnosis in Mental Healthcare Interventions

Of course, diagnosis within the mental health space is one of the most contentious issues, creating ethical, legal, scientific, and sociological debates about its viability within the future of mental healthcare. Considering the ability of Generative AI to recognize patterns within data and, therefore, possibly within human behavioral and physiological data, GAI could link these patterns to mental health diagnoses. Already, there are multiple instances where GAI is being tested for diagnosis, including an IBM speech classifier that can detect mental illness with 80% accuracy, as well as computer vision that can detect ADHD and ASD with 96% accuracy (Garg, Patil, 2021). These various areas where GAI diagnosis is being tested replicate previous developments of algorithmic analysis of patients' traits, only expanding on the processing capabilities of GAI to form efficient diagnosis tools to assist medical professionals.

Generative AI also has the power to assist with diagnosis through its analysis of neural imaging. As mentioned previously, GAI can analyze, report, and quantify minute differences within medical imaging that may be missed in other forms of image analysis. As such, GAI could be integrated into neurological biomarker analysis that links neurology to psychological symptoms, allowing psychiatrists to better prescribe medication tailored to patients (Wright, Anticevic, 2024). Similarly, a study by Saadatinia and Salimi-Badr shows that GAI can analyze electroencephalograms using GAN and VAE models to diagnose Schizophrenia with around 99% accuracy (Saadatinia, Salimi-Badr, 2024). In this way, GAI could be utilized in mental healthcare diagnosis to process biological information that can be indisputably linked to psychological conditions.

Even further, GAI can utilize existing personality diagnostics to monitor and designate patient behavior to a variety of personality characteristics. A study by Jenifa et al. describes the use of a Generative AI personality recognition model (Generative Artificial Intelligence based Learning Principles) that can recognize personality traits and match them to MBTI personality types with 97% accuracy (Jenifa et al. 2024). Considering the importance of personality assessments and diagnostics in the diagnosis of a variety of psychological conditions, including personality disorders, and also understanding the often lengthy process of diagnostics required to come to a diagnosis, the assistance of AI in completing these tasks can significantly reduce time to diagnosis, and therefore reduced time to treat and improved prognosis (Government of Canada, 2006). The diagnostic capabilities of GAI are similarly replicated in cases of depression, with automated diagnostic chatbots being a popular proposed use-case of GAI to offer treatment and psychiatric interventions with fewer delays, preventing tragedies including suicide and self-harm (Yang, Mori, 2024). Overall, the possible implementations of GAI in recognizing behavioral or biological patterns linked to psychological conditions offer many ways for medical professionals to more accurately and effectively formulate treatments for patients, and the mere variety of GAI applications in this respect makes diagnosis a prime area in which these novel technologies will be integrated.

Therapeutic Applications of GAI Accompanying Human-led Interventions

Diagnosis is not the only way GAI can be integrated into psychiatric clinical environments, and many medical professionals and biomedical researchers are investigating the possible methods by which GAI can accompany human-led mental health interventions. One method is through the integration of Medical Chatbots, a tool through which patients can gain 24-hour health support and answers to medical queries that they may need to ask when a medical professional isn't available, a technology already readily integrated into various medical fields (Tjiptomongsoguno et al., 2020). In the field of mental health, one common approach is to implement therapeutic elements into chatbots to help cope with symptoms of various mental conditions, such as the creation of Chatbots to provide aspects of Cognitive-Behavioral therapy to assist in dealing with emotional regulation (e.g., SERMO) (Denecke et al., 2020). Chatbots also have applications for conversational health education to help direct patient behavior outside clinical contexts and conduct patient surveys in contexts where seeking in-person help has a surrounding stigma (Vaidyam et al., 2019). As chatbots expand within the world of mental healthcare, implementing GAI into novel chatbot designs is a prominent possible innovation due to the prevalence of chat GAIs like Chat-GPT. These AI systems are also showing abilities to replicate empathy within a conversation, create responses that are more personalized to the user (in comparison to non-GAI chatbots), update more frequently without changes in programming to remain up-to-date and relevant, and deliver concise information that is comprehensible for patients at home (Abbasian et al., 2024). Through these capabilities, GAI could make the development of chatbots more streamlined while increasing their capabilities, offering yet another tool for professionals to utilize to provide accessible treatment.

AI image generation can also provide some utility within the therapeutic space. Though the common idea of image use in psychological contexts is often limited to Rorschach tests, images are common elements of a variety of mental health therapies. Within CBT, ACT, and DBT, among others, images can deepen conversations about behavior and emotion, especially when patients are younger and have less emotional intelligence or ability to convey emotion and therefore visuals are a prominent method of improving communication, education, and personalization within mental healthcare (Sezgin, McKay, 2024). The use of GAI to create these images could simplify integrating multiple modalities within care, creating pathways for mental health professionals to experiment with such modalities to personalize interventions further.

Another increasingly popular form of therapy is Music Therapy, in which music is integrated into therapeutic interventions to increase the efficacy of treatment for a variety of emotional disorders and psychological conditions while also improving motivation to engage in therapy (Gold et al., 2013). As GAIs become more advanced, their ability to create novel compositions by replicating melodic and rhythmic patterns increases their functionality (Louie et al., 2020). When utilized to specifically analyze therapeutic music for training and, consequently, create similar therapeutic music, GAI shows promise in both creating compositions and formulating suggestions for therapeutic music, overall increasing personalization (Hou, 2022). These implementations not only help further the usability of novel and niche forms of therapy but also reduce the time and resources required to experiment with their use in a patient's mental health intervention, consider its benefit, and reevaluate if necessary.

Ultimately, GAI's varied capabilities only assist in allowing mental health professionals to utilize technological tools to increase the accessibility, customization, affordability, and efficacy of their care. As more obscure yet crucial forms of therapeutic interventions access GAI tools to accompany them, overall patient outcomes can improve drastically.

Limitations of Generative AI in Mental Healthcare Applications

The Effects of Black-Box Modeling in GAI Creation

The majority of Generative AI can be described as a "black-box" model, which refers to how GAI processes information using extremely convoluted processing based on complex programming that prevents any significant understanding of the pathways that result in a response being created (Hassija et al. 2024). In short, the input into GAI and the output as a result of the input are known, but the process by which the output is generated is obscured. As mentioned by Hassija and others, the most apparent issue with this model is the lack of transparency in understanding how inaccurate outputs are created since it is unclear where misinformation or fallacious information was analyzed and utilized when processing the input. These issues are compounded by the frequent occurrence of misinformation being incorporated into responses created by various GAI chatbots, such as OpenAI's ChatGPT, which has been shown to include inaccuracies in a variety of topics based on the prompt given (Arvanitis et al., 2023). As such, when utilized in medical contexts, GAI informational outputs may include or be based on incorrect information while seeming legitimate. Additionally, trust put into GAI may result in such outputs being utilized with no skepticism or care, making medical harm caused by such use a real possibility.

 Volume 13 Issue 4 (2024)

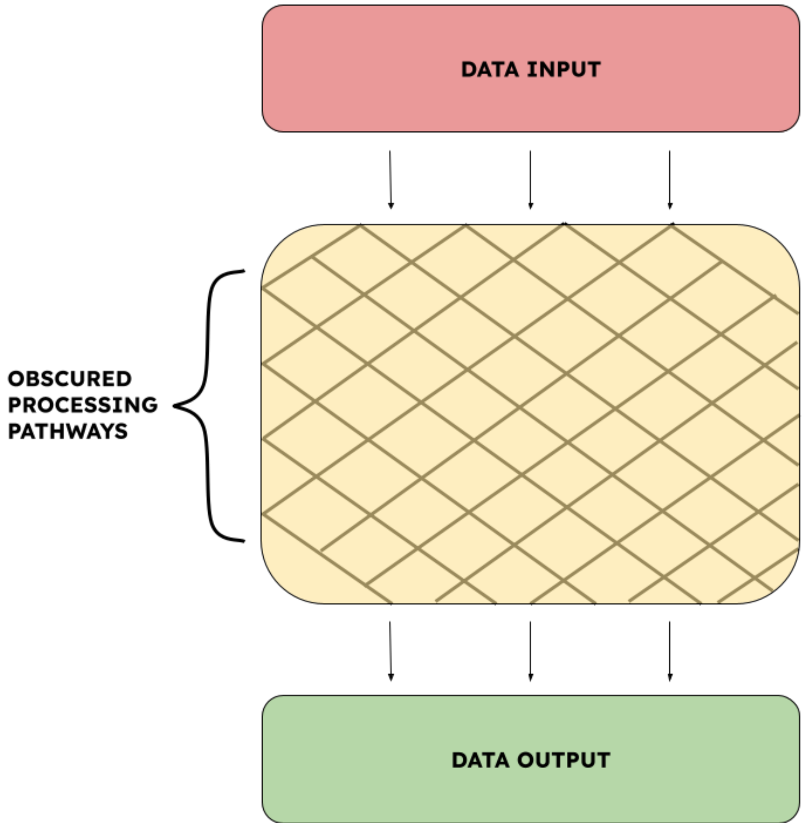

Figure 1. A visual representation of the black-box model, depicting the visible input and output and the obscured processing that occurs in between.

More specifically, the "black-box" model can perpetuate discriminatory biases (racial, gender-based, etc.) that exist within training data or are held by the creators of a GAI model. Biases within medical models and algorithms are already well documented, such as one widely-used risk-calculation algorithm that often rated Black patients and patients of color as "riskier" and, therefore, more likely to increase costs for the healthcare provider, hence justifying withholding care (Obermeyer et al., 2019). These algorithms are built off data analyzed by humans and are programmed by humans, creating many entry points for bias that ultimately result in withholding life-saving care. When these issues are incorporated into the creation of GAIs, though they themselves cannot create discriminatory bias, train by analyzing data selected by humans, finding patterns of racial bias, and incorporating such bias into their decision-making processes (Haber et al., 2024). These concerns can also accompany the concerns of inaccurate results being the result of erroneous training data. When existing GAI models utilize publicly contributed sources, both inaccurate information and biases can be incorporated into the model itself (King, 2023). Therefore, the utilization of GAI models can be especially hazardous if these biases are not thoroughly avoided. If not adequately tested, their use can endanger many people from marginalized communities when they need mental healthcare.

Of course, conscious bias against discriminated groups is not the only form of bias that can put GAI at risk of causing harm. Statistical bias, that is, statistical data that can show abstract patterns that, in some way, display some group as a detriment (Engstrom et al., 2020), can also be incorporated into GAI. Already, medical algorithms analyze medical data (including symptom frequency, prognosis, treatment course success, and more across a variety of age, gender, and racial markers) to formulate responses that can replicate discriminatory statistical bias. Pfob and Sidey-Gibbons describe a case where data that indicated black women having worse medical outcomes compared to white women (because of preexisting discriminatory conditions) led to algorithmic bias recommending such treatment plans less (due to their lowered efficacy) (Pfob, Sidey-Gibbons, 2022). In the case of algorithms, these patterns in data are analyzed without the appropriate context to explain them, resulting in an algorithm that perpetuates discrimination against vulnerable groups. Because GAI follows a similar system of pattern analysis to create responses in a variety of medical and mental healthcare applications, this risk of statistical bias being furthered by the decisions of a GAI (even if data used to train the GAI isn't purposefully skewed) is something that ought to be considered in the creation of GAI tools.

Data Privacy and Security Risks

One of the most significant apprehensions to integrating Generative AI in a multitude of fields is the lack of security and data privacy that users can face. If existing GAI models are used, since many developers continue to use inputted data to train models, sensitive data input into a GAI tool could be absorbed, utilized, and, under certain conditions, included in outputs, revealing the sensitive data inputted (Shi, 2023). This risk is especially perilous if medical professionals input HIPPA-protected information, creating the possibility that patient confidentiality will be broken. This, however, only considers the risk of an accidental data leak due to the design of a GAI tool. Malicious actors can utilize a multitude of strategies to access patient data, whether it be through feeding specific prompts to AI to reveal data or through a direct breach of training datasets or stored input data (Chen, Esmaeilzadeh, 2024).

Data protection is also only one of the many security risks in GAI development. More insidious measures that can be taken can be the insertion of fake (poison) data into GAI training data to cause medical harm, utilize the characteristics of GAI to gain access to sensitive bioinformatic or health information of significant individuals, and generally cause damage through the tweaking of AI in ways that can cause harm to patients (Harrer, 2023). Typically, security risks and possibilities of data leaks are an issue that plagues the entirety of healthcare systems that have any sort of online or cloud-based storage, but the increased pathways by which medical information can be leaked with GAI or GAI tools can be corrupted, the safety of their use is easily an essential concern for its use in any field of health, including mental healthcare.

The Importance of Human Connection

Human interaction within the field of mental healthcare is an integral facet of proper care, and adequate bedside manner can be shown through effective communication, which provides patients confidence in their provider and the healthcare they receive (Parnas, Isobel, 2017). When connected with the understanding that almost 60% of Americans would not be comfortable or confident if their healthcare provider relied on Artificial Intelligence to deliver healthcare (Tyson et al. 2023), it becomes evident that there is a gap between the potential reach that GAI can have in mental healthcare and the willingness that patients have to integrate it. One reason for this could be the lack of understanding that GAI can have when analyzing subtleties within complex, varied, and mutable human emotions, a necessary aspect of mental health therapy (Vial, Almon, 2023). In this sense, the worry behind GAI capabilities within healthcare does not lie in its accuracy but in its ability to display an essential empathetic connection within the therapeutic space.

The sacredness of the therapeutic relationship is another consideration in integrating Generative AI, as many people believe the relationship between patient and doctor would be at risk due to overbearing AI use (Sezgin, 2023). In this sense, human connection matters even more in mental healthcare than other health fields. Because of the therapeutic alliance's basis in trust and compassion, there is minimal adaptability present in current-day AI systems that can replicate such levels and complexities of emotion (Joseph, Babu, 2024). Overall, the mere therapeutic effect that human-led understanding can have within mental health fields is one of the reasons why mental health interventions are effective, and the risk of that effect's removal can make many patients uneasy at the thought of GAI-led therapeutic interventions.

Conclusion

As our technological capabilities improve over time, it is only natural that the innovation they bring is implemented within the scientific and clinical space. GAI proves to be no exception, as its varied capabilities offer it multiple points of entry in the field of mental healthcare (Marr, 2024). Considering the rapid pace at which innovations occur in the field of GAI and the integration of GAI within healthcare, while also considering the often limited scope of other reviews, we hoped to provide an up-to-date survey of current discoveries to best direct future research and integration. In this paper, we have explored the transformative applications of GAI in Mental Healthcare, noting its existing impact within these fields while expanding upon its overall potential.

The impact of GAI's novel and more efficient capabilities in comparison to existing medical technologies makes it a key innovation that could revolutionize the conduct of research and the delivery of mental healthcare interventions (Huhn et al., 2014; Sardesai et al., 2024). In clinical and biomedical research specifically, GAI can revolutionize Pharmacopsychiatry by reducing the time and resources required to design novel medications (Swanson et al., 2024). It can also improve clinical trial design to retain participants better, reduce overall financial risk, and improve overall trial success (Wang et al., 2023). In clinical settings, GAI also shows the potential to increase the efficiency of interventions by creating medical documentation in real-time (Barak-Corren et al., 2024), allowing professionals to spend time that would be spent formulating patient records instead of interacting with patients. GAI can also contribute to psychiatric diagnosis formulation by analyzing behavioral patterns (Garg, Patil, 2021) while providing some accompanying measures to professional-led therapy (Tjiptomongsoguno et al., 2020). Overall, if implemented, these various GAI applications could make these integral processes within mental healthcare less time-consuming, more resource-efficient, less costly, and more beneficial for patient prognoses.

However, the immediate integration of GAI into these settings faces some barriers. Black-box models that are representative of existing GAI models allow decisions made by these models to face little accountability, allowing cognitive and statistical bias within these systems to cause harm (Hassija et al. 2024). Data privacy risks caused by the design of GAI could result in confidential information being leaked, whether by accident or through malice, posing possible harm to patients (Shi, 2023). Finally, the integrity of mental healthcare lies within the therapeutic benefits of human connection, a facet that could become less prioritized or overlooked if GAI completely supersedes human-led interventions (Parnas, Isobel, 2017). These aspects, however, can be avoided if GAI models are implemented correctly. Through proper human oversight, precise design, appropriate transparency, and careful use that centers patients, GAI can be responsibly utilized within mental healthcare.

Ultimately, GAI, like the technological innovations that preceded it, will inevitably be integrated into the medical field to many degrees. Its capacity to improve the situations of patients, doctors, researchers, and entire healthcare systems must be addressed as AI continues to develop. As medical professionals continue to investigate its possible capabilities, roles, and limitations, GAI in mental healthcare will arrive in the near future. Overall, we hope that its implementation leads to a healthier world.

Acknowledgments

I would like to express my gratitude to my advisor, Nicholas Traugott, for his guidance during the creation of this project and for his commitment to enriching me and encouraging me to explore the subject matter of this paper. Additionally, I would like to thank my parents for their continuous support.

References

Abbasian, M., Khatibi, E., Azimi, I. *et al.* Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative AI. *npj Digit. Med.* 7, 82 (2024). https://doi.org/10.1038/s41746-024-

01074-z

Angell, B., & Bolden, G. B. (2015). Justifying medication decisions in mental health care: Psychiatrists' accounts for treatment recommendations. *Social science & medicine (1982)*, *138*, 44–56. https://doi.org/10.1016/j.socscimed.2015.04.029

Arvanitis, L., Brewster, J., & Sadeghi, M. (2024, July 15). *NewsGuard's misinformation monitor: GPT-4 produces more misinformation than predecessor*. NewsGuard. https://www.newsguardtech.com/misinformation-monitor/march-2023/

Bandi, A., Adapa, P. V., & Kuchi, Y. E. (2023). The power of Generative AI: A review of requirements, models, input–output formats, evaluation metrics, and challenges. *Future Internet*, *15*(8), 260. https://doi.org/10.3390/fi15080260

Barak-Corren, Y., Wolf, R., Rozenblum, R., Creedon, J. K., Lipsett, S. C., Lyons, T. W., Michelson, K. A., Miller, K. A., Shapiro, D. J., Reis, B. Y., & Fine, A. M. (2024). Harnessing the power of Generative AI for clinical summaries: Perspectives from emergency physicians. *Annals of Emergency Medicine*, *84*(2), 128–138. https://doi.org/10.1016/j.annemergmed.2024.01.039

Bengesi, S., El-Sayed, H., Sarker, M. K., Houkpati, Y., Irungu, J., & Oladunni, T. (2024). Advancements in generative AI: A comprehensive review of gans, GPT, autoencoders, diffusion model, and Transformers. *IEEE Access*, *12*, 69812–69837. https://doi.org/10.1109/access.2024.3397775

Blease, C., Torous, J., McMillan, B., Hägglund, M., & Mandl, K. D. (2024). Generative language models and open notes: Exploring the promise and limitations. *JMIR Medical Education*, *10*. https://doi.org/10.2196/51183

Chandra, S., Prakash, P.K.S., Samanta, S. *et al.* ClinicalGAN: powering patient monitoring in clinical trials with patient digital twins. *Sci Rep* 14, 12236 (2024). https://doi.org/10.1038/s41598-024-62567-

Chen, Y., & Esmaeilzadeh, P. (2024). Generative AI in medical practice: In-depth exploration of privacy and security challenges. *Journal of Medical Internet Research*, *26*. https://doi.org/10.2196/53008

Christensen, M., Vukadinovic, M., Yuan, N. *et al.* Vision–language foundation model for echocardiogram interpretation. *Nat Med* 30, 1481–1488 (2024). https://doi.org/10.1038/s41591-024-02959-y

Denecke, K., Vaaheesan, S., & Arulnathan, A. (2021). A mental health chatbot for regulating emotions (sermo) - concept and usability test. *IEEE Transactions on Emerging Topics in Computing*, *9*(3), 1170–1182. https://doi.org/10.1109/tetc.2020.2974478

Engstrom, L., Ilyas, A., Santurkar, S., Tsipras, D., Steinhardt, J. & Madry, A.. (2020). Identifying Statistical Bias in Dataset Replication. *Proceedings of the 37th International Conference on Machine Learning*, in *Proceedings of Machine Learning Research* 119:2922-2932 Available from https://proceedings.mlr.press/v119/engstrom20a.html.

Feuerriegel, S., Hartmann, J., Janiesch, C. *et al.* Generative AI. *Bus Inf Syst Eng* **66**, 111–126 (2024). https://doi.org/10.1007/s12599-023-00834-7

Garg, M., & Patil, H. S. (2021). Comparative Study On AI in Mental Health and Wellbeing-Current Applications and Trends. *International Research Journal Of Engineering And Technology*.

Gemini. Google DeepMind. (2024, May 14). https://deepmind.google/technologies/gemini/

Gold, C., Mössler, K., Grocke, D., Heldal, T. O., Tjemsland, L., Aarre, T., Aarø, L. E., Rittmannsberger, H., Stige, B., Assmus, J., & Rolvsjord, R. (2013). Individual music therapy for mental health care clients with low therapy motivation: Multicentre Randomised controlled trial. *Psychotherapy and Psychosomatics*, *82*(5), 319–331. https://doi.org/10.1159/000348452

Government of Canada. (2006). *The human face of mental health and mental illness in Canada, 2006*. Public Health Agency of Canada.

Haber, Y., Levkovich, I., Hadar-Shoval, D., & Elyoseph, Z. (2024). The artificial third: A broad view of the effects of introducing Generative Artificial Intelligence on psychotherapy. *JMIR Mental Health*, *11*. https://doi.org/10.2196/54781

Harrer, S. (2023). Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine. *eBioMedicine*, *90*, 104512. https://doi.org/10.1016/j.ebiom.2023.104512

Hassija, V., Chamola, V., Mahapatra, A. *et al.* Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cogn Comput* **16**, 45–74 (2024). https://doi.org/10.1007/s12559-023-10179-8

Hoffmann, J. A., Attridge, M. M., Carroll, M. S., Simon, N.-J. E., Beck, A. F., & Alpern, E. R. (2023). Association of Youth Suicides and county-level mental health professional shortage areas in the US. *JAMA Pediatrics*, *177*(1), 71. https://doi.org/10.1001/jamapediatrics.2022.4419

Hou, Y. (2022). Ai Music therapist: A study on generating specific therapeutic music based on deep generative adversarial network approach. *2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*. https://doi.org/10.1109/icetci55101.2022.9832398

Huhn, M., Tardy, M., Spineli, L. M., Kissling, W., Förstl, H., Pitschel-Walz, G., Leucht, C., Samara, M., Dold, M., Davis, J. M., & Leucht, S. (2014). Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders. *JAMA Psychiatry*, *71*(6), 706. https://doi.org/10.1001/jamapsychiatry.2014.112

Jenifa, G., K, Padmapriya., P, S., Karthika, K., Pandi, V. S., & Arumugam, D. (2024). An effective personality recognition model design using Generative Artificial Intelligence Based Learning principles. *2024 International Conference on Computing and Data Science (ICCDS)*. https://doi.org/10.1109/iccds60734.2024.10560368

Joseph, A.P., Babu, A. The unseen dilemma of AI in mental healthcare. *AI & Soc* (2024). https://doi.org/10.1007/s00146-024-01937-9

Kalman, J. L., Burkhardt, G., Samochowiec, J., Gebhard, C., Dom, G., John, M., Kilic, O., Kurimay, T., Lien, L., Schouler-Ocak, M., Vidal, D. P., Wiser, J., Gaebel, W., Volpe, U., & Falkai, P. (2023). Digitalising mental health care: Practical recommendations from the European Psychiatric Association. *European psychiatry : the journal of the Association of European Psychiatrists*, *67*(1), e4.

https://doi.org/10.1192/j.eurpsy.2023.2466

Kearney, A., Rosala- Hallas, A., Bacon, N., Daykin, A., Shaw, A. R., Lane, A. J., Blazeby, J. M., Clarke, M., Williamson, P. R., & Gamble, C. (2018). Reducing attrition within clinical trials: The communication of retention and withdrawal within patient information leaflets. *PLOS ONE*, *13*(10). https://doi.org/10.1371/journal.pone.0204886

King, D. (2023). ChatGPT Not Yet Ready for Clinical Practice. *Psychiatric News*, *58*(07). https://doi.org/10.1176/appi.pn.2023.07.7.56

Li, Y., Gao, C., Song, X., Wang, X., Xu, Y., & Han, S. (2023). DrugGPT: A GPT-based strategy for designing potential ligands targeting specific proteins. *bioRxiv*, 2023-06.

Liu, R., Rizzo, S., Whipple, S. *et al.* Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature* 592, 629–633 (2021). https://doi.org/10.1038/s41586-021-03430-5

Louie, R., Coenen, A., Huang, C. Z., Terry, M., & Cai, C. J. (2020). Novice-ai music co-creation via AI-steering tools for deep generative models. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376739

Marr, B. (2024, July 2). *How generative AI is Revolutionizing Customer Service*. Forbes. https://www.forbes.com/sites/bernardmarr/2024/01/26/how-generative-ai-is-revolutionizing-customer-service/

Microsoft copilot: Your everyday ai companion. Microsoft Copilot: Your everyday AI companion. (2024). https://copilot.microsoft.com/

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science (New York, N.Y.)*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

OpenAI API. OpenAI. (2024). https://platform.openai.com/

Parnas, S., & Isobel, S. (2017). Navigating the social synapse: The Neurobiology of Bedside Manner. *Australasian Psychiatry*, *26*(1), 70–72. https://doi.org/10.1177/1039856217726224

Pfob, A., & Sidey-Gibbons, C. (2022). Systematic bias in medical algorithms: To include or not include discriminatory demographic information? *JCO Clinical Cancer Informatics*, (6). https://doi.org/10.1200/cci.21.00146

Ren, F., Aliper, A., Chen, J. *et al.* A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. *Nat Biotechnol* (2024). https://doi.org/10.1038/s41587-024-02143-0

Rothman, J. (2024, February 16). *When A.I. can make a movie, what does "video" even mean?* The New Yorker. https://www.newyorker.com/science/annals-of-artificial-intelligence/when-ai-can-make-a-movie-what-does-video-even-mean

Saadatinia, M., & Salimi-Badr, A. (2024). An explainable deep learning-based method for schizophrenia diagnosis using generative data-augmentation. *IEEE Access*, *12*, 98379–98392. https://doi.org/10.1109/access.2024.3428847

Sardesai, N., Russo, P., Martin, J., & Sardesai, A. (2024). Utilizing generative conversational artificial intelligence to create simulated patient encounters: A pilot study for anaesthesia training. *Postgraduate Medical Journal*, *100*(1182), 237–241. https://doi.org/10.1093/postmj/qgad137

Sertkaya, A., Beleche, T., Jessup, A., & Sommers, B. D. (2024). Costs of drug development and research and development intensity in the US, 2000-2018. *JAMA Network Open*, *7*(6). https://doi.org/10.1001/jamanetworkopen.2024.15445

Sezgin, E. (2023). Artificial Intelligence in Healthcare: Complementing, not replacing, doctors and healthcare providers. *DIGITAL HEALTH*, *9*. https://doi.org/10.1177/20552076231186520

Sezgin, E., McKay, I. Behavioral health and generative AI: a perspective on future of therapies and patient care. *npj Mental Health Res* 3, 25 (2024). https://doi.org/10.1038/s44184-024-00067-w

Shi, Y. (2023). Study on Security Risks and legal regulations of Generative Artificial Intelligence. *Science of Law Journal*, *2*(11), 17–23. https://doi.org/10.23977/law.2023.021104

Skalic, M., Sabbadin, D., Sattarov, B., Sciabola, S., & De Fabritiis, G. (2019). From target to drug: Generative Modeling for the multimodal structure-based ligand design. *Molecular Pharmaceutics*, *16*(10), 4282–4291. https://doi.org/10.1021/acs.molpharmaceut.9b00634

Sora. OpenAI. (2024b). https://openai.com/sora

Stewart, D. J., Batist, G., Kantarjian, H. M., Bradford, J.-P., Schiller, J. H., & Kurzrock, R. (2015). The urgent need for clinical research reform to permit faster, less expensive access to new therapies for lethal diseases. *Clinical Cancer Research*, *21*(20), 4561–4568. https://doi.org/10.1158/1078-0432.ccr-14-3246

Swanson, K., Liu, G., Catacutan, D.B. *et al.* Generative AI for designing and validating easily synthesizable and structurally novel antibiotics. *Nat Mach Intell* 6, 338–353 (2024). https://doi.org/10.1038/s42256-024-00809-7

Teng, M. (2023, December 2). *World's first AI generative bispecific T cell engager with picomolar affinity*. Etcembly. https://www.etcembly.com/post/etcembly-reveals-world-s-first-ai-generative-bispecific-t-cell-engager-with-picomolar-affinity

Tjiptomongsoguno, A.R.W., Chen, A., Sanyoto, H.M., Irwansyah, E., Kanigoro, B. (2020). Medical Chatbot Techniques: A Review. In: Silhavy, R., Silhavy, P., Prokopova, Z. (eds) Software Engineering Perspectives in Intelligent Systems. CoMeSySo 2020. Advances in Intelligent Systems and Computing, vol 1294. Springer, Cham. https://doi.org/10.1007/978-3-030-63322-6_28

Tyson, A. (2023, February 22). *60% of Americans would be uncomfortable with provider relying on AI in their own* *health care*. Pew Research Center. https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/

Ullman, K., McKenzie, L., Bart, B., Park, G., MacDonald, R., Linskens, E., & Wilt, T. J. (2021). The effect of medical scribes in emergency departments: A systematic review. *The Journal of Emergency Medicine*, *61*(1), 19–28. https://doi.org/10.1016/j.jemermed.2021.02.024

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. *The Canadian Journal of Psychiatry*, *64*(7), 456–464. https://doi.org/10.1177/0706743719828977

Vial, T., & Almon, A. (2023). Artificial Intelligence in mental health therapy for children and adolescents. *JAMA Pediatrics*, *177*(12), 1251. https://doi.org/10.1001/jamapediatrics.2023.4212

Wang, Z., Xiao, C., & Sun, J. (2023). Spot: Sequential predictive modeling of clinical trial outcome with Meta-learning. *Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. https://doi.org/10.1145/3584371.3613001

Wright, S. N., & Anticevic, A. (2024). Generative AI for precision neuroimaging biomarker development in psychiatry. *Psychiatry Research*, *339*, 115955. https://doi.org/10.1016/j.psychres.2024.115955

Yadav, A. B. (2024). An Analysis on the Use of Image Design with Generative AI Technologies. *International Journal of Trend in Scientific Research and Development*, *8*(1), 596–599. https://www.researchgate.net/publication/378291870_An_Analysis_on_the_Use_of_Image_Design_with_Generative_AI_Technologies

Yang, M., & Mori, H. (2024). Automated diagnostics and its advantages of AI in mental health. *Lecture Notes in Computer Science*, 289–299. https://doi.org/10.1007/978-3-031-60107-1_21

Zhang, K., Riegler, G., Snavely, N., & Koltun, V. (2020). Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*.