Novel AI to avert the mental health crisis in COVID-19: Novel application of GPT2 in Cognitive Behaviour Therapy

ANONYMOUS (rajagopal.motivate@gmail.com)
Indian Institute of Technology

Research Article

Keywords: Deep Learning, mental health, GPT2, BERT, Natural Language Generation, mental health therapy, COVID-19

Posted Date: January 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-382748/v3

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

The effect of the COVID-19 pandemic on mental health is substantial. The World Health Organization has called for action to avert an impending mental health crisis. To respond to this call, this paper contributes a novel application of Deep Learning in Natural Language Generation (NLG) to seed healthy thoughts for mental health therapy.

For the 1st time in literature, a transfer learning capable large neural network with more than 100 million parameters for a NLG based mental health therapy application is proposed & demonstrated. This AI is designed to address scalable impact for millions of families with a timely health intervention in a privacy-safe approach. To the best of our knowledge, this is the first research paper to apply GPT2 (Generative Pretrained Transformer) for Cognitive Behavior therapy (CBT). Further, the paper demonstrates the proposed neural network architecture with a lab prototype implementation with reproducible results. This paper demonstrates this AI's ability to generate conditional synthetic human-like text intended to seed a healthy mental outlook. This is accomplished by fine tuning a pre-trained GPT2 language model. The source code and video demonstration is contributed at https://sites.google.com/view/ai-in-mental-health.

Also, for the 1st time in literature, a novel idea of NLU (Natural Language Understanding) activated NLG therapy is demonstrated with reproducible results using a BERT based classifier to activate the GPT2 based therapy. Performance of GPT2 models of three different sizes (124, 355, 774 million parameters) was the same for a very small dataset, thus a small GPT2 model is suggested for on-device AI inference.

This AI is a step forward in responding to WHO's call for action to avert the crisis. Towards addressing all the three dimensions of the monumental challenge, the paper designed a novel AI architecture by taking advantage of both BERT & GPT2. It also demonstrated the feasibility of Transformers-based AI for developing a mental health therapy solution. Further, this paper contributed an open-source AI prototype to support research communities to transform global mental wellness.

1. Introduction

Need: To avert the looming mental health crisis for millions of families

COVID-19 has a significant consequence on mental health across a vast population as per experts [1]. Mental health & mental wellness concern is emerging as a significant and urgent need for a vast majority of the world population as per a recently published report by the World Health Organization (WHO) [31]. The COVID-19 pandemic is causing monumental effects on mental wellbeing worldwide [32].

Experts call for an urgently deployable intervention, as per publications in JAMA [1] and Lancet [2, 32]. Recent editorials in reputed Nature Medicine [39] and Lancet [40] have highlighted the need for action as there is mounting evidence of a widespread impact of the pandemic on mental health. United Nations has recommended actions to ensure people and societies are better protected from the mental health impact of COVID-19 [3]. The Lancet position paper [2] calls for multi-disciplinary research priorities. Rising to this challenge will require integration across disciplines and sectors as per Lancet [2]. Given the need for timely intervention for millions of families, a scalable solution will be necessary. The opportunity and necessity to contribute to research to find solutions to address this challenge are immense, given the pandemic's vast impact [31]. This paper contributes to this gap and supports the research community to avert the crisis by contributing an open-source AI.

Challenge: Timely mental health intervention for millions of families

Fear about the perceived threat of COVID-19 infection, loss of a job, physical isolation from friends due to lockdown, quarantine of family members, closures of schools, fear of infecting loved ones by nurses caring for COVID-19 patients,
uncertain future, risks due to asymptomatic virus are affecting mental health [1, 2, 31, 32]. Millions of such families require early & timely health intervention [3].

Hence, the three dimensions of the challenge to be addressed by the proposed AI are

1. Scalable: The scale to impact millions of families, and
2. Timely: Early & timely intervention with 24x7 availability of help
3. Privacy: Protection of information on individual’s situations, her fears etc.

In Oct 2020, the Lancet editorial [32] said it is unclear how the world will deal with this forthcoming mental health crisis, given the shortage in the capacity to respond. WHO found [31] disruption of mental health services across many countries, especially for the vulnerable. With millions of families impacted by the COVID-19 pandemic, the shortage of capacity, and the criticality of providing timely health intervention, combined with the privacy requirements, make this challenge of averting the forthcoming mental health crisis into a monumental challenge. Also, privacy needs can't be ignored as mental health therapy often involves handling sensitive personal information.

As per WHO [38], 1 person dies from suicide every 40 seconds. Around 1 in 5 adolescents have a mental disorder as per WHO [38]. As per many editorials in Nature Medicine and Lancet Global Health journals, there is an increasing alert on the fallout of COVID-19 pandemic on global mental health [1, 2, 39, 40]. The United Nations reported that the mental health and mental wellbeing of societies have been severely impacted by COVID-19 [3]. The May 2020 editorial of Nature Medicine [39] called for researchers across multiple domains to address the impact of COVID-19 on the world’s mental health. The Nov 2020 editorial of Lancet Global Health [40] says 1 billion people suffer from a mental disorder. The Lancet editorial [40] also restates that sooner or later, health systems will be challenged to face a widespread demand for mental health due to COVID-19. By May 2020, United Nations published a policy brief [3] strongly urging the need for action on mental health. By Oct 2020, WHO calls the world to action on mental health. In short, multiple experts and leaders have forewarned about the impact of the pandemic on mental health and have called for action during the last few months.

As a response to this call for action, this paper contributes an AI to enable research communities to design a solution to avert the impending mental health crisis. Though this paper demonstrates an AI solution and develops a prototype implementation in the English language, future researchers may extend it to other languages, given the spread of COVID-19 across countries.

Existing literature and novelty in this paper

Novel application of GPT2 for CBT:

- There are not many research publications on the applications of the latest advancements in AI for mental health therapy. Specifically, the gap is in the application of state of the art NLG such as GPT2.
- While there is a lot of work in applications of state of the art NLU for mental health diagnosis [5], there is a gap in applying state of the art NLG for mental health therapy.
- While research papers had attempted therapy using NLG, old Deep Learning techniques were applied. However, the opportunity to apply the latest Deep Learning techniques remain untapped. The power of the latest advances such as GPT2 opens up the opportunity to create human-like English narratives. More significantly, the possibility of conditioning the language generated using transfer learning allows AI based therapy to help individuals. This breakthrough potential was untapped in the literature. This untapped opportunity offers a clear path to avert the looming mental health crisis from the COVID-19 pandemic. This paper presents this untapped opportunity and reports the conceptual advance by proposing an AI based architecture, its feasibility and a prototype with reproducible results, and an open-source contribution to help fellow researchers.
Very few research publications have explored the application of recent advances in NLG (Natural Language Generation) for mental health therapy. While the advancements in Deep Learning for NLG are staggering in recent years, the potential uses of applying these AI advancements to create solutions for mental health remains untapped. An advancement in the year 2019 in the field of Deep Learning based NLG is GPT-2 [25]. GPT-2 (Generative PreTrained Transformer) [25] is a powerful language model built using Attention mechanism [18] based Transformers [28]. GPT-2 is a language model that could perform natural language processing applications such as answering questions, completing text, reading comprehension, text summarization. GPT2 is capable of generating human-like language. It was demonstrated to create fake news [21] or to create poetry [15] or generate image captions [20]. This was possible due to transfer learning capability [13], recently made possible in Transformers architecture [25] based NLG, opening up the “imagenet moment in NLG”. However, there is a lack of research papers on the application of GPT-2 for mental health therapy, especially at a time when its application can transform the way the world responds to the titanic challenge of looming mental health crisis. This paper is an attempt to bridge this gap, which can lead to a breakthrough approach to avert the forthcoming crisis.

A search on Google Scholar for the keywords combination of “Cognitive Therapy” and “GPT2” shows feeble search results. A Google scholar keyword search on “Cognitive Behaviour therapy Deep Learning GPT2” or “mental health therapy Deep Learning GPT2” or “Cognitive therapy Deep Learning GPT2” or “CBT Transformers GPT2” shows a few research publications, and hence the gap identified is presented in Table 1.

To the best of our knowledge, this is the first research paper to apply latest AI approaches such as GPT2 in Cognitive Behavior Therapy [4]. This paper's work on this unexplored area opened the doors to an AI architecture that is capable of averting the future mental health crisis from COVID-19 pandemic. One of the mental health therapies is CBT or “Cognitive Behaviour Therapy” [6]. CBT is a therapy technique that can help people find new ways to behave by changing their thought patterns [7, 8, 9]. A novel NLG based Cognitive Behavior therapy model is proposed in this paper.

There is substantial research publications in NLU (Natural Language Understanding) such as BERT [10] for diagnosis of sentiment [34] or sensing of emotions [43]. However, there are only very few publications on NLG for mental health therapy. A survey paper [33] took stock of a decade of studies with 139 papers shows a lot of effort has happened on diagnosis, but there is significant scope for future research on novel applications of NLG for mental health. A paper in Nature scientific reports [34] also shows efforts on the topic of diagnosis. It shows progress in application of Deep Learning for classification problem statements to classify emotions/mental health conditions. However, the opportunity to apply Deep Learning in the synthesis using NLG is a relatively unexplored research theme in the context of mental health, and even more specifically in the context of CBT.

There are attempts to use NLG for therapy using old Deep Learning techniques such as LSTM. However, not many papers have explored the use of latest Deep Learning techniques for mental health therapy. Transformers [28] based language models represent the latest advancements in Deep Learning based NLG. While there are attempts to use NLG for therapy earlier, such as in year 2017 [35], not many research papers explored the latest advancements in NLG. Since GPT2 emerged only in 2019 [25], the opportunity to apply state of the art NLG models such as GPT2 has opened up. While 2019 saw the evolution of GPT-2, 2020 saw the introduction of a more powerful language model called GPT-3. Though GPT-3 [43] is a successor to GPT-2, it is not suited for edge AI due to its massive size. Since GPT-3 is only available as a cloud API requiring transmission of user's thoughts via the internet, it is not suitable to meet the privacy requirements in a mental health solution. Due to the power of transfer learning [13, 27] since the ‘Imagenet moment in NLP’ arrived, it is possible to produce a language model to suit a particular requirement such as poetry [15]. Transformers based NLP architectures are capable of generating AI performance as close to humans. For instance, it may be hard for humans to find if an AI generated a piece of fake news [21]. This power of AI to generate short text narratives close to human performance can be tapped into for designing novel health therapy solutions. The application of GPT2 like architectures for generating human-like narratives in mental health therapy solutions is a game-changing idea.
More scalable than existing approaches

- The need for a pandemic scale solution has been stated earlier.
- In contrast to cloud-based crawling of social media posts, the proposed on-device AI architecture opens the possibility of providing mental health solutions to a larger population.

Scalability is a significant need in the context of pandemic scale mental health crisis from COVID-19 for millions of families across the globe [2]. This paper makes a unique contribution by proposing an approach that is much more scalable than existing approaches. In addition, there is another unique dimension in this paper that enables a significantly larger percentage of the population to be offered mental health compared to approaches in the existing literature. The architecture proposed in this paper offers a leap in the addressable proportion of the world population. To the best of our knowledge, this is the first paper to propose a novel edge AI architecture to do mental health on the user's smartphone. This edge AI based architecture is presented in Figure 6. While most existing literature [5] focus on predicting mental health from publicly accessible social media posts, this limits the percentage of population to those netizens who are actively posting public tweets on personal sensitive thoughts. In contrast, this paper proposes an architecture where AI inference happens on the user's smartphone, thus allowing the opportunity to screen a larger population. Gboard by Google [14] is a downloadable predictive keyboard for smartphone users, as shown in the screenshot of Figure 4. Similar to AI embedded Google predictive keyboard on Android smartphones [14], which auto-corrects spelling errors as users type in words, the proposed approach to embed AI on smartphone will enable significant scalability.

A unique idea of intelligent activation of therapy

- The conceptual advance in “NLU activated GPT2 based therapy” is another unique aspect of the contributed AI architecture.

Once a BERT classifier detects a sequence of depressing thoughts on user's personal smartphone, the GPT2 predictive keyboard can be activated to provide self-help to weed off thoughts that can lead to habituation of depression. This is also first time in the literature where an AI based therapy is activated intelligently by another AI that keeps a tap on the candidate’s mental health. Once a BERT based classifier detects a sequence of depressing thoughts, this GPT2 predictive keyboard can be activated on the user’s personal smartphone to provide self-help to weed off thoughts that can lead to habituation depression. This novel AI architecture is presented in Figure 6.

More privacy safe than existing approaches

- While cloud chatbots based counseling has been attempted, the concern is in sending sensitive data on one's personal thoughts into a cloud based server. In contrast, this paper develops a privacy-safe approach from the grounds-up from designing an appropriate architecture such that one's sensitive data doesn't leave the boundaries of her personal smartphone.

Privacy safety design from the ground-up is a crucial requirement for digital mental health counseling solutions. The privacy safety of the proposed approach makes this approach to AI-based therapy further unique in the literature. While current attempts to counseling employ an AI chatbot [11, 19, 24, 41], the user’s private data is sent across the public internet to the cloud based chatbot engines. In contrast, in this paper’s proposed architecture, the AI inference happens on-device in the personal smartphone. Privacy safety aspect in mental health approach is illustrated in Figure 6. As per this figure, the user’s
sensitive personal information doesn’t leave her personal smartphone device's boundaries. The proposed solution is privacy safe compared to other approaches, and this is illustrated in the video in URL.

**A unique design to enable Early Intervention & Protection of sensitive data**

- Early intervention in mental health for millions is critical during a pandemic.
- Handling of sensitive personal data should be done with utmost care.
- There are limitations of employing cloud-based chatbot based approach. Cloud chatbots are not conducive for early intervention.
- In contrast to the cloud chatbots based approach, on-device AI inference is proposed. Adding a mental health AI to a smartphone's keyboard allows for widespread deployment and early intervention at scale to aver the pandemic scale crisis.

The significance of early intervention in mental health is stated in JAMA [44]. The significance of proactive and early intervention can't be understated [44]. Hence this limits the adoption of chatbots. Chatbots [11] require initiation by user base, limiting early intervention & scalable user adoption. Given the scalability needed to activate mental health wellness for millions of families, a chatbot may not yield user adoption. In the early stages of depression, the affected individuals may not be aware that they need help. This lack of awareness of one's own mental health status in the early days of depression [44] can stop an individual from asking for help from a web-based chatbot.

In contrast to chatbots, approaches such as Gboard [14, 16] can enable a systemic proactive early intervention at scale for the global population. The expert review in [4] calls out the cons of collecting personal mental health data. From a privacy point of view, patients’ personal data can’t be sent over the internet to cloud-based chatbots. In contrast, the proposed approach is AI that run on-device. This AI is an NLG based CBT inspired self-help therapy, packaged in the form of a predictive keyboard such as Gboard [14]. The proposed idea is illustrated in Figure 2 and Figure 4.

| Knowledge gap in the literature | Concept | Novelty / What is unique in this paper |
|---------------------------------|---------|----------------------------------------|
| Lots of existing literature     | **Mental Health Diagnosis** Application of NLU (BERT based analysis of a sentence spoken/tweeted) | |
| Gap/ Opportunity to contribute  | **Mental Health Therapy** Application of NLG to influence the person’s thoughts | Notable contribution is the proposed GPT2 architecture for CBT (Refer Figure 6) |

*Table 1: Related Literature and gaps for contribution*

**Contributions:**

A call for action to avert the forthcoming mental health was issued recently by the United Nations (UN) [3]. Towards addressing this mental health crisis fuelled by COVID-19 pandemic [2], contributions in this paper are as follows.

**A scalable AI solution for mental health intervention:** Given the pandemic's impact, the proposed solution needs to be scalable to millions of families. It also needs be a timely health intervention for millions of people. It also should be privacy
safe. This work proposes and demonstrates an AI-based strategy with a reference solution in response to this call for multi-disciplinary research on mental health [2]. The focus of this paper is applying the latest advancements in AI for therapy, though the mental health solution covers both therapy and screening. The solution to avert the crisis is based on both therapies using NLG (Result #1) and screening using NLU (Result #2). Both therapy and screening can be seen in Figure 1. The proposed mental health solution and the accompanying AI architecture is demonstrated with a working lab prototype with reproducible results. A lab prototype is implemented and open-sourced to enable many interested research communities to avert the crisis.

Novel application of GPT2 in therapy. This paper proposes and demonstrates a novel AI based approach to provide mental health therapy. The novelty is regarding the applications of recent advances in Deep Learning based Natural Language Generation (NLG) to one of the commonly used mental health therapy called Cognitive Behaviour Therapy (CBT). As described earlier in the literature review section, this paper explored an untapped potential of applying the state of the art advances in language models to create a mental health therapy solution. The possibility of conditioning the language generated using transfer learning allows AI based therapy to help individuals develop a positive way of thinking and seeing situations. Like auto-correction of spellings in a smartphone's predictive keyboard, this AI "auto-corrects" the narratives being typed by a mentally depressed person. Natural language generation generates text as user starts typing his thoughts, so that the user can view the situation from a "positive lens". A NLG enabled keyboard can be downloaded on to the user's smartphone that can offer care instantly, enabling early correction of unhealthy thoughts. The proposed AI embedded mental health predictive keyboard is illustrated in Figure 4. Since GPT2 inference computation happens in less than few seconds, this coaches the user's to have healthy thoughts while maintaining privacy. Since few research publications explored the applications of state of the art AI for language generation in self-help based therapy techniques such as CBT, this paper contributed to this gap. It proposed an AI architecture based on Transformers based NLG, proposed novel AI approaches such as NLU activated NLG. It also demonstrated the feasibility of the proposed AI with a prototype implementation. The proposed AI opens the door to develop a AI-based solution to avert the looming pandemic scale crisis for millions of families with a timely health intervention. It is also designed to ensure privacy through an on-device AI inference, even though CBT often involves handling of very sensitive personal information. A pre-trained GPT2 model was fine tuned to generate human like English based on training on a synthetic dataset. Three different GPT2 models were compared, concluding a small-sized language model may be good choice for on-device AI inference. Novel ideas such as NLU activation of GPT2 based therapy ensures timely intervention in a wide deployment for a large population. A prototype of BERT activated GPT2 based therapy is demonstrated. The demonstration shows the GPT2 based self-help therapy is activated only when appropriate. Experts have alerted us of a crisis sooner or later [40]. In this backdrop of the forthcoming crisis, this AI demonstrates a significant leap forward in developing an approach to prevent the impending mental health crisis. This work shows it is possible to use AI to prevent the looming mental health crisis.

Experts have called for multi-disciplinary teams [2] to get involved in averting the crisis. Accordingly, to encourage research communities to avert the forthcoming mental health crisis, the paper contributes an open-source AI prototype. While this paper focuses on the Deep Learning aspects of the solution, future multi-disciplinary research teams can extend this AI from the mental health perspective. Though this prototype is on English, future AI researchers can extend this AI to many world languages as many countries need this solution.

Idea summarized: Novel AI application to avert the mental health crisis

A key idea is an application of state of the art Deep Learning based Natural Language Generation (NLG) for mental health therapy. There has been amazing progress in attention technique based NLG models, such as Transformers based NLG architectures. A language model is basically a Deep Learning model that is able to look at part of a sentence and predict the next word. GPT-2 is a language model to produce human-like text. The opportunity to employ transfer learning techniques in NLG arrived very recently. The recent advances in Transformers based NLG model such as GPT2 allow for AI performance
similar to human like performance. However, the potential of apply these Deep Learning advancements for improving mental health remains untapped in the literature. This paper demonstrates the potential to employ Transformers based NLG architectures to produce AI solutions in mental health therapy. Using Transfer learning, a GPT2 model was fine-tuned to generate short text sentences to help depressed individuals to look at the situation from a different point of view. A BERT based NLU detects if a person is entering depression and then activates the therapy module.

The idea of applying NLG for therapy is presented in Figure 2. In this idea, an AI resides in the smartphone of the user and helps to improve mental wellness by suggesting flexible language narratives. Similar to the auto-correction of spellings in a smartphone's predictive keyboard, the proposed AI solution “auto-corrects” the narratives being typed on the user’s personal smartphone, as illustrated in Figure 4. This paper infuses AI into one of the popular mental health therapy approach, called Cognitive Behaviour Therapy (CBT) [4]. CBT is a therapy technique to help people find new ways to behave by changing their thought patterns.

With the power of transfer learning and human like text generation capabilities based on GPT2, the potential to apply NLG in therapy is significant. A lab prototype demonstrates this result, and a short video recording of a demo is presented online at URL, https://sites.google.com/view/ai-in-mental-health.

Potential impact in future: Enabling research communities

GPT-2’s ability to generate conditional synthetic text samples of unprecedented quality combined with transfer learning opens up an opportunity for mental health professionals to scale their impact to support millions of people on a continuous online basis. This is the potential that can be unlocked to avert the challenge of mental health impacted by the COVID-19 pandemic. This paper presents this potential by developing and demonstrating state of the art AI.

Researchers can take advantage of this opportunity to apply state of the art AI techniques to improve mental health. Further to presenting novel ideas, the paper also contributes an entire AI prototype solution in open source. The scope of the prototype implemented is specified in Table 2. This is intended to encourage interdisciplinary research communities in their future research.

Organization of the paper

[Table 2 is in the supplementary files section.]

As presented in Table 2, Contribution #1 is around enabling a solution to avert the crisis, Contribution #2 is around novelty. The paper’s results section is organized as follows: The primary result of this paper is Result #1. While Result #1 presents the novel application of AI for mental health, Result #3 expands the implementation of result #1 from a Deep Learning viewpoint. NLG in CBT is presented as both Result#1 and Result #3, while NLU in screening is presented as Result #2. Future research by the community can expand this AI into a real-world deployment to avert the crisis.

The next section on Results presents the three results. For each of the three results, the respective methods, discussions on literature is organized in each sub-section.

2. Results, Methods, And Discussions
Given the mounting worldwide mental health impact reported recently in Oct 2020 by Lancet’s editorial [31], the focus is on designing an AI approach that can address all the 3 dimensions of the challenge. A scalable approach to impact millions in a timely intervention in a privacy-safe way is the focus. BERT activated GPT2 for the generation of short narratives for seeding healthy thoughts is demonstrated. The sample narratives generated by AI are shown to look similar or as good as narratives generated by human counselors. To offer CBT like self-help to correct one’s outlook towards a situation, a conditional language model is created using transfer learning. A human-like text generation and the synthesis of conditioned language is demonstrated. The sample outputs shown in Table 4 show the sentences generated by AI is as good as one spoken by humans, thus showing the feasibility of employing AI to scale mental health counselors’ impact. Mental health experts can scale their productivity by training an AI, which can assist their patients in their absence. This is important given the need to support millions of families, given the reported shortage of capacity in mental health services [31, 32]. The use of such AI also ensures both early and timely activation of therapy. Further, a NLU activated NLG based therapy, ensures activation at the right time. A Human-Computer Interaction approach of a proposed mental health keyboard with an on-device AI inference ensures 24x7 mental health assistance for millions of individuals. The on-device AI inference supports the privacy and protection of one's personal information. Comparing the performance of different GPT2 model shows a small-sized GPT2 with 224 Million parameters is a choice for widespread deployment on-device on smartphones. Future research can extend the performance on smartphones using DistillGPT.

Among the 3 results, Result #1 represents the primary result of this paper. While novel ideas are presented in Result #1 and Result #3, the paper doesn’t make any novelty claims for Result #2. The two contributions and the three results are specified in Table 2.

2.1 Result #1: Novel idea of NLG in therapy to avert the crisis: Design of a AI for early intervention in mental healthcare for millions while ensuring safety of sensitive info

Challenges addressed by this result:

1. WHO (World Health Organization) reported the need for action on mental health [3]. The challenge is to design mental health therapy to help the individual look at a given situation from a different perspective, in order to lead to mental wellness [8]. As an illustrative example, in a situation of a person losing a job during the COVID-19 pandemic, the AI should help the individual develop a positive inner belief [6]. The challenge is to design an AI based therapy by applying the state of art advancements in Deep Learning in NLG.

2. The challenge is designing an approach that can meet the demands of a pandemic scale mental health solution [2]. The need is a scalable solution to serve millions of families with 24 hours of continuous support for each of the families.

3. Early intervention is an implementation challenge for 21st century mental health care as per JAMA [44]. Early intervention can be defined as diagnosis and treatment at the earliest possible point, even presymptomatically [44]. Today’s challenge in mental healthcare is that treatments are typically deployed late and without the strategic goal of reducing the progression of the illness, as per JAMA [44].

4. Privacy concerns as mental health counseling often involve sharing of sensitive personal information with the therapist [24].

5. There is a reported shortage in the capacity of mental health services as per WHO’s assessment in year 2020 [31]. Given the pandemic, the need for tools to significantly multiply mental health professionals’ productivity is essential [24]. Going forward, a mental health therapist should be able to care for a significantly larger number of patients.
Methods & Discussions:

a) A response to the call to action to avert the forthcoming crisis:

The proposed AI is designed as a response to WHO’s call to action on global mental health [3, 31]. As per WHO, almost 1 billion people suffer from a mental disorder [38]. Around 1 in 5 of the world’s adolescents suffer from a mental disorder, as per WHO facts [38]. The economy loses US$ 1 trillion every year in productivity because of depression or anxiety [38]. The Lancet editorial in Nov 2020 reports the monumental effects of COVID-19 pandemic on mental health [40]. The editorial [32] raises the following concern:- It is unclear how the world will deal with this forthcoming crisis, as the capacity of mental health services to respond in such a large scale doesn’t exist today. Hence the need for a scalable approach like the one proposed in this paper becomes significant.

b) The vast untapped potential of Transformers based AI architecture for mental health therapy:

Though Deep Learning research is progressing at an amazing pace, the opportunity to apply Transformers architecture [28] powered text generator [25] to improve mental health is not yet explored in the existing literature. Given the forthcoming crisis, it is imperative to explore the feasibility of employing the latest advances in AI for mental health. In order to avert the titanic crisis, this untapped potential is explored in this paper. This paper presents a novel idea at the intersection of CBT [17] & AI.

As per the May 2020 expert review [4, 5], the opportunity to apply a powerful deep neural network of the order of 100 Million neural network parameters for mental health therapy solutions to avert the looming mental health crisis is less explored in the literature. The benefit of such a neural network is human-like performance in language modelling [21]. The ability to fine-tune the models using transfer learning technique [13] allows the computer-based generation of language that is as close as possible to human counselors [6, 24]. One reason is progress in large NLG models such as GPT2 (Generative Pretrained Transformer 2) [25] happened very recently in year 2019. A search of “GPT2 and Cognitive Behavior Therapy” in Google Scholar yields no result.

GPT-3 [43] introduced in year 2020 is too large for on-device AI inference on smartphones. Unlike GPT-2, GPT-3 is only accessible as a cloud API. So GPT-3 can’t meet the privacy requirement of on-device AI inference. Hence GPT-3 is not suitable for consideration for this mental health challenge.

Using GPT2 to help patients practice Cognitive Behavior therapy (CBT) to change their negative view — all in real-time with privacy safety is a new idea. Further, given the scale of the problem, the therapy has to proactive rather than user-initiated. The idea of a timely health intervention in a privacy safe approach for millions of families by novel application of GPT2 is made possible now due to the advancements in AI, and this paper proposes this idea & demonstrates with a working prototype.

c) Digital interventions in mental health approaches such as Cognitive Behavior Therapy (CBT):

Experts have called for the involvement of multi-disciplinary research [2] to avert the forthcoming crisis. While this paper’s primary highlight is on Deep Learning applications, this paragraph introduces mental health therapy concepts such as Cognitive Behaviour Therapy (CBT).

CBT is a popular form of mental health therapy. Cognitive Behavior Therapy (CBT) [17] is psycho-social intervention that aims to improve mental health. Cognitive model of depressed individuals self-construct such a negative view of himself [6]. Beck’s “Cognitive triad” [7] enables the patient to correct their view of the world [8]. By changing unhelpful or inaccurate thinking, Cognitive therapy equips individuals to practice more flexible ways to think to overcome the cognitive distortion [6]. The JAMA article [9] concluded the effect of early intervention using Cognitive therapy. But the number of people who need help is multiple order of magnitude higher due to the pandemic, hence JAMA article [1] calls for creative thinking in treatment. The Lancet Psychiatry position paper calls for Digital Interventions [2]. This delineates the potential of applying
Deep Neural network based “Digital Cognitive therapy”. From the perspective of an interdisciplinary researcher, the **opportunity for AI-based CBT** is proposed in this paper. The concept idea of AI based CBT is offered in Figure 2. As shown in the figure, the circle of thoughts and inner beliefs of an individual can be ‘influenced’ by helping the person to change the way he ‘perceives’ the situation. The figure also illustrates the example where AI helps the individual to perceive his loss of a job from the viewpoint of his strengths. The video in the URL further clearly articulates how AI is able to sow healthy thoughts.

d) **AI for transforming Early Intervention in mental health care:**

**Early intervention is an implementation challenge for 21st century Mental Healthcare** as per JAMA Psychiatry [44]. Early intervention is significant in many healthcare settings. The same applies for mental healthcare too. Early intervention is giving care at the earliest possible point or pre-symptomatically [44]. In contrast to visiting counseling centers physically, AI allows for CBT inspired self-help almost instantly. Though an appointment with a human counselor may take weeks, especially with the reported shortage of capacity of mental health services, an AI based intervention can happen almost instantly. AI therapy enabled smartphone can bring care to the individual in the early stage of depression. A NLU activated therapy can turn on the therapy at an appropriate time as it detects if the person needs help. This ability for a proactive AI based CBT presents a breakthrough. **AI can thus provide a solution to the 21st century challenge of early intervention in mental healthcare.** The idea of early intervention in mental health by creatively applying AI can yield breakthrough advancements in the mental health of a person.

**Result & Methods: Novel application of GPT2 for CBT to address the 3 challenges**

- **Highlights of Result #1 :**
  1. In response to WHO and experts’ call for action to avert the looming mental health crisis, this result contributes an AI-based solution.
  2. This result contributes a solution that addressed the 3 dimensions of the challenge: scale for millions, timely intervention & privacy safety.
  3. As of 8th Jan 2021, there is a lack of research publications on applying state of the art AI models such as GPT2 for mental health therapy. To the best of our knowledge, this paper is the first to apply GPT2 for CBT.
  4. Novel application of GPT2 for mental health therapy solution is proposed.
  5. The proposed concept of AI aided therapy is articulated in Figure 2.
  6. The power of Transformers based NLG architecture allowed for the generation of human-like narrative, where fine tuning of the language model was performed by transfer learning.
  7. The proposed GPT2 based CBT was demonstrated with a lab prototype implementation.
  8. The feasibility of generating short sentences by AI that resemble human-generated was experimentally demonstrated. The video capture the live demonstration. For reproducible results, the code is shared online.
  9. Fine-tuning a pre-trained GPT2 on a synthetic dataset composed of around 5000 short-sentences generated language narratives that help the person look at a situation from a more positive mental outlook.
  10. The source code is contributed in open source code. This can enable future work by research communities to avert the forthcoming crisis.
The idea of applying GPT-2 for mental health is rather unique in the research literature. GPT2 based CBT is attempted in this work in the backdrop of lack of any research publications on GPT2 based CBT. Given backdrop of a call for multi-disciplinary priority [2] to avert the looming global mental health crisis, contributing to this gap urgently is even more significant. This paper not only address this gap, but also encourages research communities to avert the crisis by 2 contributions as specified in Table 2.

The proposed idea of **GPT2 based CBT** is introduced to address the challenge articulated earlier. The novel concept of GPT2 in CBT is articulated in Figure 2. In the proposed model, a language model listens to users situation and help her frame the narrative. A language model is simply an AI that predicts the next word in the sentence, given the previous set of words in the sentence. The example illustrated in Figure 2 showed how GPT2 helped a person who lost a job to develop a better ‘outlook’ to pervice the situation. The video demonstration in URL ([https://sites.google.com/view/ai-in-mental-health](https://sites.google.com/view/ai-in-mental-health)) shows clearly various scenarios of how GPT2 based NLG can help ‘tune’ an individual’s view or outlook. A fine tuned GPT2 model is demonstrated to generate human like text in this video as well in the screenshots in Figure 2 and Figure 4. This demonstrated the novel concept of employing a fine tuned GPT2 model towards a solution for mental health therapy. This demonstration of a novel application of the power of GPT2 to provide human like text to enable therapy will be of tremendous interest to experts and leaders who are interested to prevent the forthcoming mental health crisis caused by the COVID-19 pandemic.

Further, this paper conceptualizes and implements a working prototype of AI based solution that can offer immediate mental health “care” to improve mental wellness of millions of individuals. A **working prototype of applying GPT2 for mental health is demonstrated.** Also, a video recording of the prototype is presented as a video exhibit in the URL, [https://sites.google.com/view/ai-in-mental-health](https://sites.google.com/view/ai-in-mental-health). Further to enable reproducibility of results, the Google colab hyperlinks are shared at [https://sites.google.com/view/ai-in-mental-health/ai-to-seed-good-thoughts](https://sites.google.com/view/ai-in-mental-health/ai-to-seed-good-thoughts). The Deep Learning implementation aspects of this Result #1 are further detailed in Result #3.

Given the vast untapped potential of applying Transformers based mental health therapy solutions, this paper also contributes an AI prototype in open source. Thus this paper encourages and enables research communities to accelerate future research to avert the impending & monumental mental health crisis.

Different aspects of the idea are articulated in 4 different pictures - Figure 2, Figure 3, Figure 4 and Figure 6.

The conceptual idea of **AI in Cognitive Behavior Therapy** is proposed in Figure 2. As illustrated in the figure, the cognitive circle of thoughts and beliefs of an individual is intercepted by self-help based AI therapy. The diagram depicts how AI influence feelings & thoughts. The triangle at the center of the circle represents the person’s beliefs, which can be influenced by the AI. Inspired by CBT technique [17] to get help by correcting one’s beliefs about a situation, the AI offers self-help to correct one’s belief for every situation. An example of a situation is a person losing a job due to the pandemic, and then getting depressed. Figure 2 illustrates an instance of how AI can help a person who may feel depressed after losing his job. In the illustrated scenario, he types/speaks his situation in his smartphone as “I lost my job, I am depressed”. The GPT2 language model takes this initial phrase as an input and predicts the next words in the sentence. Based on this input, the GPT2 generates a narrative as “I lost my job, I am depressed. Let me keep remembering that I am smart”. The screenshot in figure 2 demonstrates this scenario. More examples that demonstrate the AI based self-help is shown in Table 4. A video demonstration of this AI based self-help therapy can be seen online at the URL. The conditional language generation was fine-tuned in such a way so that the positive beliefs are gradually sowed. Mental health experts compose narratives such as one generated in Table 4 containing a situation and a belief into a training dataset. The training dataset used in this prototype had around 5000 such short-sentences containing various situations. Each sentence in the dataset had a situation and the corresponding belief that can help the person to come out of depression. The dataset used in this lab prototype can be accessed at this URL. This dataset is a small dataset synthesized programatically. Mental Health Professionals compile such a dataset and use it to train the AI to create a fine-tuned GPT2 model. This helps scale the number of patients who can be cared for by every mental health expert. Thus mental health experts can leverage AI to multiply their productivity
to achieve the broader objective of preventing the forthcoming mental health crisis across countries. This addresses the shortage of capacity in mental health professionals [31] to scale to millions of families. Mental health professionals fine tune a pre-trained GPT2 model to create a new model using transfer learning. Transfer learning [13, 22] was performed on an OpenAI's pre-trained GPT2 model [25] as shown in Figure 3. This allows for human like text narratives [21] to be generated by the new model, which when read, may influence the thoughts, hence enabling the depressed individual to cultivate a positive mental outlook [8]. This Human-Computer Interaction model [33] is proposed to be similar to a user downloadable smartphone keyboard such as the popularly used predictive keyboard on android smartphones such as Google Gboard [14].

A visual of the keyboard is shown in Figure 4. The AI model is embedded in the smartphone keyboard, where the AI inference happens locally on the local smartphone device running Tensorflow Lite. So similar to how predictive keyboard such as Gboard helps auto-correct the spelling of what is being typed, the proposed keyboard helps correct the mental outlook to improve the mental health of the smartphone user. Once a smartphone is enabled with this AI, it equips the individual to think more positively, as shown in Figure 2. User's thoughts, often in the form of speech or text is fed into their personal smartphone, then analyzed in privacy safe technique. Privacy is enabled as the user's thoughts/text doesn't leave the smartphone, but inferred locally on the device as presented in Figure 6. Thus a novel conceptual advance in the application of the latest techniques in AI for averting the global mental health crisis had been contributed. Additionally, the proposed AI design is implemented and demonstrated to be technically feasible using a working implementation of the proposed AI. For reproducibility of results, the AI is contributed in open source at the website accessible at this URL. In addition, the additional novel ideas are contributed and demonstrated to evolve an AI solution to address the global mental health crisis.

If the AI identifies a trajectory towards depression, conditional language modelling (NLG) is activated, as shown in Figure 6. A NLU model activates the NLG, based on detection of mental resilience of the individual, as illustrated in Figure 6. At an appropriate instance, a GPT2 neural network-based NLG (Natural Language Generation) transforms any depression thoughts into something with a better outlook. A method to detect mental resilience is discussed later as part of Result 2. Given the set of words as input, GPT2 outputs the next set of words to auto-complete the sentence. So this AI can be used to lead a stream of thoughts away from depression. This novel idea of AI in Cognitive Behavior Therapy is proposed in Figure 2. AI was demonstrated to generate a narrative that helps influence inner belief for a situation. Also, the proposed AI architecture is presented in Figure 6.

The online demo of AI-based prototype solution is in this URL. In this video exhibit in this URL, it can be seen how a fine tuned GPT2 model can be employed by communities to help the individual to gain a better outlook by fixing the internal beliefs. The fine tuned GPT2 model is able to generate a language of gratitude and hope, even when the input is thoughts of loss and depression, as seen in the screenshot of Figure 4.

Mental health assistance in real-time, where self-correction is facilitated by AI using NLG, while another NLU model keeps a tab on the person's mental health, was demonstrated in Figure 4 and Figure 6. The comprehensive NLU activated NLG implementation method is discussed in Result 3. The NLU based mental health detection is discussed later as part of Result 2. In short, this result #1 demonstrated the potential and feasibility of applying GPT2 to offer mental health care almost instantly to a potential candidate. This kind of early intervention in mental healthcare is much needed [44].

To summarize, mental health experts train a GPT2 model to multiply their impact to avert the crisis. The novel concept of AI in CBT shown in Figure 2 was demonstrated with a prototype implementation. The potential for proposed AI architecture to solve the the 3 dimensions of the challenge is a notable discussion point. The proposed AI architecture approach presented in Figure 6 is designed to achieve the scalability to millions, achieve early intervention in mental healthcare, and ensure the sensitive personal information of the individual doesn't leave her personal device for privacy safety. Thus this result is a leap forward in the roadmap to apply recent advances in AI to avert the looming pandemic scale mental health crisis. The open-sourcing of the AI further encourages many research communities.
Result #2: Implementation of a lab prototype of a NLU based detection of the state of mental health

Challenges addressed by this result:

1. Many reports have already established the monumental scale of the mental health challenge in COVID [2]. Given the vast majority of the population across multiple countries, a systematic and scalable strategy for proactive mental health screening and non-intrusive rapid diagnosis is necessary to avert the looming mental health crisis. Figure 1 shows the framework of a solution.

2. An idea of NLU based activation of therapy is later discussed as part of the next result, result #3. This implementation of a NLU to detect the progression of mental health is discussed in result #2. This NLU module is later re-purposed as a sub-module as part of result #3.

Methods & Discussions on NLU for detecting mental health:

Very large deep neural networks such as transformer architecture based language models such as Google BERT offer a significant ability to understand English language, making it an excellent choice to understand what a candidate says using NLU (Natural Language Understanding) [12]. Transfer Learning on BERT is proven to be a viable technique for understanding sentences in any domain [13]. The abundance of literature in NLU for social media listening [5, 34] is noted in Table 1. Within the context of social media listening with NLU, the idea of use of BERT (Bidirectional Encoder Representations from Transformers) [10] over tweets is well established recently in 2019. Until 2018, Deep Neural Networks such as RNN was utilized [5]. This result shows the application of these state of art Deep Learning methods for screening or diagnostic. Given an abundance of literature in NLU based screening, this result #2 is about an open-source lab prototype implementation based on proven approaches.

Result & Method:

Highlights of Result #2:

1. Result #2 is about the application of NLU to detect the state of mental health of an individual.

2. Result #2 assumes significance in the context of the idea of NLU activated NLG, the details of which are described later as part of Result #3.

3. Result #2 is about implementing an NLU module to detect the progression of an individual's mental health.

4. The implementation is by application of BERT.

5. From a literature point of view, the paper doesn't claim any novelty in result #2 on its own. Result #2 is presented here for two reasons. Firstly, it is re-purposed as a sub-module in Result #3. Secondly, it is part of a contribution to avert the crisis as specified in Table 2.

To address the scale of a pandemic, a very different approach is necessary. Any capability to perform large scale screening or rapid pre-diagnosis enable experts is valuable to avert the looming crisis. A way to quickly analyze what a candidate is gone through during past weeks is the result obtained. An ability to screen large number of public, and analyze temporal patterns of every candidate quickly in the form of a visual report – makes it possible for rapid diagnosis by mental health professionals. An AI solution for large-scale screening is shown in Figure 7 to identify candidates who need diagnosis. For
the shortlisted candidates, a visual report, as shown in Figure 5 is generated for each candidate. The visual reports show mental resilience and time-based swings of cognitive behavior and the recovery rate. This report enables rapid diagnosis by mental health therapists. Using transfer learning, a BERT [10] based binary classifier identifies if the person is showing the language of a depressed person or exhibiting the signs of a person on a recovery path. The time to recover after a loss of job or family member is also explored. The mental resilience of a candidate can be understood by seeing the trend over a period of time. The proposed architecture for the detection of progression of mental health is shown in Figure 5. In this figure, a topic analyzer is cascaded with a mental health classifier, and then temporal modeling is performed. In short, a quick way to pre-diagnose by mental health professionals is demonstrated by the application of state of the art NLU. Thus the feasibility of screening at scale (Figure 7) along with rapid pre-diagnosis by mental health professionals using a cascade of BERTs based architecture (Figure 5) is demonstrated. An online demo is at this URL, https://sites.google.com/view/ai-in-mental-health/ai-in-diagnosis.

Result #3: Design of neural network architecture: NLG activated GPT2 based therapy

Challenges addressed by this result:

The challenge is in coming up with a neural network architecture for the solution proposed earlier in Result #1. The challenges to be addressed are

1. Scalability to a vast population
2. Privacy & protecting sensitive information: Patients reveal a lot of personal experience during a mental health counseling session. Collection of such sensitive data from many patients by a cloud system may involve risks. Such data should NOT leave the personal device boundary and should be deleted immediately after AI processing on the personal device.
3. Early intervention and 24x7 availability of care
4. Activate assistance only when appropriate: Intelligently choosing when to assist the patient in contrast to enabling her to become self-reliant.

Methods & Discussions:

The opportunity to scale impact via AI based CBT is discussed earlier in Result #1. Here familiar technique of transfer learning is used on the dataset on OpenAI's pre-trained GPT2 language model. This transfer learning idea is shown in Figure 8. Transfer learning in GPT2 has successfully produced poetry [15] and fake news [21]. The superiority of GPT2 (Generative Pretrained Transformer 2) [25] over earlier NLG techniques to produce reasonable text generation has been well established due to attention [18] based neural network architecture on model capacity above 100 million parameters.

The therapy needs to be started only when appropriate. If the person shows signs of natural recovery from a depressed mental state due to natural resilience, therapy is NOT required. So the AI therapy needs to be activated only when appropriate. It is important to keep a tab on the person's progression of mental health. Based on the progression, the decision to activate is taken intelligently by the AI. This idea of intelligent activation of therapy based on how the person is doing over few days is unique in the literature. This concept of detection by NLU, and then appropriate activation of therapy is proposed and demonstrated in this paper. The proposed intelligent activation approach is shown in Figure 6. A BERT based model
keeps track of a person’s mental health. In case this model detects that a person is trending towards depression, then it triggers activation of GPT2 based therapy model. This idea of intelligent activation of therapy using NLU activated NLG architecture is unique in the literature of mental health. The result is demonstrated in Figure 11.

Result & Method:

**Highlights of Result #3:**

- The idea of NLU activated NLG based therapy was proposed and implemented. A BERT-GPT2 cascade was implemented. In this approach, BERT detects if a person is depressed, and when appropriate, activates the AI therapy. The results of the proposed neural network architecture design presented in Figure 8 is demonstrated in Figure 11. Figure 11 shows two scenarios from a prototype of NLU activated NLG therapy. One scenario showed NLU triggered GPT2 therapy, another where therapy was NOT required. This demonstrated the proposed AI architecture of NLU triggered NLG. This intelligent activation architecture was demonstrated. This result is useful for early intervention [44], where the AI is pre-deployed in smartphones and gets automatically activated at the right time for timely activation of therapy to enable early intervention.
- For reproducible results, the source code is accessible online at URL.
- A pre-trained GPT2 model is fine tuned using transfer learning on a small synthetic dataset. Three different pre-trained GPT2 models of small (124M), medium (355M) and large (774M) pre-trained models were fine-tuned and all there achieved the same accuracy levels. (Refer Figure 9)
- Since the small (124M) and large GPT2 (774M) achieved the same accuracy levels for a small dataset of 5000 short sentences, a small GPT2 model with 124 million neural network parameters will be appropriate for on-device GPT2 inference.
- Though GPT-3 is a successor to GPT-2, GPT-3 is not suitable for on-device inference. This paper experimentally identified a small GPT2 model is sufficient to provide the performance required. In the future, researchers can explore distillation/model compaction and smaller model such as DistillGPT2 [36]
- Human-like text generation capability of NLG activated GPT2 based therapy was demonstrated with reproducible results. The sample results in Table 4 show human-like text generation is feasible with Beam search on GPT2. This demonstration of the capability to generate human like short narratives is useful as AI can offer early CBT therapy in the form of timely self-help, given mental health professionals need tools to multiply their productivity.
- This result demonstrated AI based self-help therapy. Its advantage is early intervention in mental health, user privacy & massive scalability. Thus, an AI-based mental health strategy & neural network architecture that combines the strengths of latest advances in NLU and NLG, namely BERT and GPT-2 was proposed and demonstrated. Thus this result demonstrated a way forward in preventing of looming pandemic scale mental health crisis.
- The open source contribution encourages many communities for further research, given experts are calling to action for multi-disciplinary research.

To summarize, here are unique knowledge elements and results contributed

1. **Beam search** [26] with GPT-2 decoded the next words in the working prototype. The results of beam search based GPT2 prediction is shown in the screenshot in Figure 4. The prototypes are implemented with Transformers [28, 29] based NLG and NLU using a Tensorflow/Keras library [29] on Google colab.
2. **A novel NLU-triggered-GPT2 is proposed & demonstrated with a prototype implementation.** The proposed architecture of NLU triggered NLG to selectively activate the therapy is presented in Figure 6 and more detailed in Figure 8. The simple
implementation of the proposed architecture as a lab prototype is demonstrated online at the URL, and the screenshot of this is presented in Figure 11. The screenshots in Figure 11 shows the two different scenarios, one where the AI therapy is activated, another scenario where the therapy was NOT required. The activation of GPT2 is performed only when appropriate as NLU module keeps a tab on the person’s mental health over time. By turning on/off the AI on the smartphone keyboard, families can opt in for self-help based mental wellness. So when both mental health tracking AI, and CBT AI are combined together, this allows a pervasive, non-intrusive, privacy safe approach to provide mental health care for millions. The BERT model detects if a person is depressed, then selectively activates the GPT2 based therapy. NLU module’s ability to detect the progression of mental health over a series of sentences over days was shown earlier in result #2. Figure 5 earlier showed that the NLU was able to detect if a person was recovering from depression. Hence the idea of NLU triggered NLG allows for activation of therapy only during appropriate circumstances such as the person doesn’t recover after a loss. While most people recover from a loss naturally after a duration, some may get into increasing levels of depression. This proposed idea of NLU activated therapy is able to handle such situations who need help. The proposed HCI model of smartphone keyboard makes it easy to keep a tab of the mental health of a family member, and the proposed approach can automatically activate therapy in the form of self-help. While users can opt-in to automatically download ‘mental wellness AI’ feature in their smartphone keyboard, this ensures family members stay mentally healthy. The protection of user’s sensitive information was demonstrated earlier in Result #1. Hence the path for scalable timely & early intervention with proactive deployment of the proposed ‘NLU activated NLG’ for masses can protect a country from being impacted by any future mental health crisis.

3. Three pre-trained GPT2 model of different sizes were fine tuned with same synthetic dataset. These were small, medium and large GPT2 neural network models of 174, 355, 744 million parameters respectively. The performance results of 3 different GPT2 models are tabulated in Table 3 and plotted in Figure 9 as training loss over training steps. This showed that for a small synthetic dataset, all 3 models achieved the same level of performance. Given the on-device requirement for AI inference, the GPT2 small (174M) model is suggested for smartphone deployment. In future, researchers can explore more memory and compute efficient models for smartphone such as DistillGPT2 [36] to enable deployment into real world.

4. An idea of conditioning language generation is demonstrated, demonstrating that AI is close to generating narratives like a mental health professional. The resulting conditioned text is shown in Table 4 for a couple of scenarios. It shows how AI can assist during many situations. It shows AI generates a conditional narrative that spins the words so that the user can look at the situation from a positive mental outlook. Transfer learning on Transformers based language model opens up the “imagenet moment for NLG”, so the potential of GPT2 for mental health therapy was successfully demonstrated. This result is a significance step towards AI based mental healthcare. The potential of Transfer learning in GPT2 to transforming the boundaries of what is feasible.

5. The dataset used for transfer learning for training the GPT2 model is a programmatically generated synthetic dataset. This synthetic dataset used in this paper contains 4098 records and can be accessed at the URL, https://sites.google.com/view/ai-in-mental-health/ai-to-seed-good-thoughts. Given the dataset is programmatically created as per the code shared at this URL, this gives flexibility for the Mental health therapists to quickly configure the input words to generate a synthetic dataset for a many situations. The dataset is contains short sentences such as the one shown in Table 4. Each sentence have a combination of a situation, and a belief. The situation and belief to influence is created by a mental health experts. The training time for all 3 models shown in Table 3 on this dataset was completed in less than 5 minutes on a single GPU environment for all the different sized GPT2 models, enabling swift solution deployment by mental health experts at the time of real word deployment.

6. Since mental health counselling for a set of target population related to each other by similar situations (e.g. a group of nurses overwhelmed by handling COVID-19 patients in a hospital), federated learning approach can be beneficial. To learn from community, aggregate model of Federated Learning [23] can be explored. The federation learning concept is presented in the architecture in Figure 8. Federated learning to identify positive self-help narratives that yield faster healing based on joint learning from a group of similar patients. This could be a future direction of research.
7. The choice of AI inference on the smartphone vs cloud was discussed in the context of privacy and willingness of patients to send sensitive information to cloud. A smartphone based model with AI inference on the local device enables privacy safety and user willingness to express themselves. The architecture proposed in Figure 6 ensures the user’s personal thoughts doesn’t leave her personal smartphone device, and hence enables privacy.

Based on the results demonstrated here, future researchers can develop a practical real world mental health solution to avert the forthcoming mental health crisis.

With the intelligent activation of AI therapy using a BERT-triggered-GPT2 architecture presented in Figure 6, Figure 11 shows the potential of AI based mental health solution to address the monumental challenge the paper aimed to achieve.

The power of transfer learning capable Natural Language Generation models such GPT2 when combined with on-device deployment approaches as shown in Figure 8 represent a leap forward in solution to respond to the global call for action by WHO and experts.

The design of neural network such as one in Figure 8 with an Open source lab prototype implementation demonstrated is a step forward in the response to the call for action by WHO to prevent the forthcoming monumental crisis.

The source code is available at URL https://sites.google.com/view/ai-in-mental-health/ai-to-seed-good-thoughts. The AI prototype is open-sourced. This encourages research communities for future research, given experts have called for multi-disciplinary priorities to avert the crisis.

| No of parameters in the neural network | GPT2 small | GPT2 medium | GPT2 large |
|--------------------------------------|------------|-------------|------------|
| Loss (Fine tuning)                   | 0.31       | 0.32        | 0.31       |

*Table 3: Performance of three different GPT2 models*
### Table 4. Sample output generative by GPT2 based CBT therapy

| Situation                                      | AI generated narrative for CBT therapy                                                                 |
|------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| (Input to GPT2)                                | (Output from GPT2)                                                                                    |
| 1 i lost my job, i am depressed.               | i lost my job, i am depressed. i remember my family say i am smart compared to others                 |
|                                                | i lost my job, i am depressed. my friend is kind to me                                                 |
|                                                | i lost my job, i am depressed. Let me recollect that i am cool                                       |
| 2 my dad died to COVID-19 virus infection, i am feeling so sad | my dad died to COVID-19 virus infection, i am feeling so sad, but luckily for me i have a supportive family |
|                                                | my dad died to COVID-19 virus infection, i am feeling so sad. Let me recollect that my family is so supportive |
|                                                | my dad died to COVID-19 virus infection, i am feeling so sad, i have managed worser things than this   |
| 3 i am feeling alone due to COVID19 virus caused lay off | i am feeling alone due to COVID19 virus caused lay off                                                |
|                                                | .Let me recollect that i am well qualified                                                           |
|                                                | i am feeling alone due to COVID19 virus caused lay off i feel so low . i have saved enough            |

### 3. Conclusions & Future Directions

This paper contributed a novel AI based on the applications of the recent advancements in Deep Learning to mental health. By applying state of the art in Deep Learning based language modelling [22], this paper makes a unique contribution to designing an AI that can help the community prevent the onset of a global mental health crisis. The work contributed novel ideas to the literature and developed a prototype to demonstrate the AI. The results are reproducible and are available online. Further to encourage the community to help avert the crisis, a AI prototype is contributed in open source. This supports the call to action by WHO [3, 31].

#### Overview of contributions made by this paper

The two aspects this paper developed are as follows:-

#### Contribution of novel ideas

1. This paper identified & contributed to the following gap in the literature. There are not many research papers that explored the applications of very recent advances in AI for addressing the mental health challenge. Specifically, there is
a lack of research papers on GPT-2 based mental health. This paper identified this gap and contributed a novel AI. Contributing to this gap is significant in the context of the call to action by eminent researchers [1, 2, 3, 31, 32, 40]. The work is significant as this is the 1st AI-based therapy based research that uses the latest AI techniques to design and develop a solution.

2. This paper establishes the role of an AI-based approach to address the global mental health challenge [1, 2, 3]. It proposed a novel AI architecture based on state of the art Deep Learning models.

3. Transformers based modeling [25] is making major leaps in AI since the year 2018. At its very core, Transformers based neural networks employ attention mechanism [18] for modeling. Transformers architectures are causing major shifts in natural language models more recently in the field of Deep Learning. It allowed for the development of pre-trained models, typically on massive datasets, and by exploiting the scalability of Transformers blocks to train gigantic neural networks. BERT is a Transformers based language model, and it is extensively used in understanding sentences. Typically, BERT [10, 34] is applied to build a classification model to detect depression based on words uttered in social media. Though there is substantial literature on such Natural Language Understanding (NLU) applications to diagnose mental health, there is almost not much literature on applications of Transformers based Natural Language Generation (NLG) for therapy.

4. Since the paper argued about the limitations of employing cloud-based chatbots for counseling, an alternative edge-AI based approach for therapy is discussed. The proposed approach is better than cloud chatbots as the proposed AI protects one's personal thoughts being streamed into cloud servers. The proposed AI is also better than cloud chatbots in addressing the practical challenge of early intervention in mental healthcare.

5. Transfer Learning [22, 28, 29] of language models is a mechanism to fine-tune the generation of text as per the required pattern by an NLG model. This work experimentally demonstrated that a GPT-2 model, once fine-tuned, can generate short sentences as required with Beam Search on GPT2.

6. The paper demonstrated a GPT-2 model [25] of any size (124 M or 355 M or 774 M) could be fine-tuned to generate human-like text narratives. Given the need for on-device AI inference on smartphones, the paper suggests using the small-sized GPT-2 (124M) model or further smaller sized models such as DistillGPT2. Though there are larger language models such GPT-3, it is not suitable for on-device inference.

7. A novel conceptual advance in GPT-2 based CBT was proposed and demonstrated with a working prototype. In this approach, as the user enters her thoughts into her personal smartphone keyboard, the AI inference can auto-correct her beliefs or outlook so as improve her mental wellness.

8. A conceptual advance is about the novel design of NLU triggered GPT-2 based AI therapy. While therapy is not required for everyone, it may be required to activate only when appropriate or necessary. So as BERT-based NLU module keeps a check of the progression of the smartphone user's mental health over time, it triggers the GPT-2 based NLG therapy module as needed. This novel idea of autonomous intelligent activation of therapy based on NLU activated NLG design pattern was prototyped in a simple BERT activated GPT2 architecture.

9. Human-like language generation for the purpose of therapy was demonstrated. In contrast to popular demonstrations of AI to generate fake news, this AI in this paper demonstrated the use of AI to generate short sentences similar to the ones used in CBT therapy. Based on fine-tuning a GPT-2 small model on a small synthetic dataset, the AI text output was human-like and helped the reader perceive the situation from a positive 'lens'. Examples of how the AI helps perceive the situation were experimentally shown as follows. These two examples are samples of output sentences generated by a fine tuned GPT2 language model, given an input sentence. In example #1, this GPT2 model formulated a short narrative that helps one develop and reinforce a positive mental outlook to a job loss situation. In another example #2, the model coaches the individual with an inner belief that helps her develop the courage to handle a difficult situation.

- Example #1:
  - Input (user’s speaks): “i lost my job, i am depressed.”
  - Output of AI: “i lost my job, i am depressed. i remember my family say i am smart compared to others”
• Input (user's types): “my mom died to COVID-19 virus infection, i am feeling so sad.”
• Output of AI: “my mom died to COVID-19 virus infection, i am feeling so sad, i have managed worser things than this”

In addition to the above contribution, the paper’s second contribution is to enable the community

1. A prototype AI with easy reproducibility can be accessed online to enable real-world deployment
2. The source code of the AI is released as Open source to encourage researchers

Multidisciplinary discussion: Contributions towards AI in mental health

Experts have called for multidisciplinary priorities to avert the global crisis [2]. While the above paragraphs discussed the AI-related aspects of the contribution, the following paragraph discusses how this AI addressed the global mental health challenge.

This proposed AI was designed to solve the three dimensions of the challenge.

1. Scale to millions of users: Improving mental wellness for millions of families
2. Timely intervention: Early intervention in mental healthcare
3. Privacy: Protection of sensitive personal information

The 1st dimension of the global mental health challenge is the scalability of the solution. The challenge is how to improve mental health for millions of families. The scale of impact on mental health due to COVID-19 pandemic is reported by articles [2, 32, 39, 40]. The AI demonstrated in this work allows for scalable on-device AI inference based deployment on millions of smartphones. Experimental results showed the requirement for on-device AI inference on a smartphone can be met, while still maintaining AI performance of human-like language generation. Experimental results showed a small-sized GPT-2 model provides good enough accuracy like larger sized GPT-2 models. A small sized GPT-2 neural network with 174 Million parameters was able to generate short text narratives that is as good as human spoken narratives, based on the examples. Various examples showed that text on any situation could be accepted by the AI and then transformed into a short-sentence that can help the reader view the situation from a different perspective. By fine-tuning the language model, the generated short sentence is designed to generate a positive belief for a given situation. In one example of a situation of job loss, the model generated a short sentence to help the reader remember that he is smart. A GPT-2 small model can be fine-tuned on the cloud, and then downloaded to the smartphone for on-device inference. For a typical small dataset with a size of around 5000 short sentences, a fine tuned GPT-2 small model yielded human-like short sentences.

The 2nd dimension is the challenge in the implementation of early intervention in 21st-century mental health care. This challenge of early intervention in mental health care is reported in JAMA [44]. Since this AI can be deployed on millions of smartphones, the therapy gets activated only when necessary due to the autonomous intelligent activation. If the smartphone user is depressed, the NLU activated GPT2 system tracks the mental health, and when it detects depression, it
automatically activates therapy. So once deployed to millions of smartphones, the early intervention of mental healthcare can automatically get activated for the needy individuals among the millions.

The 3rd dimension of the challenge is about adhering to rules to protect the personal information of patients requiring mental health counselling [6, 17]. Since a lot of information about the patients' various personal situations and emotions is spoken with the counselor, such information is very sensitive and personal to each individual. It is important to protect this information. Such information on personal situations should not get leaked to anyone. In the proposed approach in this paper, the AI inference happens locally on the smartphone device. So the user's sensitive information never gets sent out the smartphone into the internet. Since the inference happens on the smartphone, the personal information is deleted as soon as it is processed. So any text typed into the mental health keyboard or spoken by the user is processed by AI and deleted immediately. So such sensitive data typed/spoken is processed locally on the smartphone within a second. Thus no personal data is archived. Thus this proposed AI provides utmost protection of sensitive personal data to millions of families.

Results: Untapped potential & the demonstration of AI approach for CBT

This AI addressed the three dimensions of the global mental health challenge to avert the forthcoming crisis: namely scaling of the solution to support millions of families, early intervention in mental healthcare, and protection of sensitive personal information. Next, a detailed design of the proposed novel AI was conceptualized, prototyped & demonstrated. Finally, a working prototype was contributed to open source, enabling further progress in this field. This encourages multidisciplinary research communities to avert the crisis.

To the best of our knowledge, there is a lack of well-published research papers on the applications of the latest advancements in Natural Language Generation for mental health therapy. There is untapped potential to explore the use of the latest advances in AI for improving mental health. Since the last few months, there is a call to action by WHO and experts to avert the forthcoming mental health crisis. So the need for research on this untapped potential was crucial.

Given the mounting evidence of mental health impact due to the COVID-19 pandemic, there is a need to urgently explore the untapped potential of applying the latest progress in AI. This paper made a significant way forward in exploring this untapped potential.

The result was a conceptual advance in the way the latest progress in AI is utilized to improve global mental health. In addition to proposing a novel application of GPT-2, the paper designed an AI that meets all the dimensions of the challenge.

- The proposed novel AI allowed for scalable deployment of AI based self-help to improve global mental wellness for millions of families.
  1. This on-device AI inference of GPT2 for CBT inspired self-help allowed for scalable deployment. A GPT-2 small model with 174 Million neural network parameters offered compelling performance, and hence a compact GPT-2 language model is ideal for edge AI inference on mainstream smartphones.
- Amidst the reported shortage of mental health services and the looming pandemic scale impact on mental health, AI based approach offered a way boost the productivity of mental health experts.
1. The mental health professional composed dataset and trained an AI using transfer learning.
2. Since the text narratives were demonstrated to close to human like narratives due to conditioned language modeling, this opened the doors for early intervention for the masses.

- The 21st century challenge of early intervention in mental healthcare was addressed by a novel approach of NLU activated NLG.

1. Automatic intelligently activation of therapy allowed for early intervention of mental healthcare. NLU activated NLG ensured timely activation of self-help therapy.
2. Automatic activation of therapy on time on deployed smartphones opened

- Protection of sensitive information

1. Due to on-device inference in the proposed AI, the information typed/spoken by the user is NOT transmitted to the cloud.
2. All sensitive information typed by the user is processed by AI and immediately deleted within seconds.

**Detailed Results: Demonstration of AI architecture for mental health therapy**

The paper demonstrated the application of a Transfer Learning capable neural network for Natural Language Generation for generating human like text narratives towards a mental health therapy approach, specifically GPT2 based CBT. Inspired by a mental health therapy approach called **Cognitive Behaviour Therapy (CBT)** [17], the proposed AI explored the possibility of using human like text generated by the AI for self-help to change unhealthy ways of thinking. By conditional language modelling approach, a Transformers architecture-based GPT2 pre-trained on 8 million web pages, was fine-tuned by transfer learning on a programmatically synthesized small dataset of 5000 short sentences to create a new fine-tuned GPT2 model.

The fine-tuned GPT2 model will be activated at appropriate timing only when a Natural Language Understanding based classifier detected the need for therapy. The proposed **NLU triggered NLG** based AI is implemented with a working lab prototype. By using the power of recent advances in NLU and NLG such as **BERT** and **GPT2**, the proposed design of AI architecture was successfully demonstrated with a lab prototype. The concept of the intelligent activation of AI therapy at an appropriate time, based on each individual’s circumstance, allows for proactive deployment and early intervention in mental healthcare for millions. BERT classifier based module was able to detect if the person increasingly gets depressed over a period of time. Based on detecting the progression of mental health, intelligent activation of the therapy module was demonstrated. As a simple implementation of the proposed NLU activated GPT2 therapy, a simple prototype was implemented and demonstrated. **This is much needed given the challenge of early intervention in mental healthcare** [44].

**Once activated, the fine tuned GPT2 model generated text narratives that offered flexible ways to think about a situation.**

This work experimentally demonstrated the ability of a fine tuned GPT2 as a potential candidate for AI based therapy. Given the human like text generation capability, when combined with **transfer learning**, allowed the mental health expert to train and deploy a conditional language model to help millions of families impacted by the anxiety/depression.
By proposing an on-device AI inference inspired by the Gboard [14], a predictive keyboard on the smartphone to improve mental health in a privacy-safe way was proposed. Given a small GPT2 model with 124 Million parameters was shown to achieve same performance in fine tuning as a large GPT2 model with 774 Million parameters during fine tuning on the synthetic dataset of 5000 short sentences, it was noted that a small GPT2 model with 124M was appropriate for a smartphone based on-device inference. Though GPT-3 [43] is a successor to GPT-2 [25], it is not suitable for on-device AI inference. The proposed on-device architecture help to meet the challenge of requirement in privacy given a lot of personal information of the patient is involved. This smartphone based AI architecture also meant continual assistance enabling timely correction of unhealthy thoughts for an individual. Thus the architecture presented in this paper is a step forward in the direction of evolving an AI based novel mental health solution to avert the mental health crisis looming from the COVID-19 pandemic.

This paper addressed the monumental challenge of averting the forthcoming mental health crisis on all 3 dimensions of the challenge. By proposing state of the art Deep Learning architecture based mental health on-device AI solution, the paper demonstrated the potential of Transformers based neural network architecture to address an early and timely health intervention for millions of families across the globe. Additionally, the proposed AI solution architecture was able to offer privacy safety given the AI inference happened locally on the smartphone device. This allowed for the AI solution to handle sensitive personal information such as thoughts, personal situation within the boundaries of the personal smartphone. A BERT triggered GPT2 approach further enables such CBT-inspired self-help approaches to correct one's unhealthy thoughts. The fine-tuning of GPT2 created a conditioned language model, powerful enough to create a human like text narrative that was able to generate and sow healthy perspectives for a given situation. The beauty of the proposed architecture is its ability to handle an individual's very personal sensitive thought stream in absolutely privacy-safe ways as the AI inference deletes the spoken/typed messages as soon it is processed locally on-device. Further, since the AI-based therapy is only activated at an appropriate time by an on-device BERT classifier, the AI ensures timely therapy with privacy safety.

Future directions & enabling research communities

The doors for AI based Mental Health Therapy for pandemic level scalability has been unlocked by the ideas and the demonstrated results. This was the 1st time in the reviewed literature, an advanced transfer learning capable language model was demonstrated to generate human like narratives for mental health therapy, with the 3 challenges of solution being able to scalable to millions in a privacy safe with a timely health intervention. The UN's call to action [3] and the titanic challenge of preventing the forthcoming mental health crisis [32, 39, 40] is addressed by designing a AI based solution using the state of art Deep Learning.

Given the magnitude of the looming crisis, accelerated research by multiple research communities is called upon by experts [2]. Towards supporting this goal of enabling future research by research communities, this paper contributes the AI prototypes in open source. While this paper explored a best of its kind language model to support English language, further research is required on more languages to support the population in multiple countries. Further generation distillation approaches [36] such as DistillGTP2 and model compaction to enable deployment of the AI on commodity smartphone is required. More importantly, a first baby step in the roadmap to practical real-world deployment of state of the art AI based therapy has been attempted by this paper, but a lot of interdisciplinary research is required in the future to bring this AI based mental health solution to improve mental health for millions of people. While this papers’ lab prototype demonstrated the feasibility of applying state of the art AI innovatively for AI based therapy, the architecture proposed in this paper is provides an approach that can meet the massive challenges for a large scale real-world mental health solution. The architecture proposed, along with the working prototype, are capable of meeting the three dimensions of the challenge for a large scale
real-world deployment to avert the looming mental health crisis, namely the ability to impact millions of families in a timely health intervention in a privacy-safe approach. To build a real-world impact, multi-disciplinary researchers now have the opportunity to avert the forthcoming crisis.

4. Availability Of Source Code

Source code: Open sourced at URL, https://sites.google.com/view/ai-in-mental-health

Dataset: Dataset was synthesized programmatically and is embedded as part of source code, and can be accessed online at the URL.

5. References

1. Galea, S., Merchant, R. M. & Lurie, N. The Mental Health Consequences of COVID-19 and Physical Distancing: The Need for Prevention and Early Intervention. *JAMA Internal Medicine* (2020). doi:10.1001/jamainternmed.2020.1562

2. Holmes, E. A. *et al.* Multi-disciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *The Lancet Psychiatry* (2020). doi:10.1016/s2215-0366(20)30168-1

3. UNSDG | Policy Brief: COVID-19 and the Need for Action on Mental Health. *unsgd.un.org* (2020). Available at: https://unsgd.un.org/resources/policy-brief-covid-19-and-need-action-mental-health. (Accessed: 28th May 2020)

4. Durstewitz, D., Koppe, G. & Meyer-Lindenthein, A. Deep neural networks in psychiatry. *Molecular Psychiatry* 24, 1583–1598 (2019). https://www.nature.com/articles/s41380-019-0365-9

5. Chancellor, S. & De Choudhury, M. Methods in predictive techniques for mental health status on social media: a critical review. *npj Digital Medicine* 3, (2020). https://www.nature.com/articles/s41746-020-0233-7

6. Dobson, K. S. *Handbook of cognitive-behavioral therapies*. (Guilford Press, 2019).

7. Keser, E., Kahya, Y. & Akin, B. Stress generation hypothesis of depressive symptoms in interpersonal stressful life events: The roles of cognitive triad and coping styles via structural equation modeling. *Current Psychology* (2017). doi:10.1007/s12144-017-9744-z

8. Dobson, K. S. & Shaw, B. F. The effects of self-correction on cognitive distortions in depression. *Cognitive Therapy and Research* 5, 391–403 (1981).

9. Ehlers, A. *et al.* A Randomized Controlled Trial of Cognitive Therapy, a Self-help Booklet, and Repeated Assessments as Early Interventions for Posttraumatic Stress Disorder. *Archives of General Psychiatry* 60, 1024 (2003).

10. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv.org* (2018).

11. Denecke, K., Vaaheesan, S. & Arulnathan, A. A Mental Health Chatbot for Regulating Emotions (SERMO) - Concept and Usability Test. *IEEE Transactions on Emerging Topics in Computing* 1–1 (2020). doi:10.1109/tetc.2020.2974478

12. Doan, S. *et al.* Extracting health-related causality from twitter messages using natural language processing. *BMC Medical Informatics and Decision Making* 19, (2019).

13. Ranti, D. *et al.* The Utility of General Domain Transfer Learning for Medical Language Tasks. *arXiv:2002.06670* (2020).
14. Yang, T. et al. Applied Federated Learning: Improving Google Keyboard Query Suggestions. arXiv.org (2018). https://arxiv.org/abs/1812.02903

15. Liao, Y., Wang, Y., Liu, Q. & Jiang, X. GPT-based Generation for Classical Chinese Poetry. arXiv:1907.00151 (2019).

16. Li, X. et al. Multi-site fMRI Analysis Using Privacy-preserving Federated Learning and Domain Adaptation: ABIDE Results. arXiv:2001.05647 (2020).

17. Beck, J. S. Cognitive Behavior therapy basics and beyond. (Guilford Press, 1995).

18. Vaswani, A. et al. Attention Is All You Need. In Advances in neural information processing systems, pp. 5998-6008 (2017).

19. Holt-Quick, C. et al. A chatbot architecture for promoting youth resilience. arXiv:2005.07355 (2020).

20. Xia, Q. et al. XGPT: Cross-modal Generative Pre-Training for Image Captioning. arXiv:2003.01473 (2020).

21. Zellers, R. et al. Defending Against Neural Fake News. arXiv:1905.12616 (2019).

22. Peng, X., Li, S., Frazier, S. & Riedl, M. Fine-Tuning a Transformer-Based Language Model to Avoid Generating Non-Normative Text. arXiv:2001.08764 (2020).

23. Lim, W. Y. B. et al. Federated Learning in Mobile Edge Networks: A Comprehensive Survey. IEEE Communications Surveys & Tutorials 1–1 (2020). doi:10.1109/comst.2020.2986024

24. Gratzer, D. & Goldbloom, D. Therapy and E-therapy—Preparing Future Psychiatrists in the Era of Apps and Chatbots. Academic Psychiatry 44, 231–234 (2020).

25. Radford, A. et al. Language Models are Unsupervised Multitask Learners. OpenAl Blog 1, no. 8: 9 (2019).

26. Li, X., Li, P., Bi, W., Liu, X. & Lam, W. Relevance-Promoting Language Model for Short-Text Conversation. arXiv:1911.11489 (2019).

27. Eisenschlos, J. et al. MultiFiT: Efficient Multi-lingual Language Model Fine-tuning. arXiv:1909.04761 (2019).

28. Wolf, T. et al. HuggingFace's Transformers: State-of-the-art Natural Language Processing. arXiv:1910.03771 (2020).

29. Maiya, A. S. ktrain: A Low-Code Library for Augmented Machine Learning. arXiv:2004.10703 (2020).

30. Clark, K., Luong, M.-T., Khandelwal, U., Manning, C. D. & Le, Q. V. BAM! Born-Again Multi-Task Networks for Natural Language Understanding. arXiv:1907.04829 (2019).

31. WHO team. The impact of COVID-19 on mental, neurological and substance use services, WHO Publication, ISBN: 978-92-4-001245-5 (2020) https://www.who.int/publications/i/item/978924012455

32. Editorial. The intersection of COVID-19 and mental health. The Lancet Infectious Diseases. https://doi.org/10.1016/S1473-3099(20)30797-0 (2020)

33. Pedro Sanches et al. HCI and Affective Health: Taking stock of a decade of studies and charting future research directions. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI ’19). ACM https://doi.org/10.1145/3290605.3300475

34. Gkotsis, G., Oellrich, A., Velupillai, S. et al. Characterisation of mental health conditions in social media using Informed Deep Learning. (Nature) Scientific Reports 7, 45141 (2017). https://doi.org/10.1038/srep45141
35. K. Oh, D. Lee, B. Ko and H. Choi, “A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation,” 18th IEEE International Conference on Mobile Data Management (MDM), Daejeon, 2017, pp. 371-375, doi: 10.1109/MDM.2017.64.

36. Melas-Kyriazi, Luke, George Han, and Celine Liang. “Generation-Distillation for Efficient Natural Language Understanding in Low-Data Settings.” arXiv preprint arXiv:2002.00733 (2020).

37. Vig, Jesse. “A multiscale visualization of attention in the transformer model.” arXiv preprint arXiv:1906.05714 (2019).

38. WHO Mental Health Facts https://www.who.int/news-room/facts-in-pictures/detail/mental-health

39. [Editorial] Keep mental health in mind. Nature Medicine 26, 631 (2020). https://doi.org/10.1038/s41591-020-0914-4

40. [Editorial] Mental health matters. The Lancet Global Health 8, no. 11 (2020). https://doi.org/10.1016/S2214-109X(20)30432-0

41. Sharma, Ashish, Adam S. Miner, David C. Atkins, and Tim Althoff. “A Computational Approach to Understanding Empathy Expressed in Text-Based Mental Health Support.” arXiv preprint arXiv:2009.08441 (2020).

42. Fedus, William, Barret Zoph, and Noam Shazeer. "Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity." arXiv preprint arXiv:2101.03961 (2021).

43. Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan et al. “Language models are few-shot leaners. arXiv 2020." arXiv preprint arXiv:2005.14165 4.

44. McGorry PD, Ratheesh A, O'Donoghue B. Early Intervention—An Implementation Challenge for 21st Century Mental Health Care. JAMA Psychiatry. 2018;75(6):545–546. doi:10.1001/jamapsychiatry.2018.0621

Declarations

6 Author information

Rajagopal A, Interdisciplinary research program, Indian Institute of Technology, Madras, 600025, India. am18d301@smail.iitm.ac.in , rajagopal.motivate@gmail.com Phone: +91 9445412134

7 Ethics declaration / Competing Interest

The authors declare that they have no competing interests.

Ethics approval: Not applicable

8 Funding

None

Figures

Figure 1
**AI framework to avert the pandemic scale mental health crisis during COVID19.** The opportunity to employ AI in both diagnosis and counselling is immense. Though there are many research publications in using NLU for diagnosis, there is a significant opportunity to contribute to applying NLG for therapy. This paper develops a working lab prototype for both mental screening using BERT, and mental health counselling using a combination of BERT & GPT2.

**Figure 2**

**Novel application of AI in mental health therapy.** The novelty is in applying Natural Language Generation (NLG) to stimulate a healthy mental outlook as the user types a message about a depressing situation such as losing a job during the pandemic. Like auto-correction of spellings in a smartphone's predictive keyboard, this AI "auto-suggests" the narratives being typed by a mentally depressed person to sow healthy beliefs.

**Figure 3**

The novel application of AI in mental health therapy. The novel approach involves the use of Natural Language Generation (NLG) to stimulate a healthy mental outlook as the user types a message about a depressing situation such as losing a job during the pandemic. Similar to auto-correction of spellings in a smartphone's predictive keyboard, this AI "auto-suggests" the narratives being typed by a mentally depressed person to sow healthy beliefs.

**Figure 4**

**Result #1: Conditioned GPT2 for a Mental Health Therapy.** This figure relates to Result #1. The screenshot above shows the input and output of the fine tuned GPT2 model. While the user's situation is depressing, the AI is able to predict a better narrative for a CBT inspired self-help therapy. A smartphone predictive keyboard for improving the mental wellness of users can embed this AI.

**Figure 5**

**Result #2: Rapid AI pre-diagnosis can be utilized by mental health professionals to amplify their productivity.** This figure relates to Result #2.

**Figure 6**

The AI supports the patient's natural mental resilience development, protects her privacy & private data. Privacy safe AI is a significant progress.

**Figure 7**
Scalable screening to identify who needs help

**Figure 8**

Neural network based architecture for Cognitive Behaviour Therapy

**Figure 9**

*Performance of three different GPT2 models*

**Figure 10**

*GPT2 visualization*

**Figure 11**

*Therapy gets activated only when appropriate. This picture shows the screenshot of the prototype implementation of architecture proposed earlier in Figure 8. A fine tuned BERT model triggers activation of GPT2 based therapy. For reproducing the results, the source code is available online at the given URL*

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- **Table2.docx**