PWC
PROBLEM. SOLVED.

Introduction

The Vin Baker Bouncing Back Foundation (VBBBF) is a project spearheaded by NBA star Vin Baker and inspired by his personal journey of recovery from substance abuse. VBBBF has created several initiatives to support others in recovery, including the Vin Baker Recovery Center located in Milwaukee, WI. The Vin Baker Recovery Center's mission is to provide access to healthcare for substance use, mental health issues, and recovery support. However, the reach of a single physical location is limited and for this reason, VBBBF is interested in creating a comprehensive, online resource to provide nationwide access to support and services related to substance abuse prevention and recovery and mental health challenges.

A Clearinghouse is defined as a "central agency for the collection, classification, and distribution of information". VBBBF envisions creating a comprehensive repository of information, resources, and support services that address substance abuse and mental health issues as well as the provision of 'wrap-around' social service referrals that address the underlying issues that prevent individuals from accessing the care they need.

This online resource will be known as the Vin Baker Substance Abuse Mental Health Clearinghouse ("Clearinghouse"). Its purpose will be to provide support, direction, and guidance for individuals and families affected by substance abuse and mental health issues, with a special focus on traditionally underserved communities. The Clearinghouse as envisioned has three primary components:

- Provision of tools and resources that support substance abuse prevention, treatment, and recovery for individuals and families facing mental health challenges.
- An easily accessible directory of local and national services across the continuum of care, including hotlines, treatment facilities, and support groups.
- A central repository for evidenced-based information and resources that will serve affected individuals, their families, and those with professional interests in these topics.

Competitive Environment

There are many websites that provide information, support, and services in relation to substance abuse and mental health issues. The Substance Abuse and Mental Health Services Administration (SAMHSA) hosts a federally-funded website that, along with its many subsidiary websites, provides access to a voluminous amount of mental health and substance abuse information. For example, a search on 'fentanyl' on the SAMHSA website produces 5 pages of results. The results provided are an undifferentiated mix of items of interest to individuals, healthcare professionals, and policy makers and are not filterable or sortable, which makes finding specific information of interest challenging. The website is built using the CloudFront content delivery network powered by Amazon and is reasonably ADA-compliant, as required by federal regulations.

The National Institute of Mental Health (NIH) provides a similarly information-dense website that is primarily focused on providing information and support for healthcare professionals performing research relating to mental health, behavioral issues, and cognitive disorders. A search on "fentanyl" returns a single page of results, with scholarly content geared toward healthcare professionals and researchers. Search results can be filtered by several preset modifiers including

"Publications", "Videos", and "News". This website provides useful information for both healthcare professionals and policymakers but is not very useful for individuals. This website is built using Drupal, a content management system very similar to WordPress, and is reasonably ADA-compliant.

A quick Google search for "Substance Abuse and Mental Health Clearinghouses" returns websites that are a mix of federally funded, privately-funded, and a "hybrid" mix of funding entities that include academic institutions. These websites all contain at least some of the resources of the type that VBBBF is interested in providing through the proposed Clearinghouse website. A sampling of these sites includes:

- **PyschU:** Calling itself "The World's Largest Community & Online Resource for Mental Health", PyschU has a long list of supporting organizations, from Tuskegee University to NAMI.

  The PyschU site uses Information Tailoring to serve cohorts of users similar to those VBBBF wishes to serve, although the focus seems skewed towards healthcare professionals rather than individual users. The website includes videos, podcasts, and infographics to disseminate information in multiple modalities. A "Disease State Education Simulator" and "virtual role play for health care professionals" are gamified elements, also geared towards healthcare professionals. The Search feature is limited, presumably due to a poorly populated database. A search on "fentanyl", for example, returned only 2 resources. The website is built on WordPress and is not ADA-Compliant.

  https://psychu.org/

- **The National Mental Health Consumers' Self-Help Clearinghouse:** This website is a national technical assistance center founded and run by individuals diagnosed with mental health conditions. The Clearinghouse is operated in collaboration with Temple University, this website is very small, consisting of link lists of publications and resources with the Clearinghouse portion hosted by Temple University. At least some of the information on this website has not been updated since 2015. The Temple University website is a larger collection of link lists, mostly presented as "blog-style" posts. There is a search function labeled "Temple University Collaborative on Community Inclusion". A search on 'fentanyl" returns no information. The website is built using "frames", which is an outdated technology that has not been used since the late 90s and is not ADA-compliant.

  https://www.mhselfhelp.org/

- **Health Center Resource Clearinghouse:** Operated in partnership with National Association of Community Health Centers.

  This website encompasses a wide range of health-related topics that include substance abuse and behavioral health. The substance abuse information is presented on a single page that includes a link to SAMSHA. There is a link to a 'Clearinghouse' search that is

filterable and sortable, however, a search on 'fentanyl" returns no information. The website is built on WordPress and is somewhat ADA-compliant.

https://www.healthcenterinfo.org/priority-topics/substance-use-disorders/

These websites are all good-faith attempts to provide a resource of the type described in the RFP for the Clearinghouse – online platforms that attempt to consolidate information, resources, and services related to substance abuse and mental health support. All of these websites provide at least some evidence-based information and resources to support prevention, treatment, and recovery as well as searchable "directories" of varying quality to access information regarding treatment, support groups, and other services aligned with substance abuse recovery and mental health. However, none of these websites meet the criteria of providing a truly comprehensive and consolidated Clearinghouse that helps the target cohorts make sense of the staggering amount of information available across the internet on the topics of substance abuse and mental health. The approaches are fragmented, at best, and some, as in the case of The National Mental Health Consumers' Self-Help Clearinghouse, provide outdated information using technology that has long been considered obsolete.

Clearinghouse Objectives

It is clear that a true Clearinghouse as described in the VBBBF RFP is needed. It is a Herculean task for anyone to sift through the results of a Google search on any topic related to substance abuse and mental health and find information that is up-to-date, accurate, and truly helpful. This task is nearly impossible for people in active distress, who need immediate access to support and services and who are not going to spend hours sifting through websites that do not provide instant, actionable answers to their pressing questions. The Clearinghouse website examples given above do not meet this need and create frustration that they do not deliver the comprehensive results that are implied by the use of the word "Clearinghouse". A true Clearinghouse resource should act as a 'one-stop shop' that reduces the burden of time spent endlessly searching the web for information, support, and services and increases the time spent accessing and using the services that create a path to wellness and recovery. The question then becomes how to create a Clearinghouse that accomplishes the goals of:

- Consolidating information, resources, and services related to substance abuse and mental health.
- Provision of accurate, up to date, evidenced-based information and resources that support prevention, treatment, and recovery.
- Creation of a directory of local and national support services, including hotlines, treatment facilities, and support groups.
- Inclusion of interactive tools, educational materials, and personal stories to aid in understanding and coping with substance abuse and mental health issues.

VBBBF has identified three target audiences for the Clearinghouse website:

- Individual Users
- Service Providers
- Policymakers and Funders

The information needs and information-seeking behaviors of these three cohorts are varied and diverse. For example, an individual searching for information on fentanyl may be a person who is concerned about their own or a loved one's use of this substance and is seeking information about intervention and treatment options. A Service Provider may be interested in best-practices information about treating a patient who presents in a clinical setting having used fentanyl and is possibly in a life-threatening overdose situation. A policy maker may be seeking information that includes statistics about the effects of fentanyl on communities of interest along with information about trends in access to treatment and recovery resources. While each of these cohorts chose the same search term, their expectations for the results presented are quite different.

A challenge facing VBBBF in the creation of a Clearinghouse resource is the design of a service that meets the information needs of these target cohorts within a user-friendly, accessible platform that not only provides comprehensive content and resources but also serves as a bridge to close gaps in understanding and access. In order to better understand the challenges in achieving these objectives, this proposal addresses three distinct components that must be addressed in order to develop a system that meets the stated objectives :

1. Development of the underlying Clearinghouse dataset
2. Delivery of the content through interaction with a public-facing website
3. Tailoring of the content delivery to meet the information needs of each of the target cohorts.

A project of this scope is immensely challenging. Technical resources across a range of disciplines will be required to plan and execute this project. Each component of this project will require funding that may, in aggregate, be beyond the current funding capacity of VBBBF. This proposal is designed to provide in-depth information related to each of these objectives in order to provide VBBBF decision makers with the background information needed to determine the best course of action, which may include rethinking the project in terms of phases tied to fund-raising goals.

Scope of this Proposal

The majority of the Key Factors and Tasks as outlined in the RFP involve operational and marketing activities and are outside the scope of this response. Additionally, the parameters of the technical aspects of this project as presented in the RFP are too broadly defined to enable price-and-time specific recommendations and costing. This response contains detailed information to support decision makers and stakeholders in four areas highlighted in the RFP:

- The Clearinghouse Dataset
- Information Tailoring for Multiple Target Cohorts
- Gamification
- Artificial Intelligence

This information in each of these areas is provided in order to educate VBBBF decision-makers and stakeholders about the choices available to achieve the goals and objectives of the Clearinghouse.

Datasets

The heart of a Clearinghouse is a dataset. For the past 50 years, datasets have been built and stored using Relational Data Base Management Systems ("RDBMS") such as SQL/MySQL. VBBBF has expressed an interest in using Artificial Intelligence ("AI") to provide information resources to their target cohorts via a public-facing website. AI requires a different type of architecture called "Vector Database" to store data. Although Vector Databases were conceived in the early 2000s, they did not gain traction for general use until the rise of AI in recent years. Vector Databases are an immature technology that is continuing to rapidly evolve on what feels like a daily basis. While RDBMS technologies are mature and stable, Vector Databases are not. One of the challenges of choosing a Vector Database platform is that today's technology can easily become tomorrow's obsolescence – an appropriate analogy for understanding the dilemma facing decision makers considering the use of AI technologies is the Beta vs VHS "wars" of the 1970s.

A dataset of the type proposed for the Clearinghouse will require specialized database resources including dedicated database hosting and management, regardless of which database platform is selected. If an RDBMS database is selected, this project may also require data warehousing services that can combine multiple, disparate sources and types of information into a single database using a common data model. Until the size and composition of the dataset can be reasonably estimated, it is impossible to estimate the amount of resources required or the costs associated with their provision.

Choices made about the design, population, and management of the VBBBF dataset are the most important and fundamental decisions that will be made about this project. Many of the choices, such as the type of database (RDBMS vs Vector) and the choice of platform (Open Source, Proprietary, Self-Hosted, Cloud-based, etc.) are intertwined with choices about the technology used to deliver the information on the website. But no matter the outcome of these choices, the fundamental work of choosing what information will be included in the Clearinghouse and how that information will be collected for insertion into the database is of the highest priority as no other work can begin until these decisions have been made.

Creating the Dataset

VBBBF is responsible for setting the basic parameters of the data to be ingested by their database system, whether an RDBMS or vector database is selected. There are several methods for creating the initial dataset including 'web scraping', application program interface (API) calls to the source (where allowed), and of course, manual data entry by human beings.

The first step in creating the dataset will be the identification of the types, categories, and sources of data for use in the Clearinghouse. These sources may include information collected from government-created websites such as SAMHSA, academic resources, non-profits engaged in the areas of interest, and other credible sources of information as identified by VBBBF. After identifying the information to be collected for inclusion in the Clearinghouse, a method of retrieving this information must be selected.

"Web scraping" is the most common method for gathering data to create a dataset. Web scraping is an automated method for extracting information from websites or data stores. Web scraping software may directly access the internet using the Hypertext Transfer Protocol ("http") or a web browser. While web scraping can be done manually by a software user, the term typically refers to automated (scripted) processes implemented using a bot or web crawler. It is a form of copying in which specific data is gathered and copied from the web, typically into a central local database or spreadsheet, for later retrieval or analysis.

Newer forms of web scraping involve monitoring data feeds from web servers using APIs to automatically retrieve data from the target server. Not all servers offer an API option and those who do often charge a fee based on the number of times the data are accessed via the API. Although web scraping is a common activity that forms the basis for all content presented by search engines such as Google, there are many uses of web scraping that are considered unethical, illegal, and nefarious. Many servers actively discourage data sharing and web scraping by detecting and disallowing bots from crawling (viewing) their pages. In response, there are web scraping systems that rely on using techniques to simulate human browsing that enable gathering web page content for offline parsing. The ethics and legality of these methods is far from settled and if VBBBF is contemplating using scraping programs to gather data for inclusion in the Clearinghouse, it should also consider whether such action is authorized by reviewing the terms of use and other terms or notices posted on or made available on each site it plans to scrape.

There are several ways to approach web scraping. VBBBF could also choose a service such as Nimble to perform the scraping and data retrieval. They could also choose to hire a programmer to write a customized scraping program to retrieve these data. Either of these alternatives will be much faster than the third alternative, which is building a dataset manually, using human beings to input the data. This is, however, an option, and given the low cost of overseas labor, may be an alternative VBBBF wishes to consider. Until the scope, potential size, and composition of the dataset is determined such that an time-to-crawl can be estimated, it is impossible to provide anything other than a high-level 'guesstimate' of costs for this service . Costs in the range of $10,000 to $20,000 to crawl the data and deposit it in a central location for further processing are a reasonable guide, although this may change as the contours of the dataset are more clearly defined. Depending on the size and number of sources selected for scraping, this process could take anywhere from several days to several weeks or more.

The scraped data will likely require further 'cleaning' in order to be usable, especially if an RDBMS is selected. Additional time and costs may be incurred to clean and prepare the data for insertion into the database.

A Primer for Database Technologies

Vector Databases vs. Relational Databases

The contrast between vector and traditional relational databases (RDBMS) lies in their foundational architecture and intended use cases. Relational databases organize data into tables with rows and columns, ideal for structured data that fits well into schemas. RDBMS enforces a rigid schema, necessitating predefined structures for data. While this ensures data integrity, it may hinder adaptability to diverse and evolving data types. RDBMS databases perform well for structured

tasks such as managing user profiles and logins, storing content for use in a content management system, and managing simple functions such as storing and retrieving media assets or managing transaction data. Traditional relational databases are not well-suited to applications that involve the use of AI and in fact, we may be witnessing the beginning of the end of this type of data store for anything other than the most basic 'transactional' data due to the rapid adoption of AI technologies.

Relational databases rely on indexing the contents of the database to perform keyword search – that is, they retrieve all data that matches the exact words or phrase entered by the user. It is incumbent on the user to know what they are searching for and to choose the correct search terms to generate the desired results. All of the comparison websites referenced in this proposal are built using traditional relational database systems that are accessed using keyword searches, which is why retrieving tailored, specific information is nearly impossible without an understanding of the expected results. While this type of search may work reasonably well for researchers and policy makers, it can be incredibly frustrating for individual users and service providers who are often approaching search with a vague understanding of the topic they are investigating and have no way of predicting which keywords will produce results will be useful or relevant to them.

Content management systems ("CMS") such as WordPress or Drupal are based on an RDBMS that provides storage and management of the basic website components such as the content (About, Contact, Staff Profiles, etc), forms and their submitted data, events calendars, multimedia galleries, and other types of simple information and content. However, the RDBMS that provides basic website functionality will not provide the robust capabilities required for managing, storing, and serving a large, diverse, and growing Clearinghouse dataset. Unless the dataset is very small, a separate RDBMS will be required to manage, store, and serve the Clearinghouse data.

Vector search, on the other hand, will find relevant results even when the entered search terms do not provide exact keyword matches due to a process called 'embedding'. Embeddings encapsulate the essence of data items using a mathematical representation rather than 'keywords'. A fundamental distinction between RBDMS systems and Vector systems lies in the way data are organized. While RDBMS arranges data in tables, Vector Databases utilize multidimensional arrays. The array model allows for the efficient handling of data that has a large number of features or attributes, making Vector Databases particularly suited for scenarios involving machine learning models.

There are vector-enhanced RDBMS systems available that are able to support some level of vector search, however, their performance and function are inferior to those of pure vector databases when used for AI applications. Moreover, traditional relational databases are not designed to handle non-structured data types like images, audio, and video. These "hybrid" RDBMS databases are not scaling well and the websites using either traditional or hybrid types of RDBMS systems will need to rethink and rebuild their entire data structure in order to take advantage of AI capabilities.

The most important aspect of vector databases is their seamless integration with the AI ecosystem. Vector databases offer native support for embedding generated by popular machine learning tools, providing a bridge between AI models and database management. This integration is not only about storing vectors but also about efficiently retrieving and processing them in the context of AI applications.

AI relies on vector embeddings, a type of mathematical data representation that carries within it semantic information that's critical for the AI to gain understanding and maintain a long-term memory they can draw upon when executing complex tasks. Embeddings are generated by AI models (such as Large Language Models) and have many attributes or features that represent different dimensions of the data and are essential for understanding patterns, relationships, and underlying structures.

An example to help you think about data dimensions is the word "basketball". The data dimensions would not only be basic attributes about the basketball itself, like size, color, and brand but would also include information the places where a basketball is used, facts and information about the people who use basketballs, basketball game strategy, basketball game skill development, and so on. Each of these data points will have additional related data points that create a web of information that is synthesized by AI to organize and contextualize information in response to a query about some aspect of basketball. As you think about all of types of information that could be attached to a seemingly simple word like 'basketball', you begin to understand the depth and breadth of the data dimensioning required to make AI possible. This type of data dimensioning cannot be performed by an RDBMS; only a Vector Database provides the capabilities to perform the sophisticated mathematical operations that make dimensioning possible.

There are many factors that must be considered when choosing the type of database that will form the backbone of the Clearinghouse. An RDBMS is a safe, mature, solidly supported choice but it is not forward-looking. Choosing a Vector Database is forward-looking but has significant risks as an immature technology. We are currently experiencing a seismic shift in methods of information retrieval and presentation due to the explosion of AI. AI is moving forward at an astonishing pace. In the time it takes VBBBF to consider the options for creation of the Clearinghouse dataset, "next-gen" Vector Database technologies are already rolling out at breakneck speeds. Several emerging trends in Vector Databases are especially noteworthy:

Serveless Databases

Serverless represents the next evolution of vector databases. First-gen architectures created vector database architecture that is accurate, fast, scalable, but expensive. With the rise of AI use cases where cost and elasticity are increasingly important, a second generation of serverless vector databases is currently rolling out and available for production use.

First-generation vector DBs have three critical pain points that a serverless vector database solves:

• **Separation of storage from computation:** To optimize costs, computation should only be used when needed. That means decoupling the database's index storage from queries and searching only what is needed.

• **Multitenancy**: Ensuring infrequently queried data do not increase costs by remaining 'always on'. In laymen's terms, this means that the database keeps certain data 'warm' because it is frequently accessed while keeping other data 'cold' because it is rarely used. This reduces costs by reducing the amount of data that must be maintained in an 'always on' state.

• **Freshness:** A Vector Database needs to provide fresh data, meaning within a few seconds of inserting new data, it is queryable. The freshness layer acts as a temporary "cache" of vectors that can be queried until the index builder places the new vectors into the index.

Taking a first-generation vector database and adding separation of storage from computing, multitenancy, and freshness results in the next-gen of modern Vector Databases. This architecture is preferred for the modern AI stack, at least in the moment. Performance improvements continue apace and it will not be long before we are on our third-, fourth-, and fifth-generations of Vector Database architecture. This highlights the risks inherent in making choices about any new technology – whatever you choose today is very likely to be supplanted by something faster, better, and cheaper tomorrow.

Retrieval-Augmented Generation (RAG)

Vector databases are also used to implement Retrieval-Augmented Generation ("RAG"), an approach that has been developed over the past year to improve the usefulness of large language model (LLM) systems such as ChatGPT. RAG solves two important problems with these chatbot systems. First, systems implementing RAG can provide users with the "references" that back up the LLM's answers. Second, RAG can help keep the ChatBot systems on target and aligned with the goals of the organization deploying the LLM.

With RAG, the search engine first takes the user's question and searches through a large set of documents, finding parts of each document that might be responsive. These document parts and the user's question are then provided to the LLM with a prompt that says something along the lines of, "given the following documents, please answer this question."

RAG might seem like the obvious solution for customizing an LLM. But RAG development comes with a steep learning curve and requires specialized technical knowledge. Due to the newness of RAG technology, there does not yet exist a body of best practices nor is there a large pool of candidates with the requisite knowledge to implement this technology effectively. That said, RAG is rapidly supplanting "First Gen" AI data architecture and VBBBF should have a basic understanding of its benefits in order to inform its decision-making process.

The RAG chain takes the user query (text) and retrieves relevant data from the database, then passes that data and the query to the LLM in order to generate a highly accurate and personalized response.

There are a lot of complexities in this architecture, but it does have important benefits:

- It grounds your LLM in accurate proprietary data, thus making it so much more valuable.
- It brings your models to your data rather than bringing your data to your models, which is a relatively simple, cost-effective approach.
- Efficient cost and scaling enable lower-cost knowledge support for AI.

RAG is already a reality in the Modern Data Stack. The biggest players are working at a breakneck speed to make RAG easier by serving LLMs within their environments, where organizational data are stored. Snowflake Cortex now enables organizations to quickly analyze data and build AI apps directly in Snowflake. Databricks' new Foundation Model APIs provide instant access to LLMs

directly within Databricks. Microsoft released Microsoft Azure OpenAI Service and Amazon recently launched the Amazon Redshift Query Editor. RAG is of interest to VBBBF not only because of its improved accuracy and lower costs, but because of its ability to provide attributions to the source material, enhancing the credibility of the information produced by its AI model.

Challenges of Modern Database Solutions

Both serverless vector databases and RAG are examples of technologies that are rapidly becoming standards for the development of AI infrastructures, even though they have been introduced to the public-at-large only in the last 12-18 months. This is the challenge inherent in selecting a technology to support an AI-focused website; any choice you make is likely to be at least somewhat obsolete in a very short time, meaning that multiple rounds of investment in improvements may be required to continue to keep pace with a rapidly evolving technical landscape.

Those familiar with RDBMS as they are currently used in application development may have the impression that a database is just like a car – turn the key and go. This is because issues like backup, data integrity, data management, and performance tuning are so well-understood and well-documented that database users and owners are barely aware of them. Standard procedures exist for every aspect of RDBMS management and there is a large pool of capable database professionals to manage these systems. However, a vector database is not like that.

Vector Databases have many issues related to the newness of technology. As compared to the mature RDBMS systems, vector databases often lack robust data management capabilities and transactional support, making it more challenging to ensure data integrity, consistency, and scalability. Modifying or deleting data in Vector Databases can be challenging, as dimensional data are highly complex and small changes can create outsized impacts. There have been not-infrequent reports of unrecoverable data loss in Vector Databases, which is obviously problematic and will require an additional layer of technology to provide backup and recovery. There is a dearth of professionals with the requisite skills to manage Vector Databases, though this is expected to improve dramatically over the next several years. Given the early stage of adoption and limited shared experiences, any decisions made by database professionals regarding the management of Vector Database technologies will be based on incomplete and rapidly evolving information and the environment will often be one of ‘figure it out as you go’. This is obviously not ideal for an organization lacking the deep pockets that will be required to compensate for the lack of mature features and low-knowledge environment surrounding the use of Vector Databases.

Estimated Costs to Develop the Clearinghouse Dataset

As stated previously, it is not possible to develop anything other than a high-level ‘guesstimate’ of costs to develop the Clearinghouse dataset, due to the lack of specifics about the size and composition of the Clearinghouse data. It is not unreasonable to expect to spend in the range of $5,000 - $25,000 simply to identify, extract, and organize the resources for use by the selected database technology.

Summary of Next Steps in Developing the Clearinghouse Dataset

- Identify sources of data to be included in the Clearinghouse.
- Determine the method for collecting and organizing the data, including addressing accuracy and data redundancy.
- Take any additional steps such as 'cleaning' to make the data usable.
- Select a database technology.
- Populate the database with the Clearinghouse data.
- Index the data and prepare if for use with the selected retrieval technology.

If VBBBF decides to forgo AI capabilities in favor of a more traditional website such as those presented in the examples, an RDBMS system will be an appropriate choice. If data warehousing capabilities are also required – and again, this cannot be ascertained without an understanding of the size and composition of the dataset - initial development costs for the database could be in the range of $30,000 to $150,000. In addition to the costs of preparing the dataset and the database for production use, VBBBF will incur monthly costs that include database hosting, backup and security, and the engagement of a full or part-time database professional to ensure the ongoing accuracy, availability, and reliability of the data. These ongoing costs could be in the range of $1,000 to $10,000 per month, but as with estimating the initial cost, it is impossible to provide an accurate picture of the total costs of ownership without a detailed specification for the dataset.

If VBBBF commits to the inclusion of AI capabilities for the Clearinghouse, a Vector Database will be required and its cost will be included in the overall development costs for the AI functionality, which are discussed in the next section. Note that including "next-gen" capabilities such as 'serverless' or RAG in the specification will likely increase costs above the high-level estimates provided in this proposal.

Information Tailoring

What is "Information Tailoring"?

Information Tailoring is a means of customizing messaging to meet the needs and values of specific target audiences. Tailoring aims to enhance the relevance of the information presented and thus to produce the desired changes in response to the messaging. Many studies have shown that tailoring health messages to personal values, beliefs, preferences, and characteristics increases the perceived relevance of the information, thereby increasing the motivation to pay attention to the information.

The primary goal of Information Tailoring is to 'trap' the attention of the user. Tailoring strategies are used to increase the likelihood of behavioral modification in response to the received communication and may include content specifically designed to elicit an emotional response such as fear or hope.

Information Tailoring can also contain contextualization, for example, framing information through a localized lens of location, language, or culture, through the lens of age and/or gender, or by inferences made through the self-identification of a user within one of the target cohorts. For example, messaging about fentanyl use would be presented in different contexts depending on

whether the user has identified themselves as an individual, a caregiver, a member of a healthcare organization, or as a policy maker.

Information Tailoring Strategies

Primary strategies used in Information Tailoring include:

- **Personalization:** When individuals perceive messages to be personally relevant, they are more likely to be motivated to process information more deeply. Personalization often uses shared characteristics, such as age, geographic location (urban vs. rural), or lifestyle/cultural preferences to craft a compelling message for a large subset of users.
- **Content Preference Matching:** When users can control the mode by which they receive messaging, i.e. through text, audio, or visual modes, the messaging is more likely to be evaluated, processed, and remembered. Additionally, having control over the pace and order of the content has been shown to improve retention.
- **Feedback:** Creating feedback loops that use information from past actions to influence future actions in a cycle that encourages continuous improvement.

Extensive Information Tailoring will be needed for the areas of the website that include gamified elements, outreach/marketing campaigns, and for the results of keyword searches performed against an RDBMS, in order to present relevant, motivating messaging to each of the target cohorts. Less Information Tailoring will be required for a website based primarily on AI-based interaction as the customization of messaging is built into the AI response mechanism.

Developing tailored healthcare messaging is a costly process, especially as methods of tailoring become more sophisticated. Content tailoring increases development time and costs due to the creation of a much larger message library of customized content, the formulation of tailoring criteria and algorithms, and the addition of programming to support the selection and presentation of content according to the defined criteria for each of the target cohorts.

Provision of Healthcare Tailoring Expertise

Because PWC does not have expertise in the realm on Information Tailoring in a healthcare setting, we consulted with Dr. Juan Sebastian Muhamad, who has specialized knowledge in this field and whose doctoral dissertation is titled “Exploring Factors Related to Mental Health Information and Mental Health Help Seeking in Young Adults” Dr Muhamed has agreed to provide consultation services at the rate of $125/hour for this project to:

- Create a website design that tailors the presentation of Clearinghouse information and messaging for each of the target cohorts.
- Create gamified elements for the website that are tailored to engage each of the target cohorts.
- Provide assistance in tuning search results, especially those based on “keyword search”, to provide appropriate responses to queries from each of the target cohorts.

Because no specific information is provided in the RFP about the types of messaging to be provided for each of the target cohorts, the specifications for any gamified elements, or the selection of the

presentation mechanism for Clearinghouse search results, it is impossible to provide an estimate of the costs or time frame that will be required for the Information Tailoring.

Gamification

Gamification has emerged as a novel approach for improving mental health interventions, with two major applications in current practice. One approach involves directly developing gamified mental health interventions to create novel and more engaging treatments, whereas the other involves integrating gamification elements into existing mental health interventions to enhance their efficacy.

The following benefits are just some of the many ways in which gamification tools enhance learning and training content:

- Allows learning to be fun and interactive.
- Provides the opportunity to see real-world applications of learning content.
- Real-time feedback that encourages continuous improvement.
- Friendly competition can improve motivation to learn.

The essential first steps are to explore what behaviors you want to promote and what learning strategies and training objectives you are trying to achieve. Once objectives are established, you can proceed to set up your game mechanics which can include:

- Points
- Badges/Achievements
- Levels
- Progress bars
- Leaderboards
- Rewards

One approach to incorporating gamification into mental health interventions, as discussed on the NIH website, is to add gaming elements to improve the effectiveness of-behavioral intervention designed for clients diagnosed with depression. These elements may include quizzes, matching games, progress bars, challenges, and other 'gamified' presentations of information. The gamification components studied by the NIH were well-received by clients and suggest that clients generally accept the incorporation of gamification into existing treatments for mitigating depressive symptoms.

Gamification With Avatars

An Avatar is a character who can act as a guide through a game or a process. Those of us old enough remember Clippy the helpful/annoying assistant included in early versions of Microsoft Office understand both the potential usefulness and annoyances of Avatars. As a concept, the intention of providing help through a friendly character is laudable, though having the character randomly pop-up makes it annoying to many users. Some organizations forgo character representations entirely and opt instead for text-based tools with speech bubbles indicating what you should do next while others use character avatars representing the organization to guide

people through a process. In some learning-related quests, relatable or specifically chosen characters give instructions. Examples of the use of avatars for the VBBBF may include an avatar representing Vin Baker giving advice and information to visitors, an avatar representing the 'demon' of addiction who can be defeated through positive actions taken by a visitor, or a customized avatar created by the visitor to represent themselves as they embark on their quest for well-being.

To make therapy more accessible, certain apps introduce avatars as virtual therapists. Users can interact with these avatars, which guide them through exercises, meditations, or even cognitive behavioral therapy sessions. This can make the therapy app feel more personal and relevant to a user, driving engagement by creating an emotional connection to the "therapist".

Unlike standardized designs, the world of avatars thrives on diversity. Users aren't restricted to a one-size-fits-all model; they're presented with an array of choices, each allowing them to infuse a part of their identity into their digital counterpart. This act of choice offers visitors a harmonious blend of creative freedom and immediate feedback. Whether it's adjusting facial features, selecting unique attire, or even changing skin tones, the customization options are vast, enabling users to exercise their creativity and see their choices come to life.

Potential Uses of Avatars

Avatars can serve as personification of either something desired or undesired. "Away from" and "towards" motivation are strong reasons why we do or don't do some activities. Gamification can be used to steer people away from undesirable behaviors and towards desired behaviors. By creating a character that represents this undesirable, we can create an engaging narrative to help people build a new habit. An example of an "away from" avatar would be developing a character who represents and externalizes the "monster within" that individuals facing substance abuse and mental health issues experience. The goal is to keep the monster in check by engaging in guided behavior modifications designed to produce positive steps towards recovery that reduce the 'control' of the monster over the individual..

At the crux of the avatar's appeal lies a powerful psychological mechanism: the core drive for ownership and possession. Users, when presented with an avatar that they can customize, inherently feel a sense of ownership. This digital representation becomes a part of their identity, a virtual self that they can mold, nurture, and showcase. Many users choose avatars that mirror their real-world appearance, while others opt for idealized or aspirational versions of themselves. This connection is so potent that it often translates into heightened engagement and loyalty to the platform.

Further extending the avatar's psychological reach is its ability to foster social connections. Beyond serving as mere representations, avatars have the potential to redefine user interactions and experiences. They can be tailored to facilitate social interactions, allowing users to communicate, collaborate, and even compete with one another. Avatars can play a pivotal role in fulfilling the human need for social influence and connection. In digital spaces, where users might be hesitant to reveal their true identities due to vulnerability concerns, avatars offer a safe middle ground. Instead of uploading personal photos, which might feel too revealing, or choosing impersonal images like landscapes, which lack identity, avatars provide a personalized yet guarded

representation. They strike a balance, ensuring users can be sociable and expressive without feeling overexposed.

Moreover, the avatar's inherent flexibility paves the way for diverse experiences. Platforms can introduce customization options, challenges, and narratives centered around avatars, enhancing user engagement and retention. The possibilities are virtually limitless, from unlocking unique abilities and features to participating in virtual events and competitions.

The realm of gamified design has seen a surge in the implementation of avatars. While the primary function of avatars is to symbolize users in digital spaces, a deeper exploration reveals their potential to tap into various core drives, enhancing user engagement and transforming the overall experience.

The Playfulness Dilemma

One of the primary reservations surrounding avatars and other gamified elements is their inherent playfulness. A gamified environment can bring a sense of whimsy and fun to otherwise serious subject matters. However, this very characteristic can sometimes be their downfall. For platforms targeting users from cultures or professional backgrounds that emphasize formality, the introduction of avatars can come across as overly playful or even childish. In such scenarios, the playful nature of avatars can divert users from the platform's primary objectives, making them feel like they're merely "playing a game" rather than engaging in a serious activity. It's imperative, therefore, for designers to make extensive use of Information Tailoring to ensure that avatars and other gamified elements align with the platform's intended tone and purpose.

Creating dynamic avatars and other gamified elements that offer customization and interactivity can elevate the user experience and dramatically improve user engagement and 'time-on-platform'. The features also necessitate a significant investment of time and money including the development of custom artwork and intricate programming to allow users to customize their avatars.

Despite the challenges, avatars, when executed correctly, can be a game-changer in gamified design. Gamification and avatars cater to an audience with a playful mindset, foster a sense of community where users can socialize and view each other's avatars, and ensure longevity in the experience so users can form a lasting bond with their avatars and with each other. When these elements align, avatars can truly be a vehicle for engaging users and encouraging them to use the tools and resources providing by the Clearinghouse.

Tailoring Gamification and Information Elements

A study examining the use of gamification elements in unguided cognitive-behavioral therapy for adolescents diagnosed with depression uncovered an important issue as it relates to gamification designed for an adolescent cohort. The gamification elements included playful quizzes about therapeutic content, puzzles, and mini-games involving rewards. Teens expected the game formats to be as competitive and engaging as commercial games, highlighting an important challenge for the gamification landscape as it relates to healthcare. The availability of high-quality entertainment games has raised clients' expectations for gamification components in mental health interventions, especially among adolescents who frequently engage in leisure gaming. The

implication is that gamification elements, especially those designed for a younger cohort, will require a substantial investment in order to reach the level of sophistication expected by teens who have spent their formative years playing sophisticated commercial games and have performance expectations that are much higher than other target cohorts.

Other studies have found gender differences in gaming motivation, with male gamers generally exhibiting greater motivation for competition and challenge than female gamers. To enhance motivation and engagement in gamified interventions, game developers often implement reward systems such as game points, badges, medals, trophies, and leaderboards. These systems primarily emphasize the recognition of accomplishments through competitive, tangible rewards which are typically more appealing to male gamers who tend to have a stronger drive for achievement and competitiveness. Hence, gamified interventions that use tangible reward systems to reinforce competitiveness may be effective in motivating male clients and providing them with a more engaging and rewarding experience. In contrast, female gamers tend to have different motivational needs and preferences, such as being more inclined toward cooperative gaming and games that emphasize social affiliations. Hence, it is necessary to explore alternative approaches to gamification that align better with the motivations and preferences of both male and female visitors.

Gamification elements will require extensive Information Tailoring in order to be positively received by the target audiences. Because VBBBF has defined an extremely broad group of target cohorts, Information Tailoring will be critical in driving adoption of any gamification elements among the identified target groups. However, a lack of clearly defined specifications, objectives, and desired outcomes makes it impossible to determine what elements of the website should be gamified and how the gamification should be implemented, making any cost or time-to-launch estimates impossible.

A Guide to Artificial Intelligence ("AI")

VBBBF has expressed an interest in deploying AI to create a suite of services for accessing Clearinghouse data and resources. Plato Wynne Consulting ("PWC") took several meetings with Abdallah Musmar, a professor in the areas of machine learning and AI strategy at the University of South Florida, to discuss the development and costs of an AI-based system to perform the Clearinghouse functions. Professor Musmar provided useful information about the fundamental aspects of AI as well as some broad recommendations and cost estimates related to its use in developing the Clearinghouse. His recommendations must be considered against a backdrop of a multitude of choices regarding the design and implantation of an AI-based system in a rapidly evolving AI ecosystem. In order to understand the opportunities and challenges of any AI-based system, as well as the choices and trade-offs that must be considered, it is imperative to have a basic understanding of the current state of AI technology. An extensive primer for AI technology is provided as part of this RFP to help guide VBBBF in its decision-making process.

Introduction to AI

Artificial intelligence ("AI") is technology that enables computers and machines to simulate human intelligence and problem-solving capabilities. On its own or combined with other technologies (e.g., sensors, geolocation, robotics) AI can perform tasks that would otherwise require extensive human intelligence or intervention. Digital assistants, GPS guidance, autonomous vehicles, and generative AI tools (like Open AI's Chat GPT or Google/Bing Search) are just a few examples of AI in the daily news and our daily lives.

As a field of computer science, artificial intelligence encompasses (and is often mentioned together with) machine learning and deep learning. These disciplines involve the development of AI algorithms, modeled after the decision-making processes of the human brain, that can 'learn' from available data and make increasingly more accurate classifications or predictions over time. It is important to note that no one, not even the developers of these technologies, understand exactly how AI 'learns'. A new field of study is arising that is attempting to gain insight into how AI 'learns' but at this point in time, AI remains a "black box" whose inner workings are mysterious and poorly understood.

Artificial intelligence has gone through many cycles of hype, but even to skeptics, the release of ChatGPT seems to mark a turning point. The last time generative AI loomed this large, the breakthroughs were in computer vision, but now the leap forward is in natural language processing ("NLP"). Today, generative AI can learn and synthesize not just human language but from other data types including images, video, software code, and even molecular structures.

Applications for AI are growing every day. But as the hype around the use of AI takes off, conversations around the uses of AI, its implications for what it means to be "human", the ethics of the various uses of AI, and the responsibility to prevent AI from doing more harm than good have become critically important. These conversations all have implications for the provision of mental healthcare using AI and therefore, are an important consideration for the VBBBF decision makers. In order to understand the costs, benefits, and trade-offs in using AI for the Clearinghouse, this proposal provides review and discussion of some fundamental AI concepts.

What is Machine Learning?

Machine learning ("ML") is a subset of artificial intelligence that uses algorithms to analyze data, learn from it, and make predictions and informed decisions based on the data.

Instead of needing to explicitly program every single rule for how a computer should behave, machine learning algorithms are designed to automatically learn and improve from experience. The ML models learn by examining training data and using statistical techniques to identify patterns, extract insights, and find relationships between inputs and outputs.

As new data is fed to the algorithms, they continue updating and improving their capability to make accurate predictions and analyses. Machine learning enables computers to solve problems in a way that reflects the natural learning processes of humans. The large datasets that ML algorithms use would be nearly impossible to program by hand.

What is "Deep Learning"

Machine learning and deep learning are sub-disciplines of AI, and deep learning is a sub-discipline of machine learning.

Both machine learning and deep learning algorithms use neural networks to 'learn' from huge amounts of data. These neural networks are programmatic structures modeled after the decision-making processes of the human brain. They consist of layers of interconnected nodes that extract features from the data and make predictions about what the data represents.

Machine learning and deep learning differ in the types of neural networks they use, and the amount of human intervention involved. Classic machine learning algorithms use neural networks with an input layer, one or two 'hidden' layers, and an output layer. Typically, these algorithms are limited to supervised learning: the data needs to be structured or labeled by human experts to enable the algorithm to extract features from the data.

Deep learning algorithms use deep neural networks—networks composed of an input layer, three or more (but usually hundreds) of hidden layers, and an output layout. These multiple layers enable unsupervised learning: they automate extraction of features from large, unlabeled and unstructured data sets. Because it doesn't require human intervention, deep learning essentially enables machine learning at scale. Again, it is important to understand that this deep learning is currently a 'black box', that no one – not even its creators - understands.

Key Differences Between Machine Learning vs. Generative AI

Machine learning and Generative AI both learn from data, but their purposes and strategies differ. Here's how:

- **Goal**: Machine learning is focused on analyzing data to find patterns and make accurate predictions. GenAI, on the other hand, is focused on creating new data that resembles training data.
- **Training Techniques:** Machine learning models use a variety of approaches depending on the use case, such as supervised, unsupervised, and reinforcement learning. Generative AI mainly uses techniques that involve dual-learning. One part learns to generate data and the other part learns how to critique it.
- **Outputs:** Machine learning models output inferences, classifications, or predictions based on learned relationships. Generative AI models output completely new assets like text, image, or music.
- **Performance Metrics**: The success or failure of a machine learning model is based on predictive accuracy metrics like precision and recall. GenAI models are judged on qualitative metrics assessing realism, coherence, and diversity.

Machine Learning Foundation Models

Prebuilt ML models that have been extensively trained on data are known as Foundation Models ("FM"). FMs can be tailored to a variety of downstream activities with additional training on datasets designed for specific tasks. In short, a Foundation Model is a type of machine learning or deep learning model that is trained on broad and diverse data so that it can be applied across a

wide range of use cases. Foundation Models serve as a starting point for more specialized and sophisticated models. Foundation Models have transformed AI, powering prominent generative AI applications like ChatGPT1.

The Foundation Model has diverse applications:

- **Chatbots:** It can engage in natural conversations with users.
- **Content Creation:** It writes essays, poems, stories, and more.
- **Code Generation:** It can produce code snippets.
- **Language Translation:** It translates text between languages.
- **Creative Writing:** It invents fictional characters and scenarios.

The Foundation Model leverages its pretrained knowledge and fine-tuning to generate human-like text, making it a versatile tool for the aggregation and dissemination of data from multiple, disparate resources.

Legal Environment of Foundation Models

The Stanford Institute for Human-Centered Artificial Intelligence's (HAI) Center for Research on Foundation Models (CRFM) coined the term "Foundation Model" in August 2021 to mean "any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks". This was based on their observation that preexisting terms, while overlapping, were not adequate, stating that "'(large) language model' was too narrow given [the] focus is not only language; 'self-supervised model' was too specific to the training objective; and 'pretrained model' suggested that the noteworthy action all happened after 'pretraining." The term "Foundation Model" was chosen over "foundational model" because "foundational" implies that these models provide fundamental principles in a way that "foundation" does not. After considering many terms, they settled on "Foundation Model" to emphasize the intended function (i.e., amenability to subsequent further development) rather than modality, architecture, or implementation. As governments regulate Foundation Models, new legal definitions have emerged.

- In the United States, the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence defines a foundation model as "an AI model that is trained on broad data; generally uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts".
- In the United States, the proposed AI Foundation Model Transparency Act of 2023[20] by House Representatives Don Beyer (D, VA) and Anna Eshoo (D, CA) defines a foundation model as "an artificial intelligence model trained on broad data, generally uses self supervision, generally contains at least 1,000,000,000 parameters, is applicable across a wide range of contexts, and exhibits, or could be easily modified to exhibit, high levels of performance at tasks that could pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters."
- In the European Union, the European Parliament's negotiated position on the E.U. AI Act defines a foundation model as an "AI model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks".

- In the United Kingdom, the Competition and Markets Authority's AI Foundation Models: Initial Report [1] defines foundations model as "a type of AI technology that are trained on vast amounts of data that can be adapted to a wide range of tasks and operations."

Overall, while many of these definitions stick close to the original Stanford definition, they do introduce some subtle distinctions. For example, the U.S. definitions are the sole definitions to make reference to the size of a foundation model, though they differ on an exact magnitude. Beyer and Eshoo's definition also specifies that foundation models must achieve a level of performance as to be a potential danger. In contrast, the E.U. definition includes mention of whether the model is designed for generality of output. Nonetheless, all definitions share that Foundation Models must be trained on a broad range of data with potential applications in many domains.

How Are Foundation Models Created?

Creating Foundation Models involves using massive amounts of unlabeled data to train a model, which is subsequently applied to a variety of applications. “Pretraining” is the process of converting unlabeled data to the Foundation Model (“FM”).

A Foundation Model represents words as dense vectors (embeddings) in a high-dimensional space, meaning each word is mapped to multiple meanings and context, with each mapping representing a dimension of the word. An “attention mechanism” allows the model to focus on relevant parts of the input sequence during both training and inference. Embeddings and attention weights enable the model to understand context and relationships.

Foundation models are poised to significantly change the machine learning lifecycle. Although it currently costs millions of dollars to develop a Foundation Model from scratch, they’re useful in the long run. It’s faster and cheaper for data scientists to use pre-trained FMs to develop new ML applications rather than train unique ML models from the ground up.

Examples of Foundation Models

- **BERT:** Released in 2018, Bidirectional Encoder Representations from Transformers (BERT) was one of the first foundation models. BERT is a bidirectional model that analyzes the context of a complete sequence then makes a prediction. It was trained on a plain text corpus and Wikipedia using 3.3 billion tokens (words) and 340 million parameters (modifiers). BERT can answer questions, predict sentences, and translate texts.
- **GPT:** The Generative Pre-trained Transformer (GPT) model was developed by OpenAI in 2018. It uses a 12-layer transformer decoder with a self-attention mechanism. and was trained on the BookCorpus dataset, which holds over 11,000 free novels.

  GPT-2 released in 2019. OpenAI trained it using 1.5 billion parameters (compared to the 117 million parameters used on GPT-1). GPT-3 has a 96-layer neural network and 175 billion parameters and is trained using the 500-billion-word Common Crawl dataset. The popular ChatGPT chatbot is based on GPT-3.5. And GPT-4, the latest version, launched in late 2022 and successfully passed the Uniform Bar Examination with a score of 297 (76%).

- **Amazon Titan:** Amazon Titan FMs are pretrained on large datasets, making them powerful, general-purpose models. They can be used as is or customized privately with company-specific data for a particular task without annotating large volumes of data. Initially, Titan will offer two models. The first is a generative LLM for tasks such as summarization, text generation, classification, open-ended Q&A, and information extraction. The second is an embeddings LLM that translates text inputs including words, phrases, and large units of text into numerical representations (known as embeddings) that contain the semantic meaning of the text. While this LLM will not generate text, it is useful for applications like personalization and search because by comparing embeddings the model will produce more relevant and contextual responses than word matching. To continue supporting best practices in the responsible use of AI, Titan FMs are built to detect and remove harmful content in the data, reject inappropriate content in the user input, and filter the models' outputs that contain inappropriate content such as hate speech, profanity, and violence.
- **Claude:** Claude 2 is Anthropic's state-of-the-art model that excels at thoughtful dialogue, content creation, complex reasoning, creativity, and coding, built with Constitutional AI. Claude 2 can take up to 100,000 tokens in each prompt, meaning it can work over hundreds of pages of text, or even an entire book. Claude 2 can also write longer documents—like memos and stories on the order of a few thousand tokens—compared to its prior version.
- **Hugging Face:** Hugging Face is a platform that offers open-source tools to build and deploy machine learning models. It acts as a community hub, and developers can share and explore models and datasets. Membership for individuals is free, although paid subscriptions offer higher levels of access. There is currently public access to nearly 200,000 models and 30,000 datasets.

Training Foundation Models

Foundation Models are built by optimizing a training objective(s), which is a mathematical function that determines how model parameters are updated based on model predictions on training data. Foundation Models are trained on a large quantity of data, working under the maxim "the more data, the better". Performance evaluation does show that more data generally leads to better performance, but other issues arise as data quantity grows. Tasks like managing the dataset, integrating data across new applications, ensuring adherence to privacy and security safeguards, and maintaining data quality all become more difficult as data size grows. The specific demands of Foundation Models have only exacerbated such issues, as it remains the norm for large Foundation Models to use public web-scraped data. Public web data remains a plentiful resource, but it also demands stringent moderation and data processing from Foundation Model developers before it can be successfully integrated into the training pipeline.

"Pretraining" is the creation of an FM by training a model with terabytes of unlabeled text or multi-modal data (such as images, audio, or video). During pretraining, the model is exposed to a vast amount of unlabeled text data and the goal is to learn patterns, context, and relationships within this data. Specifically, the model is trained to predict missing words or the next words in

sentences. It learns to generate coherent text by understanding the context provided by preceding words.

The quantity and quality of training data as well as the training infrastructure are critical components in the development of a large model. Before AI can be deployed, useful, highly reliable, consolidated, well-documented datasets must be created using a modern data platform. Modern data structure and platforms are discussed in the Datasets section.

Training Foundation Models often runs the risk of violating user privacy, as private data can be disclosed, collected, or used in ways beyond the stated scope. Even if no private data is leaked, models can still inadvertently compromise security through learned behavior in the resulting Foundation Model. Data quality is another key point, as web-scraped data frequently contains biased, duplicate, and toxic material. Information about web scraping is presented in the Datasets section. Once Foundation Models are deployed, ensuring high-quality data is still an issue, as undesirable behavior can still emerge from small subsets of data.

After pretraining, the Foundation Model is fine-tuned on specific downstream tasks. Interacting with users seeking mental healthcare and substance abuse information is an example of a "downstream task". Fine-tuning adapts the pretrained model to perform specific tasks, such as text completion, translation, or question answering. During fine-tuning, the model learns task-specific patterns and biases. In order to fine-tune an AI system, organizations must develop specific task scenarios along with desired outcomes in order to guide the system towards producing high-quality responses.

When given a prompt, the Foundation Model generates coherent and contextually relevant text. It extrapolates from its learned patterns to create new sentences, paragraphs, or even entire articles. Users can interact with the model by providing prompts, and it responds with generated content. The search space has been completely upended by the power of LLMs to find exactly what users want based on the meaning of their queries versus matching keywords. This upending is relevant to the Clearinghouse because Google, Bing, Meta, and OpenAI have all deployed AI to produce results for searches that are in the domain of the Clearinghouse. The results produced by Clearinghouse searches will be 'competing' against the results produced by these search 'giants', who are already ahead of the Clearinghouse in every way, from the construction of their datasets to the deployment of advanced AI capabilities.

Foundation Model Challenges

Foundation Models can coherently respond to prompts on subjects they haven't been explicitly trained on. But they have certain weaknesses. Here are some of the challenges facing foundation models:

- **Infrastructure requirements.** Building a Foundation Model from scratch is expensive and requires enormous resources, and training may take months.
- **Front-end development**. For practical applications, developers need to integrate Foundation Models into a software stack, including tools for prompt engineering, fine-tuning, and pipeline engineering.

- **Lack of comprehension.** Although they can provide grammatically and factually correct answers, Foundation Models have difficulty comprehending the context of a prompt. It is important to remind ourselves that these are machines and as such, they aren't socially or psychologically aware.
- **Unreliable answers.** Answers to questions on certain subject matter may be unreliable and sometimes inappropriate, toxic, or incorrect. Unreliable answers are commonly known as "Hallucinations", which are discussed in the section titled Issues With LLM Output Reliability ("Hallucinations").
- **Bias.** Bias is a distinct possibility as models can pick up hate speech and inappropriate undertones from training datasets. To avoid this, developers should carefully filter training data and encode specific norms into their models.

Choosing a Foundation Model

Open Source VS Proprietary Models

The open-source approach of freely distributing technology for the public to use, share and modify helped create the modern internet and cloud-computing. Open-source AI models are appealing because they offer a means of using Large Language Models without paying and sharing data with a vendor like Microsoft. And because they are shared for public dissection, open-source models typically include their inner workings, which companies can use to build their own models customized to meet their specific needs.

On the other hand, proprietary models like OpenAI or Google's Bard and Gemini are more like black boxes. They don't let you see or change how they work, but they usually perform well because they are backed by companies who can underwrite the substantial upfront costs of creating an FM. Initial costs required to train their models can run into the hundreds of millions of dollars, so these companies protect their investment by setting up paywalls that control who uses their models and how.

How a model handles your data is important, especially when working with sensitive, health-related data. With open-source models, you can keep all your data in-house. Assuming that data controls and proper security measures are in place, open source models offer greater privacy than proprietary models. Proprietary models store your data on their servers, presenting multiple risks, including the accidental exposure of your data and the retention and use of your data without your permission as inputs for the creation and improvement of other Foundation Models produced by proprietary model makers.

When evaluating a model, you must consider what data it was trained on and who made it. Models trained on unverified data can be biased or raise ethical flags. The model's origins, like where it was made and who made it, can also influence the outcomes of any downstream activities, including whether accurate responses are provided to user queries. An example in the news as recently as May 22, 2024 is the acquisition of News Corp resources by OpenAI. As the owner of the Wall Street Journal, Fox News, and other outlets considered to have a 'right-leaning' bias, it can be expected that outputs from future OpenAI models will contain at least some of the biases of these organizations. Searching for information such as 'election integrity' using a system that is built

using an OpenAI Foundation Model is likely to produce vastly different results than a model that has not been trained so heavily on News Corp information.

Also of concern is the speed of the model and whether it can handle scaling up. The level of speed and 'scale up' needed will be determined by the estimated number of users during various stages of the Clearinghouse lifecycle, from launch through maturity. Questions regarding speed can be discussed after estimates about the dataset size and the number of daily or monthly queries have been made.

How much power a model uses also matters because of its environmental footprint. There are countless ecological and social impacts in the creation and maintenance of AI data infrastructure, from the conflicts and conditions of the workers mining the valuable materials that go into making chips to the pollution caused from the processing of these materials. There are also reports of underpaid, overexploited workers being used to support the responses in Generative AI systems that raise serious ethical concerns. OpenAI, the creators behind the wildly popular ChatGPT, say that the energy used to train the average AI model will increase 10 fold each year. Some believe machine learning is on track to consume all the energy that can be supplied.

Considerations when using Public AI Models

Generative AI models like ChatGPT are available for companies to leverage as-is. While the time-to-launch and the barriers to entry are incredibly low (all you need is an account and the ability to ask questions to get started), public models are quite generic and pose multiple risks.

Because ChatGPT has been trained on the entirety of the internet, it struggles with specialization, nuance, and complex "thinking," especially when tasked with niche problems. A human expert in the field is necessary to guide ChatGPT to useful outputs, to verify them, and to turn the raw text into a polished final version.

Additionally, organizations risk copyright infringement if models generate content wholesale from a copyrighted source—a problem that's not often apparent to the end user. Organizations also risk open-sourcing their software or content if it's not sufficiently human-directed, as AI outputs cannot be copyrighted as of May 2023.

There are other risks besides inaccurate or even harmful results and various legal liabilities, including copyright infringement. While OpenAI does allow organizations to turn off logging, the default ChatGPT settings capture inputs and outputs for training data—meaning proprietary information can be accessed by other parties without your permission.

Organizations should also consider the implications of relying too much on public models. Because the outputs tend to be generic (even with advanced prompting), organizations risk losing a competitive advantage if other entities take advantage of these models in the same way. Even if content outputs aren't identical, public models struggle with true innovation that can set your product and organization apart.

Considerations for using Customized AI Models

Custom AI models are trained on your data, enabling much more subtle and nuanced responses specific to your needs. With custom models, AI can act in a much more advanced capacity, replicating the work of a human in the field.

For example, a chat support feature could be enhanced with custom models trained on an entire library of customized information relating to substance abuse and mental healthcare. Armed with internal knowledge, the chatbot can serve visitors reliable, credible, and accurate information in a way that simply isn't possible with public models.

Custom models allow an organization to fully control all aspects of the AI application including controlling data sources and inputs, customized model training, limiting access to the underlying data, and greater oversight of the overall privacy and security infrastructure surrounding the AI application. However, there is a trade-off as custom models also require technical resources that are scarce, expensive and time-consuming. An organization considering a custom model must be prepared for a large upfront investment of both time and money before any benefit is realized as well as large ongoing costs to sustain their AI ecosystem.

Types of AI and Their Uses

Generative AI ("GenAI") is a category that contains a myriad of tools built to use information from Large Language Models ("LLM") and other types of AI models using machine learning to generate new content. An LLM is a type of AI model that uses machine learning built on billions of parameters to understand and produce text.

Three major things stand out when you compare Generative AI and LLMs.

1. Not all Generative AI tools are built on LLMs, but all LLMs are a form of Generative AI.

Generative AI is a broad category for a type of AI, referring to any artificial intelligence that can create original content. Generative AI tools are built on underlying AI models, such as a large language model (LLM). LLMs are the text-generating part of Generative AI.

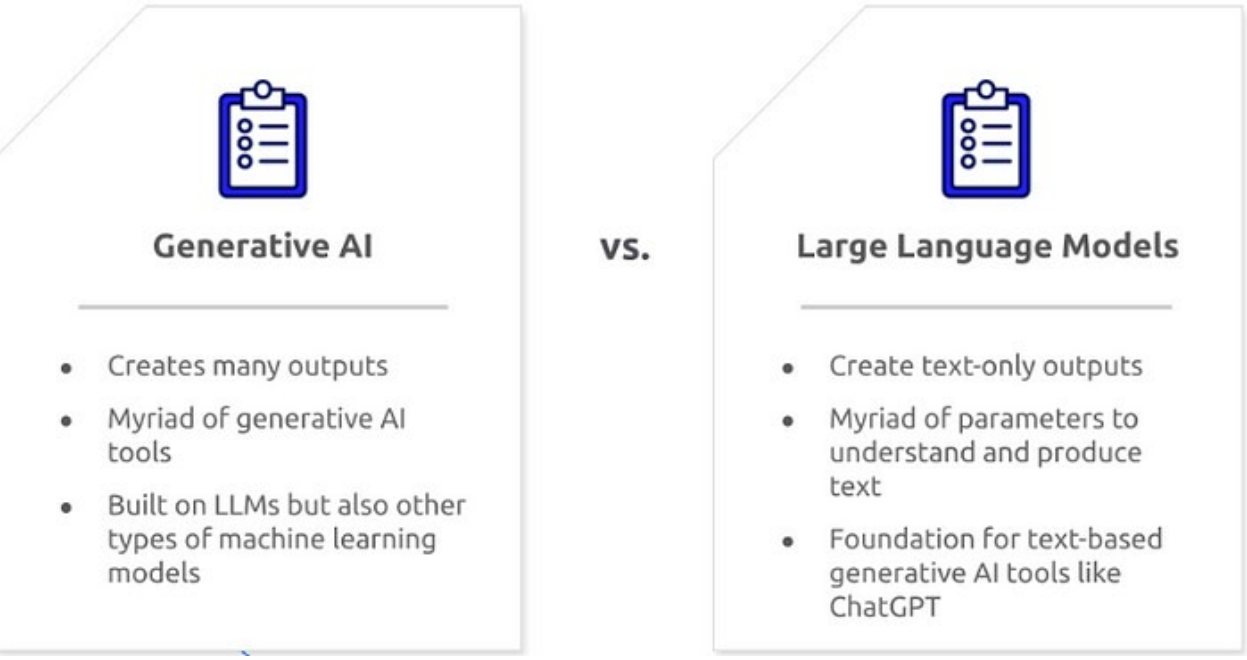
2. LLMs create text-only outputs.

LLMs can only create text outputs, and they used to only be able to accept text inputs, as well. When OpenAI first released ChatGPT in 2022, it was built on a text-only LLM, GPT-3. But now, with the development of "multimodal" LLMs, these LLMs can accept audio, imagery, etc. as inputs. OpenAI's next iteration, GPT-4, is an example of a multimodal LLM.

Both generative AI and LLMs will revolutionize industries, but they will do so in different ways. Generative AI could change the way we do 3D modeling, generate video output, or create voice assistants and other audio. LLMs will focus more on text-based content creation but still have other significant uses and may play a role in wider generative AI options like voice assistants.

3. LLMs are growing.

AI tools like ChatGPT and Google's Bard launched only within the last two years. Everest Group notes that one reason 2023 saw such exponential growth is the expansion of parameters in large language models, with GPT-4 having more than 175 billion parameters .



Source: https://medium.com/@prash.parihar/generative-ai-vs-large-language-models-llms-whats-the-difference-9a4fa6e2ab40

Large Language Models

By now, most of us are familiar with LLMs: neural network-based language models trained on vast quantities of data to mimic human behavior by performing various downstream tasks, like question answering, language translation, and content summarization. LLMs have disrupted the world with the introduction of tools like ChatGPT and GitHub Copilot. Their varieties include adaptable Foundation Models.

These large models achieve contextual understanding and remember things because memory units are incorporated in their architectures. They store and retrieve relevant information and can then produce coherent and contextually accurate responses.

Open Source vs. Closed Source LLMs

Open source LLMs differ from their closed counterparts regarding the source code (and sometimes other components, as well). With closed LLMs, the source code—which explains how the model is structured and how the training algorithms work—isn't published. A closed LLM cannot be fine-tuned because the system is entirely closed. Additionally, closed systems do not disclose the resources used in their creation, making it difficult to assess the quality and veracity of the inputs that will be used to create downstream outputs.

Open source LLMs help the industry at large: because so many people contribute, they can be developed faster than closed models. They can also be more effective for edge cases or specific applications (like local language support), can contain bespoke security controls, and can run on local models.

But closed models—often built by larger companies—have advantages, too. For one, they're embedded in systems with filters for biased information, inappropriate language, and other questionable content. They also frequently have security measures baked in. Plus, they don't need fine-tuning, a specialized skill set requiring dedicated people and teams.

Despite the popularity of ChatGPT and OpenAI, concerns like data privacy have led many organizations to be cautious about adopting them internally.

There are several open source commercially licensed LLM Foundation Models available. These include:

- **OpenLLaMA:** An open source reproduction of Meta's LLaMA model, developed by Berkeley AI Research, this project provides permissively licensed models with 3B, 7B, and 13B parameters, and is trained on one trillion tokens..
- **Falcon-Series**: Developed by Abu Dhabi's Technology Innovation Institute (TII), Falcon-Series consists of two models: Falcon-40B and Falcon-7B. The series has a unique training data pipeline that extracts content with deduplication and filtering from web data. The models also use multi-query attention, which improves the scalability of inference. Falcon can generate human-like text, translate languages, and answer questions.
- **MPT-Series:** A set of decoder-only large language models, MPT-Series models have been trained on one trillion tokens spanning code, natural language text, and scientific text. Developed by MosaicML, these models come in two specific versions: MPT-Instruct, designed to be task-oriented, and MPT-Chat, which provides a conversational experience. It's most suitable for virtual assistants, chatbots, and other interactive user engagement tools.
- **FastChat-T5:** A large transformer model with three billion parameters, FastChat-T5 is a chatbot model developed by the FastChat team through fine-tuning the Flan-T5-XL model. Trained on 70,000 user-shared conversations, it generates responses to user inputs autoregressively and is primarily for commercial applications. It's a strong fit for applications that need language understanding, like virtual assistants, customer support systems, and interactive platforms.

Utilizing open-source projects, you can readily obtain a pre-trained model with performance comparable to ChatGPT, at zero training cost for your organization. Furthermore, a private, self-hosted LLM can be fine-tuned with proprietary data, significantly enhancing performance on specific tasks. This fine-tuning will incur additional costs and require a specialized technical skillset, which will not come cheaply. Having your own LLM provides numerous advantages across various areas, which can be broadly categorized into data protection, performance, and strategic benefits.

With increasing regulations like the AI Act and user data protection measures, having an in-house LLM allows organizations to maintain control over sensitive data. Publicly available LLMs are likely to face more regulations, leading to reduced performance and increased costs due to the inclusion of disclaimers and safety measures. By training the LLM on trusted and verified sources, organizations can ensure that models used in sensitive areas like healthcare do not exhibit bias based on unreliable data.

LLM Content Generation

LLMs and Generative AI models can produce original, contextually relevant creative content including images, music, and text. For example, a Generative AI model trained on a dataset of paintings can be enhanced by an LLM that "understands" art history and can generate descriptions and analyses of artwork.

By drawing on both Generative AI and LLMs, content can be personalized for individuals. LLMs can make sense of user characteristics and generate personalized recommendations in response, while Generative AI can create customized content based on the preferences, including targeted search results.

LLMs can enhance the conversational abilities of bots and assistants by incorporating Generative AI techniques. LLMs provide context and memory capabilities, while Generative AI enables the production of engaging responses. This results in more natural, humanlike, interactive conversations.

Large language models can be combined with Generative AI models that work with other modalities, such as images or audio. This allows for generation of multimodal content, with the AI system being able to create text descriptions of images or create soundtracks for videos, among other multimodal capabilities. By combining language-understanding strengths with content generation, AI systems can create richer, more immersive content.

When combined with Generative AI, LLMs can be harnessed to create stories and narratives. Human writers can provide prompts and initial story elements, and the AI system can then generate subsequent content, all while maintaining coherence and staying in context.

LLMs can be utilized alongside Generative AI models to improve content translation and localization. A Large Language Model can decipher the nuances of language, while Generative AI can create accurate translations and localized versions of the content. This combination enables more-accurate, contextually appropriate translations in real time, enhancing global communication and content accessibility.

Both Large Language Models and Generative AI models can generate concise summaries of long-form content. Their strengths: LLMs can assess the context and key points, while Generative AI can develop condensed versions of the text that capture the essence of the original material. This ensures efficient information retrieval and lets people quickly grasp the main ideas laid out in lengthy documents.

Examples of LLMs include:

- **GPT-3 (Generative Pre-trained Transformer 3**): Developed by OpenAI, this is one of the most prominent LLMs, producing coherent, contextually appropriate text. It's already being widely used in applications including chatbots, content generation, and language translation.
- **GPT-4:** This successor to GPT-3 supplies advancements in contextual understanding and memory capabilities. As an evolving model, the goal is to further improve the quality of generated text and push the boundaries of language generation.
- **PaLM 2 (Pre-trained AutoRegressive Language Model 2):** Here's a non-GPT example of an LLM that's focused on language understanding and generation, offering enhanced performance in tasks such as language modeling, text completion, and document classification.

Issues With LLM Output Reliability ("Hallucinations")

Large Language Models (LLMs) are at the forefront of technological discussions, known for their proficiency in processing and generating text that resembles human communication. They are transforming our interactions with technology. However, these models are not without their flaws. One significant issue is their tendency to produce "hallucinations," which affect their reliability and trustworthiness. This is especially concerning when accuracy is paramount, such as in healthcare.

"Hallucinations" in the context of LLMs refer to the generation of content that is irrelevant, made-up, or inconsistent with the input data. This problem leads to incorrect information, challenging the trust placed in these models. Hallucinations are a critical obstacle in the development of LLMs, often arising from the training data's quality and the models' interpretative limits.

The causes of hallucinations in Large Language Models (LLMs) are multifaceted and stem from various aspects of their development and deployment. Some of the key causes of hallucinations in LLMs include training data issues. LLMs, such as GPT, Falcon, and LlaMa, undergo extensive unsupervised training with large and diverse datasets from multiple origins. Verifying this data's fairness, unbiasedness, and factual correctness is challenging. As these models learn to generate text, they may also pick up and replicate factual inaccuracies in the training data. This leads to scenarios where the models cannot distinguish between truth and fiction and may generate outputs that deviate from facts or logical reasoning. LLMs trained on internet-sourced datasets may include biased or incorrect information. This misinformation can propagate into the model's outputs, as the model doesn't distinguish between accurate and inaccurate data.

LLM "Hallucination" Categories

Hallucinations in Large Language Models (LLMs) are categorized into factuality and faithfulness hallucinations.

- **Factuality Hallucination:** This occurs when an LLM generates factually incorrect content. For instance, a model might claim that Charles Lindbergh was the first to walk on the moon, which is a factual error. This type of hallucination arises due to the model's limited contextual understanding and the inherent noise or errors in the training data, leading to responses that are not grounded in reality.

- **Faithfulness Hallucination:** These are instances where the model produces unfaithful content or is inconsistent with the provided source content. Faithfulness Hallucinations fall into three broad categories:

    O **Instruction Inconsistency:** The LLM ignores the specific instructions given by the user. For example, instead of translating a question into Spanish as instructed, the model provides the answer in English.

    O **Context Inconsistency:** The model output includes information not present in the provided context or contradicting it. An example is the LLM claiming the Nile originates from the Great Lakes region, as mentioned in the user's input, while omitting that the Nile is actually in Egypt and not in America.

    O **Logical Inconsistency:** The model's output contains a logical error despite starting correctly. For instance, the LLM performs an arithmetic operation incorrectly in a step-by-step math solution.

The scope of hallucinations in LLMs is broader than in task-specific models due to the diverse range of applications and the complex nature of the models. Intrinsic hallucinations often contradict the original text or external knowledge, while extrinsic hallucinations introduce new, unverifiable information. This phenomenon is observed across various generative tasks, from summarization to dialogue generation and question answering, each posing unique challenges in maintaining accuracy and consistency.

Hallucinations can also arise from model architecture flaws or suboptimal training objectives. For instance, an architecture flaw or a misaligned training objective can lead the model to produce outputs that do not align with the intended use or expected performance. This misalignment can result in the model generating content that is either nonsensical or factually incorrect.

The way prompts are engineered can also influence the occurrence of hallucinations. The LLM might generate an incorrect or unrelated answer if a prompt lacks adequate context or is ambiguously worded. Effective prompt engineering requires clarity and specificity to guide the model toward generating relevant and accurate responses.

LLMs may generate hallucinated content when faced with unclear or imprecise input. In the absence of explicit information, models can and do provide inappropriate outputs. There have been multiple reported instances, such as when an AI-powered chatbot encouraged a user to commit suicide, that have led directly to disastrous outcomes. Sometimes, LLMs are over-optimized for certain outcomes, such as longer outputs, which can lead to verbose and irrelevant responses. This over-optimization can cause models to stray from providing concise, accurate information to producing more content that may include hallucinations. AI hallucinations are a very real and serious problem, requiring deep thinking about the possible harms caused by AI and the mitigation strategies required for prevention.

Mitigating LLM Hallucinations

Mitigating hallucinations in LLMs involves a multifaceted approach, including using scoring systems where human annotators rate the level of hallucination, compare generated content

against baselines, and implement various remediation strategies. “Red Teaming”, where human evaluators rigorously test the model, is crucial in identifying and addressing hallucinations. Product-level features like user editability, structured input/output, and user feedback mechanisms also effectively reduce the risk of hallucinations.

Efforts to mitigate hallucinations are pivotal for maintaining the credibility and functionality of LLMs. Key methods for identifying and reducing these errors involve a combination of sophisticated metrics and critical human evaluations. Looking to the future, the development of LLMs is steering towards greater robustness and safety, with a strong emphasis on grounding responses in verified information. But we are not yet at a point where AI can be trusted to always provide accurate, credible information without intervention and oversight by human beings. For the foreseeable future, VBBBF would need a team of evaluators to constantly monitor any AI application for inappropriate or harmful responses so that they can be mitigated quickly, before they propagate further within the model itself.

Generative AI

Generative AI (“GenAI”) is a branch of AI that uses machine learning techniques to generate new content that mimics the data it was trained on. By ingesting vast amounts of training data, GenAI models can employ complex machine-learning algorithms in order to understand patterns and formulate output. The key idea is that GenAI models can learn patterns and relationships in a dataset well enough to create entirely new images, text, audio, and video that still adhere to the underlying patterns.

GenAI has become possible because of advancements in deep learning techniques. These new methods enable models to develop a rich understanding of large datasets and produce strikingly realistic synthetic content.

GenAI opens up endless possibilities for several industries. One example is the healthcare sector, which is data-rich; with clinical notes, diagnostic images, medical charts, and recordings. GenAI can take these unstructured data sets and create new organized datasets, then combine them with other large, structured datasets like insurance claims to create entirely new ways of accessing and viewing healthcare-related data.

Key examples of Generative AI models include:

- **DALL-E:** This platform developed by OpenAI, trained on a diverse range of images, can generate unique and detailed images based on textual descriptions. Its secret: understanding context and relationships between words.
- **Midjourney:** This generative AI platform focused on creative applications lets people create imaginative artistic images by leveraging deep-learning techniques.
- **Dream Studio:** This generative AI platform (which also offers an open-source free version), enables music creation. It employs machine-learning algorithms to analyze patterns in music data and generates novel compositions based on input and style preferences. This allows musicians to explore new and lateral ideas and enhance their creative processes.
- **Runway:** This platform provides a range of generative AI tools for creative professionals. It can come up with realistic images, manipulate photos, create 3D models, automate

filmmaking, and more. Artists incorporating generative AI in their workflows can experiment with fine-tuning a variety of techniques.

Practical Uses of Generative AI

The field of Generative AI will progress rapidly in both scientific discovery and technology commercialization, but use cases are emerging quickly in creative content, content improvement, synthetic data, generative engineering, and generative design.

In-use, high-level practical applications today include the following.

- **Written content augmentation and creation:** Producing a "draft" output of text in a desired style and length
- **Question answering and discovery:** Enabling users to locate answers to input, based on data and prompt information
- **Tone:** Text manipulation, to soften language or professionalize text
- **Summarization:** Offering shortened versions of conversations, articles, emails and webpages
- **Simplification**: Breaking down lengthy content and creating outlines then extracting key insights
- **Classification of content for specific use cases**: Sorting by sentiment, topic, etc.

Risks of Generative AI

The risks associated with Generative AI are significant and rapidly evolving. A wide array of threat actors are already using the technology to create "deep fakes" or copies of products and people saying and doing things that do not exist in reality, and to generate artifacts to support increasingly complex scams.

ChatGPT and other tools like it are trained on large amounts of publicly available data. They are not designed to be compliant with General Data Protection Regulation (GDPR) and other copyright laws, so it's imperative to pay close attention to your organization's uses of the platforms.

Oversight risks to monitor include:

- **Lack of transparency**. Generative AI and ChatGPT models are unpredictable, and not even the companies behind them always understand everything about how they work.
- **Accuracy.** Generative AI systems sometimes produce inaccurate and fabricated answers. Assess all outputs for accuracy, appropriateness and actual usefulness before relying on or publicly distributing information.
- **Bias.** You need policies or controls in place to detect biased outputs and deal with them in a manner consistent with company policy and any relevant legal requirements.
- **Intellectual property (IP) and copyright**. There are currently no verifiable data governance and protection assurances regarding confidential information. Users should assume that any data or queries they enter into the ChatGPT and its competitors will become public information and appropriate controls should be put in placeto avoid inadvertently exposing protected information.

- **Cybersecurity and fraud**. Organizations must prepare for malicious actors' use of Generative AI systems for cyber and fraud attacks, such as those that use deep fakes for social engineering and ensure mitigating controls are put in place. Confer with your insurance provider to verify the degree to which your existing policy covers AI-related privacy and security breaches.
- **Sustainability**. Generative AI uses significant amounts of electricity. Choose vendors that reduce power consumption and leverage high-quality renewable energy to mitigate the impact on your sustainability goals.

Best practices For Implementing Generative AI

Technologies that provide AI trust and transparency will become an important complement to Generative AI solutions. Organizational leaders should follow this guidance for ethical use of LLMs and other generative AI models:

- **Start inside:** Before using generative AI to create public-facing content, test extensively with internal stakeholders and specific use cases. You don't want hallucinations to harm your credibility.
- **Prize transparency:** Be forthcoming with people with users about the fact that they are interacting with a machine by clearly labeling any conversation multiple times throughout.
- **Due diligence:** Set up processes and guardrails to track biases and other issues of trustworthiness. Do so by validating results and continually testing for the model going off course.
- **Address privacy and security concerns:** Ensure that sensitive data is neither input nor derived. Confirm with the model provider that this data won't be used for machine learning beyond your organization.
- **Take it slow:** Keep functionality in beta for an extended period of time. This helps temper expectations for perfect results.

Uses of AI for the Clearinghouse

The potential uses for AI in the development of the VBBBF Clearinghouse include:

- Aggregation of multiple sources of substance abuse and mental healthcare information using an LLM to create a single resource that can be searched using natural language queries.
- Use of GenAI to automate multilingual translation of queries and responses.
- Intelligent processing of aggregated data to produce both streamlined summaries and in-depth analysis using LLM and GenAI capabilities.
- Responses tailored to the needs of diverse audiences that includes individuals, institutions, and policy makers.
- Automated continuous improvement of the LLM model through on-going fine-tuning based on observations of real-time user activities, as well as the continual insertion of fresh data.
- Development of comprehensive 360° analytics that will identify trends and deliver real-time insights into Clearinghouse activities.

Challenges of AI Implementation for the Clearinghouse

- Developing strategies for competing with multiple, well-funded organizations that have already launched AI -powered applications in the public domain including search engines, healthcare systems/agencies, and privately funded solutions such as Woebot.
- Securing adequate funding for the expected initial and on-going costs to support an AI initiative. Developing AI is expensive at all stages and requires on-going investment over the entire lifecycle. Because AI is a rapidly shifting landscape, it is likely that the Clearinghouse will require more than one round of major investment over the next 3-5 years in order to keep pace with the evolving technology. Before taking the step to use AI as the basis for the Clearinghouse, VBBBF must clarify whether its fundraising will keep pace with the expected costs for an undertaking of the proposed scope and capabilities.
- The decisions required to select an AI model for the Clearinghouse initiative will require in-depth evaluation of Open Source vs Proprietary models as well as the "Build vs Buy" decision. These initial choices create a level of 'lock-in' that will be difficult to change at a later stage of development, and therefore should not be rushed. These decisions will require what will likely be an extended period of discussion, meetings, and product demonstrations in order to select a model with reasonable confidence that it will meet the needs of the Clearinghouse. This may be challenging against the backdrop of the suggested timeline of 3-6 months to launch as specified in the RFP.

Factors for Further Consideration

Regardless of the data platform selected, the following factors must be addressed by VBBBF in order for PWC to make specific recommendations regarding technology choices and to provide initial cost estimates for the implementation of an customized AI component in the Clearinghouse:

- Selection of sources that will comprise the initial dataset. This is discussed in detail in the Dataset section.
- Development of specific 'use cases' in order to understand the criteria and performance specifications of the Clearinghouse.
- Articulation of desired outcomes for each cohort of users to guide the programming or training process.
- Benchmark measures of success for the performance of the Clearinghouse model. The benchmarks and objectives should be clearly articulated from the beginning of the project, in order to prioritize development that meets and exceeds expectations.

Privacy Concerns

LLMs, with their training on extensive datasets, may raise concerns regarding privacy violations as they may contain personal information about individuals. Research findings from Blackberry indicate that an overwhelming 74% of IT decision-makers not only recognize the potential cybersecurity threat posed by GPT-4 but also express significant concern about its implications. Furthermore, according to recent ChatGPT and Generative AI statistics, approximately 11% of the information shared with ChatGPT consists of sensitive data. Instances have occurred where LLMs inadvertently leaked specific personal details, such as social security numbers, home addresses,

cell phone numbers, and medical information. These privacy breaches underscore the importance of implementing robust measures to safeguard personal data when deploying LLMs, considering the serious implications that can arise for individuals.

HIPPA Considerations

One of the biggest concerns surrounding the use of AI in healthcare is whether or not the tool complies with the Health Insurance Portability and Accountability Act ("HIPAA"). HIPAA requires certain "Covered Entities" to protect sensitive patient data. Any health insurance company, clearinghouse, or provider that handles Protected Health Information is required to comply with HIPAA regulations.

In the past, OpenAI warned against uploading confidential information to ChatGPT in its Terms of Service and on its FAQ page. However, on March 1, 2023, OpenAI updated its policies by noting that certain data submitted by users will not be used to train or improve its models unless individual users explicitly opt-in to share that data. OpenAI has also announced that it will be able, upon request, to sign BAAs in support of an applicable customer's compliance with HIPAA. If a BAA is not in place between OpenAI and the Covered Entity, any Protected Health Information uploaded to ChatGPT will violate HIPAA. Whether the Clearinghouse would be subjected to HIPAA regulations is not clear at this time, but it would be wise to err on the side of safety and adhere to HIPAA requirements during development and deployment.

Preliminary Cost Estimate for AI Implementation

Professor Musmar estimates that the cost to train a customized LLM 'proof of concept' model for VBBBF, based on an open-source LLM such as OpenLLaMA would be in the range of $125,000 to $150,000. This cost does not include the provision by VBBBF of the dataset that will be used to train the LLM. He estimates that additional costs to take the proof of concept to a fully-functioning model ready for public deployment are estimated in the $30,000 to $50,000 range. Please note that my extensive experience working with programmers tells me that this is likely a low-ball estimate. Programmers are always more optimistic than they should be about time and cost estimates. In the past, my practice has always been to double their number to arrive at a more realistic estimate of actual costs.

Additional costs for acquiring the data, whether through web scraping, human collection, or some combination of both would need to be estimated and added to the cost of development. The costs of acquiring the data will depend on both the source and the amount of data to be acquired. These specifications need to be provided by VBBBF in order for PWC to provide a cost estimate. Other costs during the AI development phase would include the human costs of "Red Teaming" the outputs of the model for accuracy, as well as the provision of on-going technical support from Professor Musmar or another AI professional during the implementation phase.

Long-terms costs include the costs of hosting, augmenting, and fine-tuning the data. Hosting costs can easily run into the tens of thousands of dollars per month and can be especially difficult to estimate using costing information provided by services like Amazon, who tend to be obtuse and less-than-transparent about total costs of ownership. Contracting with an experienced data scientist to oversee and fine-tune the dataset on a regular basis will add another layer of costs to the monthly outlay.

Summary of AI Considerations

This section has introduced a very high level discussion of the many factors that must be taken into account regarding the use of AI in the development of the Clearinghouse. Factors for immediate consideration include:

- Does VBBBF have the financial foundation and resources to sustain a long-term investment in AI technology?
- If AI technology is chosen, is VBBBF prepared to make choices regarding platform selection in an environment that contains ONLY imperfect information about the available options?
- How will VBBBF mitigate the technical, ethical, and legal risks associated with the use of AI?
- How will the Clearinghouse navigate the intensely competitive environment that includes AI-powered search engines and AI Therapy Agents to trap their target cohort's attention?

To summarize, AI is a complex and rapidly changing technology that presents enormous opportunities to provide comprehensive help, support, and services in the areas of mental healthcare and substance abuse. At the same time, it also presents a high level of risk to do harm if it is not implemented and maintained with the appropriate oversight, intervention, and safeguards. Building a custom AI solution is an enormous undertaking that will require a substantial commitment of time and resources before any benefit is realized. Armed with a more nuanced understanding of what is required to develop a custom AI-powered Clearinghouse in-house, I would like to turn the discussion in the next section to the use of an existing, off-the-shelf approach that could likely be implemented within the stated time frame of 3-6 months, would provide a highly engaging user experience, and will have a fairly stable and reliable estimate of costs both to develop and maintain.

AI Therapy Agents

Overview of AI and Digital Well-Being

Generative AI and digital well-being have a complicated relationship. AI has the potential to revolutionize the delivery of healthcare services and interventions, but it also poses a significant threat to the ability of individuals to disconnect from digital technologies. It is important for individuals, organizations, and developers to consider the potential effects of Generative AI on digital well-being and to take steps to ensure that it is used in a way that promotes positive outcomes.

The logging-off movement, also known as the digital detox movement, is a growing trend of people disconnecting from technology from time-to-time to improve mental and physical health, productivity, creativity, and relationships. The movement is driven by concerns about the negative effects of technology - especially the effects of social networks - on well-being, with a cascade of reports of worsening of mental health challenges such as addiction, anxiety, and depression, driven by near-constant engagement with social networking platforms.

Generative AI could also impact digital well-being negatively, through the creation of highly personalized digital experiences that are, themselves, addictive. With the ability to generate content tailored to a person's individual preferences and interests, Generative AI could make it even more difficult for individuals to disconnect from technology, as they may get hooked on cleverly personalized content, leading to an endless cycle of reliance on screens. We have already seen how this plays out with ad targeting and the attention-based economy of social media platform, which induces users to continually engage, even when it is not in the best interests of their emotional well-being.

Generative AI also has the potential to help improve mental health outcomes, precisely because of many of the same factors that can negatively influence well-being. This is the double-edged sword that hangs over all technology – for every negative outcome, there is almost always a corresponding and equally compelling positive outcome. There is current statistical evidence that content tailored to the individual and designed to be engaging and encouraging of repeated use can promote positive behavioral changes can reduce harmful behaviors and improve overall well-being. There is a growing body of evidence that indicates that AI Therapy Agents have a very strong potential to create transformational change in the realm of mental healthcare, providing services to a huge number of people who currently have no access to services at all. The truth is that Generative AI is here, and here to stay. It is incumbent on the good and caring people of this world to harness its potential for the betterment of humankind, in spite of its corresponding potential to do harm.

Transformational Effects of AI in Mental Healthcare

The advent of AI has ushered in a new era of technological advancements, profoundly impacting the provision of mental healthcare. AI Therapy Agents – sophisticated programs capable of performing tasks autonomously – are increasingly being recognized for their potential to revolutionize healthcare in general and mental healthcare in particular. AI Therapy Agents are advanced computational entities capable of performing tasks autonomously, equipped with learning and decision-making abilities, often but not always, based on LLMs and Generative AI. In mental health, these agents range from chatbots for initial patient interaction to more complex systems for therapy assistance and data analysis. AI Therapy Agents are poised to fundamentally reshape the landscape of mental health care. They offer the potential for more accessible and effective mental health services, a significant step forward in addressing the global mental health crisis, most especially for those in currently underserved populations.

Evolving digital technology and AI are transforming the field of mental health in multiple areas, including:

- **Prediction and detection:** AI, particularly machine learning, is increasingly used for prediction, detection, and treatment in mental health care.
- **Natural language processing:** Analyzing clinical texts and social media content provides a means to spot mental health states and supports the development of conversational agents for therapeutic intervention.
- **Chatbots and virtual agents:** These offer accessible therapy options for various mental health conditions, with approaches such as Cognitive-Behavioral Therapy and other therapeutic techniques.

- **Real-Time interventions:** Mobile devices can support real-time psychological interventions and behavioral prompts. They frequently use user feedback and behavior to inform their highly personalized therapy recommendations. Many AI Therapy Agents have "SOS" features that detect suicidal ideation and redirect users to immediate assistance to avert a crisis. They also have daily 'check-in' features that can help users remember to take medications, attend scheduled appointments, or simply remind to remind them that they are not alone and that the Therapy Agent is always available to provide support.

In recent years, thousands of apps have been released to help people manage their mental health. These apps are designed for various issues, including stress, anxiety, sleep, and daily functioning. In addition, there is a whole category of apps specifically to help autistic people communicate and address sensory issues.

Although an increase in telehealth has significantly improved access to care for many Americans, a shortage of qualified professionals makes it difficult for people to access the services and care they need. Additionally, many Americans find themselves living in a "healthcare desert" with no access of any kind to healthcare services within a reasonable distance of their home. Many health insurance plans do not cover mental healthcare or substance abuse services or treatments, making the cost of accessing these types of healthcare prohibitive. While an AI Therapy Agent is not a substitute for human intervention, they can supplement therapy and offer support that would not otherwise be available, especially in underserved communities. For many of the individuals targeted by the VBBBF website, they may be the ONLY intervention available, and for this reason, the use of an AI Therapy Agent to provide Clearinghouse services should be seriously contemplated.

The Rise of Digital Mental Healthcare 'Therapy Agents'

Although we are in the early stages of a revolution in the provision of mental healthcare using Therapy Agents, there are many advantages to this approach that all but guarantee rapid, widespread adoption of these technologies in the near-term including:

- 24/7 availability regardless of geographic location
- Cost effectiveness
- Stigma reduction
- Efficient diagnosis and assessment
- Real time monitoring and outreach
- The potential to transform mental health and substance abuse research through the collection and analysis of the massive data trove generated by interactions with the Therapy Agent.

There are also disadvantages to AI Therapy Agents, including:

- "Dehumanization" of mental health interventions
- Biases in the underlying model, especially biases around cultural differences in symptom expression, that impact diagnosis and treatment
- Data collection, privacy, and security concerns

- Inability to diagnose and treat multiple, complex issues that may include both mental health and substance abuse challenges.
- Potential for harmful outcomes due to hallucinations and problems with underlying response-generating algorithms
- Concerns about long-term effectiveness of non-human interventions

FDA Regulation of Therapy Agents

There are a number of Therapy Agents that have come to market recently, including simple apps like Calm, Headspace, and Smiling Mind, as well more complex and heavily-regulated apps such as EndeavorRX, ReSET, Woebot . There is also an app in the 'middle ground' of the regulatory environment called Wysa.

Applications such as Calm are not considered medical interventions and can be thought of in the same way that dietary supplements such as Turmeric are not considered "food". The applications are considered 'wellness apps' and focus on meditation, mindfulness, the use of ambient sounds and other non-medical modalities to induce a sense of well-being in the user.

EndeavorRX, ReSET, and Woebot, on the other hand, are considered medical interventions to manage or "treat" health conditions and are heavily regulated by the FDA, with stringent and complicated rules as to what they can and cannot do as a Software as a Medical Device ("SaMD") product.

Wysa is in the middle ground, considered a 'Breakthrough Device" by the FDA, which is more regulated than "wellness apps" but less regulated than SaMD applications.

Woebot, which is a heavily-regulated product, uses AI to guide the relational agent (also called "Woebot") to understand what users type in and to respond in an appropriate manner. It does this using natural language processing--which is a rules-based method that is considered 'non-generative AI". Rather than allow the model to determine appropriate responses to user inputs, Woebot has a clinical content team who programs various types of text that a user could type in and instructions for generating a response using a set of human-defined "rules" (e.g., if the user types "upset" and "partner", Woebot may respond "ok, it sounds like you are upset with your partner and having a relationship issue, is that correct?")

Unregulated apps such as Calm do not use AI or provide "therapy" as it is traditionally understood, which is why they are not considered "healthcare interventions". These apps provide guidance to the use of benign behavior modifications such as daily meditation, scheduled technology "breaks", or "sound therapy" to reduce symptoms of anxiety and depression.

Users who turn for help to widely available Generative AI tools like ChatGPT that do not use rules-based processing are in danger of receiving inaccurate, unverified, and potentially harmful or even deadly information. The lack of a rigorously enforced response paradigm of generalized AI to users experiencing mental health distress can pose risks when dealing with people who have mental health issues, particularly for those at high risk for self-harm.

There is deep concern that the current public-facing LLM/GenAI models do not have the safeguards that are needed when dealing with a population of users experiencing mental distress. Both

Woebot and Wysa have safety net plans ("SOS features") in place that are triggered when the user enters certain textual information. Publicly facing, general-use chatbots do not have robust safety features that are trained to detect user distress and in fact, there have been a number of stories in the media about general-use chatbots encouraging harmful behavior rather than mitigating it. This speaks to the need for a widely-available Therapy Agent with built-in safeguards.

The Use of A Therapy Agent for the Clearinghouse

In the course of researching possible solutions for the development of the Clearinghouse, PWC spoke at length with experts in the areas Information Tailoring, Machine Learning and AI Development, and with representatives from both Woebot and Wysa, in order to provide VBBBF with advice and guidance and the best course of action to take regarding the creation of the Clearinghouse website. Woebot is not available for public-facing deployment, due to the heavily regulated status of their application. Agents such as Calm do not provide the types or level of intervention and support services that VBBBF seeks to provide through the Clearinghouse. After speaking extensively with Rhea Yadav of Wysa, PWC is of the strong opinion that VBBBF should consider using Wysa as the foundation technology for the Clearinghouse, as it provides a nearly ideal middle ground for the use of an AI Therapy Agent for mental health support and intervention, is robust, scalable, and proven to work well across the needs of a diverse user population, eliminates the need to consider any custom AI development in-house, and can accept the types of information customizations that would allow it to meet the particular information needs of the Clearinghouse as defined in the RFP.

Wysa AI Coach

Wysa AI Coach is an artificial intelligence-based 'emotionally intelligent' service which responds to the emotions expressed by the user and uses evidence-based cognitive-behavioral techniques (CBT), DBT, meditation, breathing, yoga, motivational interviewing and micro-actions to help users build mental resilience skills and feel better.

Please refer to this listing of peer-reviewed studies for Wysa for more information about Wysa's research and development: https://blogs.wysa.io/wp-content/uploads/2023/11/Wysa-Bibliography-1.pdf

Is Wysa a Human or a Robot?

The answer is both! Wysa has an artificial intelligence (AI) coach that learns and grows wiser with the user. Wysa's human Emotional Well-being professionals are mental health professionals trained to listen, support and encourage users.

Together with the Wysa AI coach, Emotional Well-being Professionals provide support in achieving defined goals. The AI Coach is restricted in the means of response and the intended use of Wysa is to provide evidence-based tools and techniques to manage emotions and encourage mental well-being in a self-help context. It is not intended to provide diagnosis, treatment, or cure of a condition or disease or disorders, which is why it is not considered a SaMD. The Wysa AI Coach cannot and will not offer advice on unrecognizable issues.

Wysa is the result of a year-long co-design effort between a 15-person team of psychologists, designers, developers, and over 500,000 users to understand how AI chat can help users learn skills to build emotional resilience.

Wysa is focused on privacy. All Wysa chats are private and anonymous. There are no advertisements or tracking mechanisms on the Wysa platform and users are encouraged to uninstall other mental health applications that may be tracking their activities.

Wysa understands that most people don't want their problems 'fixed'. Mostly, they just want to talk through them, with someone who doesn't judge. This is borne out by a growing body of research that indicates fear of judgement is a key driver in the reluctance to seek mental healthcare and substance abuse support. Wysa is available 24/7, meeting users in their time of crisis, anywhere, anytime, and in several languages, including English, Spanish, and Hindi.

Wysa is not designed to assist with crises such as physical abuse, severe mental health conditions that may cause feelings of suicide, harm to self and any other medical emergencies. In any of these cases and otherwise, Wysa cannot and will not offer medical or clinical advice. It can only suggest that users seek advanced and professional medical help though it's automated "SOS" features, which have been activated by about 5% of Wysa's users. When activated, the SOS feature can take a variety of actions, including activating a personal safety plan, directing users to crisis helplines, providing grounding exercise, or directing the user to custom resources, which is relevant to the VSS initiative underway as part of the VBBBF.

Integrating Wysa Into the Clearinghouse

PWC spoke at a high level with Rhea Yadav of Wysa about the possibilities for integrating Wysa into the Clearinghouse website. Of critical interest to PWC were the following features:

- Wysa can be deployed as a stand-alone app or fully integrated into a WordPress-based website of the type that PWC recommends to provide the base functionality for the Clearinghouse.
- Wysa's AI can be trained on any additional data and resources identified by VBBBF for inclusion in their dataset, without the need to manage and direct the technical aspects of creating an AI application and training it in-house.
- Wysa is designed to tailor information to the user, beginning with the initial interaction with the agent.
- Wysa is fully ADA-Compliant, requiring no further modifications to ensure equitable access to users across a range of assistive technologies.
- A third-party solution such as Wysa eliminates many of potential pain points of a custom solution discussed in this proposal, including the need to retain technical professionals to build, maintain and improve the AI system over time, the need to invest in AI-related infrastructure such as hosting and backup, and the provision of AI-expert technical support that is a phone call away to resolve issues and problems.
- Wysa gives the VBBBF an incredible competitive advantage as a 'first-to-market' application based on modern AI technologies that will be hard for other providers currently in the market (such as the example websites provided in this response) to

emulate as they simply do not have the technology stacks in place to support a similar offering.

The integration of the Wysa Therapy Agent into the Clearinghouse, while of course, not without risk, is a path to realizing the majority of the objectives as stated in the RFP within a very similar cost window to building a custom AI application. Wysa's pricing, which was provided as a high-level estimate in the range of $250,000 to $500,000, with the caveat of 'flexibility', is very similar to the high-level costs PWC has estimated to develop an in-house solution, but with much a much faster deployment window. It is the recommendation of PWC that, armed with the understanding of the technology options for the Clearinghouse as presented in this RFP, VBBBF consider having a first conversation with Wysa about the possibilities of a partnership sooner rather than later. Understanding what is possible through Wysa will help VBBBF develop a standard against which any custom development activities can be measured.

The Role of Plato Wynne Consulting

Melissa Morisco Raulston, the principal of Plato Wynne Consulting ("PWC") has been building, hosting, and maintaining websites and related technology, including server and database management, since 1998. Her extensive technology background in commercial, non-profit, and academic settings provides a baseline for understanding the evolution of web-based technologies over the past 25 years – both where it has been and where it is going. While not a subject-matter expert in many of the specific emerging technologies discussed in this proposal, such as vector databases, LLMs, and GenAI, Ms. Raulston's ability to survive and thrive in a rapidly changing technology environment for almost 25 years is a testament to her ability to rapidly understand and assimilate new technologies.

The services provided by Ms. Raulston will be as follows:

- The development, hosting, and management of the basic website infrastructure
- Consulting services to provide guidance and objective insight into proposals provided by third-party consultants such as Professor Musmar and Dr Muhammed, as well as the consideration of commercial services providing similar capabilities such as Wysa.
- Oversight and management of all technical aspects of the project, from refinement of the project specification to the launch of the Clearinghouse, to include continuing oversight and management services post-launch.

The Clearinghouse Website

No matter which technologies are selected for the provision of the core Clearinghouse functionality, a basic website framework will be required to serve as the foundation upon which all other services are provided. The basic functions of this website may include:

- Providing information about the Foundation including its creation, mission/vision, staffing, and fund-raising activities.
- Descriptive information tailored to each of the target cohorts, describing the resources available through the Clearinghouse and providing help accessing those resources.

- Gamified elements such as quizzes, surveys, challenges and other interactive elements that engage and encourage visitors to take positive actions related to their well-being
- Presentation of audio and video multimedia content
- Live chat capabilities
- Forum capabilities
- Advertisement of events related to VBBBF media campaigns and other 'awareness-raising' activities.
- User management services including 'membership' style logins that form the basis of any gamification and tailoring services.
- Other features and services that have not yet been defined

The WordPress Platform

The website would be built using the WordPress platform, which provides the flexibility, portability, and maturity required for the VBBBF website:

- **Flexibility:** WordPress is an open-source CMS developed in the late 1990s. It has evolved to become the standard CMS technology in use today and a majority of the websites you have likely visited in the past, such as the New York Times, CNN Pressroom, Time Magazine,and NASA, are all built on the WordPress platform, due in great part to its superior flexibility.

  The basic WordPress install provides the content management tools that are a necessary component of every website. In order to customize a website to an organizations needs, components known as 'plug-ins' are downloaded into the base WordPress installation to create new functionality. There is a large, well-supported, and mature ecosystem of WordPress plugins that can provide almost any functionality you can imagine from 'forms builders' to gamification elements. Several examples of plugins that extend WordPress capabilities include:

  - **GravityForms:** In addition to providing functionality for basic forms such as Contact forms, GravityForms can be used to build quizzes, polls, and surveys and other 'forms-based' functionality. GravityForms integrates with a range of services from mailing list servers such as MailChimp to payment processors such as Stripe. PWC has used GravityForms for over 10 years and is confident in its performance and reliability. GravityForms popularity has led to an entire ecosystem of related plugins that extend the functionality even further – for example, PWC built an entire suite of COVID screening services for use during the pandemic in NYC that met all requirements at the federal, state, and local levels including privacy and confidentiality, using only the tools in the GravityForms ecosystem. The system was built in less than a month.
  - **BuddyPress:** BuddyPress powers the creation of an online community and includes features such as user profiles, groups, forums, activity streams, and notifications. BuddyPress is a mature and well-supported plugin that is supported and endorsed by the creator of WordPress, Matt Mullenweg. PWC has used BuddyPress in the past and is familiar and comfortable with the setup and configuration of this plugin.

    - **GamiPress:** This plugin allows the creation and extension of gamification elements on the website and can be integrated with BuddyBoss to create a wrap-around experience for users. PWC has not used this plugin previously, however, it is well-supported and documented and well within our capabilities to deploy.
- **Portability:** This is one of the most important website features that most people don't think to ask about – can we take it with us when we leave? With WordPress, the answer is always YES. Unlike proprietary systems like Wix or GoDaddy Website Builder, WordPress websites can be moved, intact and fully functioning - to any web server using the base technologies of PHP and MySQL required to run WordPress. There are literally hundreds of thousands of hosting options who provide these capabilities, though as with all things, some are better than others. To move a WordPress website, simply select a new host, then transfer the downloaded database and associated website assets to them. Often, the new host will move the WordPress website for you, free of charge, and of course, there are plugins designed to simplify the task for novice users of moving a WordPress installation. In short, WordPress is not JUST portable, it is fast and easy to port, meaning you are never locked in to a single host or server – not even PWC.

- **Maturity:** WordPress differs from many of the other technologies we have discussed in this proposal in that it is mature, stable, and supported by a worldwide consortium of open source developers. In 2009, Ms. Raulston built a new website for the FSU College of Communication and Information on the WordPress platform, becoming the first unit on campus to deploy this technology. Although there was much gnashing of teeth amongst faculty, staff, and administration regarding Ms. Raulston's deployment of what was then a new and emerging technology, she has long-since been vindicated. In 2024, about 43% of ALL websites on the internet use WordPress, which also includes the entirety of the FSU website ecosystem. WordPress and the WordPress ecosystem evolve alongside changing technology, continually introducing new capabilities to meet the needs of website developers. For example, there are already plugins available that use AI to assist with content creation and curation and it is expected that at some point in the not-to-distant future, native AI capabilities will be included in the base WordPress installation.

Service Provision

PWC can provide basic design services to include site layout/theme customization, photo manipulation, content placement and management, organic SEO optimization services, and ADA compliance. PWC or VBBBF would need to contract for services that include sophisticated branding elements such as logo design, graphical element design, illustrations, infographics, and other artistic elements outside the capabilities of PWC.

PWC hosts its websites on the Kinta platform, which is designed specifically for the needs of WordPress websites and developers. The capabilities provided by Kinsta include fast caching and CDNs (Content Delivery Networks) for optimal page speed performance, automated backups, staging infrastructure that allows changes to be made to a website, then tested, before being pushed into the 'live' environment, application performance monitoring, and many other optimization features that ensure the best performance possible in an 'always up' environment.

Website analytics are generally provided by Google using their analytics platform. There is, of course, a plugin for this as well, allowing WordPress administrators to access analytic information directly on the website without having to go to Google.

PWC provides on-going website management services that include WordPress and Plugin upgrades, the addition of new features and services and technical support that includes user training/support, troubleshooting, and issue resolution for related products such as email. PWC can provide references on request for clients who have been part of the PWC family for 12-15 years and can attest to the level of quality and attention we provide to our clients before, during, and after their website launch.

Website Costs and Time-to-Launch

Because no specific information has been provided by VBBBF regarding these basic website functions and features, this section does not provide a comprehensive set of all features that can be provided by the base website. Time-to-launch depends on a number of factors outside the control of PWC including:

- Provision of all graphical elements including logo, imagery, videos, etc
- A 'style guide' indicating preferred fonts and colors
- Examples of websites that are similar in look, feel, and layout to the proposed website to be used as a starting point for developing 'wire frames'
- A comprehensive list of desired features at launch
- All content to be placed for the initial launch of the website

PWC has developed hundreds of websites on the WordPress platform since 2009. The basic WordPress installation is free, though many of the plugins are not. Costs for plugins mostly range from $20 to under $500 each. The total costs that are incurred for WordPress and any additional plugins will be in the range of several hundred dollars to several thousand dollars, depending on the licensing model selected for purchased plugins. The majority of the costs incurred to build the website will be the labor required to create the design, transform it to a WordPress theme, add and configure the necessary plugins, place content, and ensure ADA compliance. Web sites of medium complexity generally take 4-8 weeks to launch from initial conception, assuming all of the above factors are in place. Total costs for initial launch can be expected to run in the range of $10,000 to $20,000. Monthly hosting costs for the base website will be $50-$100 per month. Most PWC clients pay an additional monthly retainer in the range of $100 to $1000 per month for ongoing development and maintenance of their website, though this costs are difficult to estimate without an understanding of whether PWC will be performing the majority of this work or training others within the VBBBF organization to maintain and update content.

Consultation, Management, and Oversight of the VBBBF Clearinghouse

For a project as large and complex as the proposed Clearinghouse, it is imperative that a Project Manager with a complete understanding of the goals and objectives of the Clearinghouse oversee the entire lifecycle of development. PWC proposes to provide consultation, oversight, and management services for VBBBF on an ongoing basis to assist in the selection of technologies, oversee their implementation, and ensure that all of the parts form a seamless whole. Additionally, the development of the project must be clearly documented such that the VBBBF has a complete

set of information regarding the development and operability of the project. Documentation must cover the decision-making process that led to each technology choice, technical information about the functionality of each portion of the technology, and a guide to the maintenance required to sustain functionality over time. Failure to choose a single Project Manager who is invested with at least some decision making authority to oversee the entire project will result in a fragmented and poorly executed final product that may not function as expected – or at all, will not meet or exceed expectations as set forth in the RPF, and will produce time-and-cost overruns that may doom the entire project.

PWC will provide all necessary consultation, management, and oversight services at the hourly rate of $100. As the project progresses, deliverables such as technical documentation, development schedules, user manuals, etc will be specified for inclusion in the final project deliverables. It is highly recommended that a project management portal such as Slack be used to manage and centralize all project communications and documents, Pricing for a project management portal can be expected to cost in the range of $15-$30 per month, per user and can be expected to rise over time, as consultants, VBBBF team members, marketing personnel and so on are added to the portal to maintain seamless communication across all domains of the project.

Conclusion

PWC has thought deeply about the proposed Clearinghouse and the multiple technical paths to achieving the objectives as stated in the RFP. Although a fully customized solution for the Clearinghouse seems attractive at first glance, it is my experience that fully customized solutions often double or even triple in time-and-cost estimates over what is presented by the consultants whose job will be to produce them. Scope -creep, consultant over-optimism, and the changing needs in light of a more detailed specification all lead to substantial increases in both time and cost, creating frustration as money continues to be spent without any tangible, immediate benefit to your organization.

Integrating multiple technologies in a rapidly evolving ecosystem such as the one we find ourselves in today is challenging and expensive, with no guarantee that the investment you make today will still provide value tomorrow. Shifting the burden of technology integration to a third-party provider such as Wysa is often the smartest move, as VBBBF is not precluded from developing a custom app, should it desire to do so, at a later date, when many of today's urgent questions regarding the direction of website applications is more settled.

PWC is available for review and discussion of this proposal at your convenience, as many of the concepts presented may require further clarification during the decision-making process. We thank you for this opportunity to respond to the VBBBF RFP and look forward to our continued collaboration.

PWC
PROBLEM. SOLVED.

Resumes and CVs

Melissa Morisco Raulston

melissa@platowynne.com
(850) 491-6702

Chief Problem Solver, Plato Wynne Consulting

WordPress developer since 2012, serving clients in a broad range of both for-profit and non-profit organizations.

- WordPress website development and hosting. Custom theme creation, simple plugin and PHP scripting. Deeply experienced in Genesis framework.
- Email management including Exchange administration.
- Content creation for organic SEO optimization across multiple industries
- Print collateral creation using Photoshop, InDesign, Adobe Suite
- Microsoft-Certified Word expert, providing range of document and training services
- https://platowynne.com/wordpress-portfolio/
- https://Word.Expert

Chief Financial Officer, Truline Construction Services, Inc

Back office administration from 2016 – 2013 for Manhattan-based General Contractor with revenues in excess of $10M

- AR/AP/Reconciliations. Project budgeting and management, Job costing
- RFP response management including content creation/packaging in InDesign
- Contract negotiation with clients and vendors. Insurance review and management
- Procore implementation (2017) and support. Procore-Certified.

Webmaster, Florida State University, College of Communication and Information

From 2006 – 2011, built several iterations of CI/CCI website and maintained all web-facing technology. Created and supervised internship program.

- Managed website conversion from ColdFusion to first WordPress website on campus
- Implement and maintain Section 508 accessibility standards.
- Implement LDAP-based College-wide faculty directory.
- Maintained Windows servers in a virtual environment, including for student use in classroom activities.
- Maintained SVN repositories.
- Participated in IT planning and budget process.
- Supervised student interns working in the Office of the Webmaster
- 2007 Staff Member of the Year

Melissa Morisco Raulston

melissa@platowynne.com
(850) 491-6702

Service Learning Coordinator, Florida State University, College of Communication and Information

Founder/Director for Project SPARTA, a technology-oriented service learning/apprenticeship program for the College of Communication and Information

- Recognized on President's 2009 Higher Education Community Service Honor Roll for exemplary service efforts and service to America's communities.
- First FSU nominee for The Thomas Erlich Civically Engaged Faculty Award
- 2011 State of Florida Public University Community Engagement Educator

Adjunct Faculty, Florida State University, College of Communication and Information

Developed and taught information technology courses, including Capstone courses

- Developed and taught LIS4910 IT Project (Capstone)
- Developed and taught LIS4708 IT Perspectives (Capstone)
- Redesigned and taught LIS4366 Web Development for Information Professionals
- Developed and taught LIS4941 IT Practicum

Founder and President, ITG Solutions Inc

From 2002 – 2008, developed and managed websites, email, DNS and related services for 400+ clients in a variety of industries

- Developed and maintained InfoToGo CMS on ASP/SQL platform
- Managed physical Windows servers for web hosting. IIS, DNS, Email (IceWarp), SQL. Devised method for running PHP/WordPress on Windows platform in 2011.
- Domain name registration, management, and migration
- Content creation across a variety of industries. Writing in a diverse range of voices.
- Personnel management including offshore team
- Contract negotiations, AR/AP, tax returns, financial reporting, payroll, marketing, sales, and escalated technical support

University of Florida

2001: B.S.B.A Cum Laude, Decision and information Science

ABDALLAH MUSMAR

AI CONSULTANT
12+ YEARS OF EXPERIENCE

amusmar1@gmail.com
(412) 552-8721
Tampa, FL
in/abdallahmusmar/
github.com/abdallahmusmar

SUMMARY

With a solid foundation in advanced analytics and generative AI, my professional journey encompasses a wide range of roles, from hands-on technical projects to strategic leadership positions. My expertise in developing and deploying large language models (LLMs) and generative AI chatbots has enabled me to tackle complex analytics challenges using innovative AI and machine learning techniques. I have played a pivotal role in the creation and enhancement of generative AI applications and products, utilizing advanced technologies such as Tensorflow, Keras, and Pytorch. Additionally, I am proficient in managing CI/CD pipelines for machine learning operations with tools like Docker, Github, and Azure Kubernetes, and my expertise in cloud technologies, particularly Azure and GCP, has been crucial in developing web applications and APIs powered by machine learning models. My deep interest in the technical aspects of generative AI drives me towards opportunities where I can contribute significantly, either as a technical specialist or in a leadership role

EXPERIENCE

OCTOBER 2023 – PRESENT
INDEPENDENT AI CONSULTANT (CLIENTS: CATCH AI| PULMONARY & SLEEP OF TAMPA BAY| KAYCHA LABS)

- Spearheaded multiple Generative AI projects as an entrepreneur, leveraging cutting-edge technologies to deliver bespoke AI solutions tailored to client needs, demonstrating versatility and innovation in AI applications.

- Implemented diverse cloud solutions across multiple projects, utilizing Azure, GCP, and AWS for AI deployment, and integrating Cosmos DB, Docker, Terraform, Kubernetes, and Neo4j to enhance performance and scalability

- Utilized Python and Go for backend development, ensuring high-performance and reliable infrastructure for AI applications, and demonstrating a strong command over key programming languages in the AI ecosystem.

SKILLS

Generative AI

LLM – GPT – Llama – Mistral AI

Deep Learning:

Keras – Tensorflow - Pytorch

Cloud

Azure - Google Cloud (GCP)

Flask - Docker – Kubernetes

Chatbots (Dialogflow)

Analytics

Github

Neo4j

Team Building

Verbal Communication

Research

Python

Tableau

Power BI

SQL

CERTIFICATIONS

Deep Learning Specialization - Coursera

Data Mining Specialization – Coursera

SEPTEMBER 2018 - SEPTEMBER 2023
PWC
Promoted 3 times from ML Engineer to Director

- Demonstrated strong hands-on Python experience while leading the development and deployment of NLP and NLU models, specifically leveraging BERT and GPT architectures.

- Effectively generated AI models using frameworks like Pytorch and Tensorflow to automate text labeling processes through Active Learning with Deep Learning.

- Developed an Intelligent Search feature using Neo4j on GCP, integrating NLP techniques for enhanced search functionality across the organization..

- Structured the technical infrastructure for scalable ML projects with an emphasis on MLOps. Led efforts in building CI/CD pipelines and cloud-based deployments on Azure, ensuring efficiency and scalability.

- Designed and implemented deep learning models, including state-of-the-art NLP models and recommender systems for various applications.

APRIL 2017 - AUGUST 2018
DXC TECHNOLOGY
Apr 2017 - Sep 2018 | Data Scientist, Client: Federal Reserve Bank
- Engineered advanced analytics features for the Treasury BI System and predictive models for service desk ticket escalations.

- Played a pivotal role as a Subject Matter Expert in Tableau Desktop, enhancing visualization capabilities.

JULY 2013 - MARCH 2017
HEWLETT-PACKARD/ HEWLETT PACKARD ENTERPRISE
Jun 2016 - Mar 2017 | Data Scientist, Client: OHIO MITS
- Led the creation of Tableau dashboards that unearthed critical trends in healthcare data.

- Conducted advanced analytics to explore infancy mortality rates, using R for in-depth statistical analysis.

- Developed Python scripts for ETL processes, ensuring data integrity and compatibility with various data sources.

Nov 2015 - May 2016 | Data Scientist, Client: Hewlett-Packard

EDUCATION

UNIVERSITY OF SOUTH FLORIDA
GPA: 3.97
Tampa, FL
Doctor of Philosophy (Ph.D.) Business Administration - Information Systems (Dec 2021)

Relevant Coursework
- Machine Learning
- Deep Learning

CARNEGIE MELLON UNIVERSITY
GPA: 3.6
Pittsburgh, PA
Master of Science (M.S.) Information System Management (May 2013)

Relevant Coursework
- Independent Study on Recommender Systems
- Object-Oriented Programming in Java
- Database Management

- Pioneered standardization efforts for data extraction, analysis, and reporting processes, ensuring consistent data-driven insights.

Nov 2013 - Sep 2015 | Data Scientist, Client: EOHHS
- Managed ETL development activities for data warehousing, focusing on long-term & inpatient claims data.

PUBLICATIONS

FIGHTING MISINFORMATION ON SOCIAL MEDIA: YOUTUBE CANCER VIDEOS
Americas Conference on Information Systems / Aug 2021

CRYPTOCURRENCY MARKET PRICE SIGNALS FOR SIGNAL-TO-NOISE RATIO ON TWITTER
Winter Conference on Business Analytics / Mar 2019

TALKS

WHAT DOES IT MEAN TO BE A DATA SCIENTIST?
The Banana Data Podcast / Dec 2020

WHAT CAN YOU DO WITH UNSTRUCTURED TEXT DATA?
BrightTALK / Oct 2020

Curriculum Vitae

Juan Sebastian Muhamad

General Information

Address: 2509 Chamberlin Dr.
Tallahassee, FL 32308
Phone: (305) 303-3797
Email: Juansmuhamad@gmail.com

Education

2023	Doctor of Philosophy	Information Science and Technology College of Communication and Information Florida State University Tallahassee, Florida. Dissertation: *Exploring Factors Related to Mental Health Information Seeking and Mental Health Help Seeking in Young Adults.*
2016	Masters	Mental Health Counseling and Music Therapy Lesley University Cambridge, Massachusetts
2013	Bachelors of Arts	Anthropology Florida International University Miami, Florida
2013	Certificate	African Diaspora Studies, Global Black Studies Florida International University Miami, Florida

Professional Preparation

2016	Certificate	Neurodevelopmental Disabilities Miller School of Medicine, University of Miami Miami, Florida
2013	Certificate	Communication Development Hanen Program
2005	Certificate	Popular Music Thames Valley University London, England

Language Proficiency

English - advanced in speaking, reading, and writing
Spanish - native in speaking, reading, and writing
Portuguese - conversational

Professional Experience

2013-2018 Music therapist and mental health counselor (intern), United Way Center for Excellence in Early Education, Miami, Florida

Conduct developmental assessments, support and guide families in accessing resources, provide counseling and music therapy services, create plans/strategies to be implemented in natural environments.

2016-2018 Facilitator, *Padres Activos*, Miami, Florida

Assist in development of sessions for 4-week intervention program focused on increased engagement of fathers for the prevention of overweight/obesity among Hispanic youth. Engage with community-based organizations and members to secure sites for the program. Provide targeted educational resources to participants (father and child).

2016-2018 Consultant, community music circles for individuals with disabilities for District Five, Miami, Florida

Consult on the development of activities to increase engagement and sense of community among individuals with disabilities that are often disenfranchised. Programs focused on the use of music for wellbeing and bridging relationships to others.

2015-2018 Mental Health Expert, United Way Early Head Start Childcare Partnerships, Miami, Florida

Visit underserved early childhood centers, provide assessment, strategies, and resources. Provide report to secure funding and implement interventions.

2015-2016 Research Associate, Mediating Resilience: A study of information Communication technologies and media use among immigrant youth, Miami Florida

2010-2011 Patient Accounts, Mercy Hospital, Miami, Florida

2009-2010 Health Information Management, Sylvester Cancer Center University of Miami, Miami, Florida

2007-2009 Percussion teacher and community music group facilitator, Universidad del Atlantico, Bellas Artes, Barranquilla, Colombia

2004-2007 Curriculum Director, London Music School, London, England

Teaching

Courses Taught (as Lead Instructor and/or Teaching Assistant)

LIS 5271 Research in Information Studies (Fall 2021, Summer 2022, Fall 2022)
LIS 5411 Introduction to Information Policy (Fall 2018, Summer 2022, Fall 2022)
LIS 3021 Technical Communication in the Information Professions (Fall 2019, Spring 2022)
LIS 3353 Information Technologies (Spring 2021)

LIS 5603 Introduction to Information Services (Spring 2021)
IDS 3493 Empowering Health Consumers in the eHealth Era (Spring 2020)
IDS 2683 Unintended Consequences of Information Technologies (Spring 2020)
SPA 4556 Practicum in Developmental Disabilities (Summer 2018)

Contributions to Science

My work has focused largely on vulnerable populations often affected by health/mental health disparities. These populations represent the intersectionality of multiple vulnerabilities, which can include low socio-economic status, gender, inaccessible educational resources, disability, age, and/or country of origin. These populations require cultural-competent interventions that recognize their specific needs, beliefs, and values, as well as the importance of the localization of campaign materials that acknowledge systemic constraints.

Publications

Invited Book Chapters

Muhamad, J.S., Wendorf Muhamad, J, & Villar, M.E., (2021). *Effects of Chronic Exposure to Invalidation on People of Color in Academia: An Exploratory Study*. In Teresa Heinz Housel (Ed.), Mental Health Among Higher Education Faculty, Administrators, and Graduate Students: A Critical Perspective (pp. 105 – 120). Lexington Books.

Conference Proceedings

Muhamad, J.S. (2019). Examining the Role of Culturally-bound, Experiential Training for Increased Engagement. Proceedings of the Communication Department and the Center for Health and Risk Communication at George Mason University, USA, 5, 107. https://dchc.gmu.edu/wpcontent/uploads/2019/08/2019-Proceedings.pdf

Muhamad, J.S., Wendorf Muhamad, J., Tian, M., Gunaydin, F., Merle, P., Huse, L.K., Wibowo, P., & Agharazidermani, M. (2020). “Chinese Virus” as Anchor for Engaging with COVID-19 Information: Anchor Bias leading to Racism and Xenophobia. Proceedings of the Association for Library and Information Science Education Annual Conference: Alise 2020, USA, 385. http://hdl.handle.net/2142/108854

Wibowo, P., Wendorf Muhamad, J., Muhamad, J.S., Gunaydin, F., Merle, P., Huse, L.K., Tian, M., & Agharazidermani, M. (2020). Analysis of Public Perception of Multiple Community Issues through Social Media Mining during a Pandemic. Proceedings of the Association for Library and Information Science Education Annual Conference: Alise 2020, USA, 389. http://hdl.handle.net/2142/108854

Refereed Papers Presented at Conferences

Muhamad, J.S. (2014, March 21). The Syncretic Feedback [Conference presentation]. The 12TH Annual South Florida Latin American and Caribbean Studies Graduate Student Conference, Boca Raton, FL, United States. http://files.ctctcdn.com/e3e644ba101/bcaa8b5c-9c01-498c-a673-efc4976df1e6.pdf

Muhamad, J.S. (2019, March 30). The Syncretic Feedback: From Africa to the Americas and Back [Conference presentation]. Balancing the Mix: A Conference on Popular Music and Social Justice, Memphis, TN, United States. https://stjenglish.com/balancing-the-mix-a-conference-on-popular-music-and-social-justice/

Muhamad, J.S. & Wendorf Muhamad, J. (2019, May 24 – 28). Music and Minority Health: Examining the Role of Music-based Approaches for Mental and Physical Health among Hispanics [Conference presentation]. The 69th Annual International Communication Association Conference, Washington, D.C., United States.
https://cdn.ymaws.com/www.icahdq.org/resource/resmgr/conference/2019/2019printprogram.pdf

Muhamad, J.S. (2019, October 22 – 23). Boundary Object between Faculty and Students' Social Worlds: Boundary Crossing for Mental Health [Conference presentation]. Florida International University Colloquium Mental Health on University Campuses: Current Issues and Challenges, Miami, FL, United States.
https://carta.fiu.edu/communication/?s=colloquium

Muhamad, J.S. (2019, November 14 – 17). Surviving the Label: Examining the Burden of the Socially Constructed Disabilities Label [Conference presentation]. NCA 105th Annual Convention: Communication for Survival, Baltimore, MD, United States.
https://www.natcom.org/convention-events/conventionresources/archives-historical-information/past-future-conventions

Wendorf Muhamad, J. & Muhamad, J.S. (2019, November 14 – 17). Unpacking the Gendered Digital Divide: Exploring Women's Physical and Intellectual Access to ICTs [Conference presentation]. NCA 105th Annual Convention: Communication for Survival, Baltimore, MD, United States.
https://www.natcom.org/convention-events/convention-resources/archives-historical-information/pastfuture-conventions

Muhamad, J.S. (2020, April 1 – 5). Interpersonal Communication as Disruption of the Medical Model Approach to Therapy Services for Children with ASD [Conference presentation]. The 90th Annual Southern States Communication Association Convention, Frisco, TX, United States.
https://www.ssca.net/2020convention.

Wibowo, P., Muhamad, J.S. (2020, July 12 – 17). The Need for Open Research Data and Research Data sharing Policy to Enforce Collaboration for Global Emergency Response: The Fight against Infectious Diseases [Conference presentation]. IAMCR 2020 Annual Conference, Tampere, Finland.
https://iamcr.org/tampere2020/abstract-books

Muhamad, J.S., & Lanh, A. (2020, August 6 – 8). Using Cognitive Behavioral Therapy to Enhance Communication Efforts in Schools to Prevent Suicidality [Conference presentation]. APA 2020, Online.
https://irp-cdn.multiscreensite.com/a5ea5d51/files/uploaded/APA_2020_Program.pdf

Muhamad, J.S. (2020, September 26 – 28). Information Sharing in an Information Society for the Prevention of Suicide for Young Adults [Conference presentation]. International Colloquium on Mental Health in College Campuses, Toulouse, France.
https://mental-health-on-campuses.org/en/

Muhamad, J.S., Huse, L.K., Wibowo, P., Tian, M., Gunaydin, F., Wendorf Muhamad, J., Merle, P., & Agharazidermani, M. (2020, November 19 – 22). Information Sharing for Collective Action: Exploring Retweets during the COVID-19 Pandemic [Conference presentation]. The 106th National Communication Association conference. Indianapolis, IN, United States.

https://www.natcom.org/sites/default/files/annualconvention/2020%20Program%20PDF%20for%20website.pdf

Tian, M., Wendorf Muhamad, J., Yang, Q., Muhamad, J.S., Wibowo, P., Huse, L.K., Merle, P., & Gunaydin, F. (2020, November 19 – 22). How to persuade people in China and in the U.S. to stay at home? A content analysis between self- protection information on Weibo and Twitter regarding the spread of coronavirus [Conference presentation]. The 106th National Communication Association conference. Indianapolis, IN,
United States.
https://www.natcom.org/sites/default/files/annualconvention/2020%20Program%20PDF%20for%20website.pdf

Muhamad, J.S. (2021, July 11 – 15). Cognitive Reframing and Health Information Seeking Behavior in Young Adults: Supporting Change and Shifting Negative Perception [Conference presentation]. IAMCR 2021 Annual Conference. Nairobi, Kenya.
https://iamcr.org/node/17292

Participatory Based Interventions

Participatory-based, or community-focused research, considers partnerships with community members essential, as they are experts and often key to the success of interventions. Instrumental to PAR/CBPR efforts in a holistic approach to understand phenomena, including but not limited to ecological, cultural, social, and psychological contexts. This methodology is particularly effective for interventions that require continued support for the community. As a mental health professional, my work has focused primarily on service to communities through keen understanding of held attitudes, beliefs, and values. In that capacity, I contribute to the field by keeping other practitioners abreast of localization issues of health-based campaigns. Further, my work seeks to bridge the gap between health professionals and community members through the development of appropriate and collaborative teams that recognize local expertise and population-level factors.

Muhamad, J. (2015). Educare Family voices: Engagement Takes a Village. Educare Learning Network proceedings.

Muhamad, J. Enhancing Communication in Rett Girls through Community Music Therapy.

Professional Memberships and Affiliations
American Music Therapy Association
American Psychological Association
National Communication Association

Community Engagement and Service

2011-2018	Policy Member, United Way Center for Excellence in Early Education (Educate)
2011-2018	Community Advocate, Children with Special Needs Council
2015-2017	Chair. Policy Committee, United Way Center for Excellence in Early Education
2015-2016	Representative. Educate Learning Network
2015	Representative. National Educare Learning Network
2014-2015	Vice Chair. Policy Committee, United Way Center for Excellence in Early Education
2013-2015	Representative. Head Start and Early Hear Start Policy Council
2011-2014	Community Representative. Policy Committee, United Way Center for Excellence in Early Education