FarmAssist: Leveraging Large Language Models to Empower Persons With Disabilities in Assisting Small and Mid-size Farmers with Sustainable Practices and Funding Access.

Proposal submission on grants.gov [DEADLINE: Tuesday, September 17, 2024.]

- ☐ Project Summary: limited to 250 words in 12-point Times New Roman, must align with SBIR/STTR Phase 1 Program goals, follow NIFA's guidelines and template, and be submitted as a PDF.
- ☐ Project Narrative [17 pages max].
- ☐ Bibliography
- ☐ Research.gov documentation
  - ☐ Collaborators and other affiliations
  - ☐ Current and pending support
  - ☐ Bio sketch
  - ☐ Budget
  - ☐ Data management
  - ☐ Equipment and facilities
  - ☐ Sub-award documentation

Project Narrative - Seventeen (17) pages max

- ✓ 1. Responsiveness to USDA SBIR/STTR Program Priorities (suggested length: 1 page)
- ✓ 2. Identification and Significance of the Problem or Opportunity (suggested length: 1-2 pages)
- ✓ 3. Background and Rationale (suggested length: 2-3 pages)
- ✓ 4. Relationship with Research or Research and Development (suggested length: 1-2 pages)
- ✓ 5. Technical Objectives (suggested length: 3-4 pages)

- ✓ 6. Work Plan (suggested length: 1 page)
- ✓ 7. Related Research or Research and Development (suggested length: 1-2 pages)
- ✓ 8. The Market Opportunity (suggested length: 1 page)

Letters

- ✓ Letter of support from Kirchner.
- ✓ Letter of support from Knowbility.
- ✓ Letter of support from SUNY Assistive tech.
- ✓ Letter of support from South Dakota.
- ✓ Letter of support from TAMU.
- ✓ Letter of support from farmer Michele.
- ✓ Letter of support from farmer Joe.

FarmAssist: Project summary.

Empowering **Persons with Disabilities (PWDs)** through emerging technologies presents a unique opportunity to strengthen agricultural resilience to climate change, drive the adoption of sustainable practices, and foster innovation. Small and mid-sized farmers often face significant challenges in securing the funding needed to support sustainable agricultural practices. At the same time, **PWDs**, particularly those who were previously employed in agriculture but can no longer perform labor-intensive tasks, are an underutilized resource. According to the USDA's AgrAbility program, over 21 million Americans with disabilities live in rural areas, many of whom are engaged in agriculture or have prior experience in the field. However, these individuals often lack the necessary tools, resources, or accommodations to continue contributing to the industry. **FarmAssist** bridges this gap by empowering these individuals to leverage their experience to assist farmers in securing vital funding through telework. **FarmAssist** aims to explore emerging technologies such as **Machine Learning** (ML), **computer vision**, and **large language models (LLMs)** to create new job opportunities for emerging teleworkers, while also providing farmers with the support they need to navigate complex financial and technical requirements to access incentive programs.

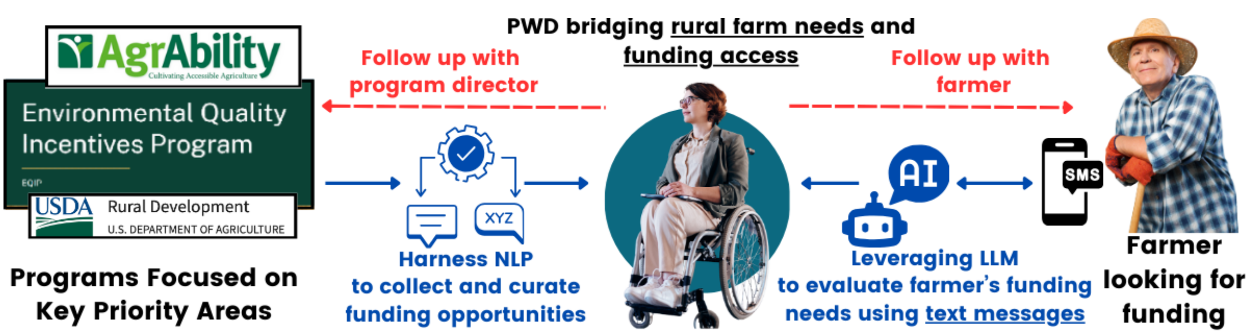

Figure 1: **FarmAssist** aims to support small and medium-sized farms in accessing critical funding resources by leveraging advanced technologies and creating meaningful telework opportunities for Persons with Disabilities (PWDs). (Left) Our proposed **Natural Language Processing (NLP) block** is intended to **collect and curate funding opportunities** for small and medium farms. This system processes **unstructured data**, simplifying it for easier analysis by a PWD. (Center) **FarmAssist** is the core technology designed to **empower PWDs** to bridge the gap between **small farmers** and **funding opportunities**. (Right) Our proposed **Large Language Model (LLM) block** will serve as the **first point of contact** with the farmer, gathering information on the farmer's **needs and farm/crop characteristics**. The LLM will then **organize this information** and suggest a **funding path** based on programs identified by the NLP system. The **PWDs** will play a crucial role in **validating the selections** made by the LLM, serving as the essential **human connection** between farmers and funding program directors. By collaborating closely with farmers, they will help streamline access to financial resources, promoting the adoption of **sustainable agricultural practices**.

FarmAssist: Leveraging Large Language Models to Empower Persons With Disabilities in Assisting Rural Farmers with Sustainable Practices and Funding Access.

1 Responsiveness to USDA SBIR/STTR Program Priorities

The proposed project, **FarmAssist**, empowers persons with disabilities (PWDs) in rural areas to assist farmers in securing funding, enhancing agricultural resilience to climate change, and promoting inclusion through innovation. Aligned with the **USDA's 2022-2026 Strategic Plan**, FarmAssist addresses climate challenges while promoting rural prosperity. By streamlining access to financial assistance, **FarmAssist** supports the adoption of **sustainable and energy-efficient practices** and the creation of a **climate-resilient agriculture** system. Through programs like the Environmental Quality Incentives Program (EQIP) and the On-Farm Energy Initiative, PWDs can help farmers navigate and capitalize on these incentives. The USDA's focus on **improving program delivery and customer experience** aligns with FarmAssist's use of **emerging technologies**, such as **Machine Learning** and **Large Language Models (LLMs)**, to simplify access to incentive programs and enhance farmers' efficiency in adopting **climate-smart agriculture** practices.

Figure 1: **FarmAssist** aims to create meaningful telework opportunities for **Persons with Disabilities (PWDs)** by harnessing **Emerging Technologies** such as **Natural Language Processing (NLP)** and **Large Language Models (LLMs)**. These technologies empower **PWDs** as **Emerging Workers**, enabling them to focus on critical tasks that require human oversight and expertise, significantly enhancing their ability to help farmers in rural areas secure funding. This initiative bridges the gap between farmers and the essential resources they need, fostering **Emerging Work** opportunities that drive sustainable agricultural growth.

2 Identification and Significance of the Opportunity

In the United States, PWDs have the power to transform the agricultural landscape by leveraging cutting-edge technologies to help bridge the gap between farmers in rural areas and the essential funding needed to advance agricultural conservation practices. Between 2008 and 2016, over half a million farmers reported a form of disability in the US [1]. On average, almost two out of ten farmers (19.2%) and nearly one out of ten farmworkers (9.0%) had a disability [2], with mobility issues and musculoskeletal challenges being the most common [3]. Despite policies aimed at integrating people with disabilities into fair, stable, and well-compensated employment [4], workers with disabilities still have been found to earn 17.1% less than their non-disabled counterparts [5]. While emerging technologies like Starlink [6] are improving rural internet access, many farmers with disabilities still lack adequate support to access new emerging virtual work. This gap leads to lower incomes, higher poverty rates, and fewer career opportunities, as measured through programs like AgrAbility [7]. A limited access to off-farm emerging employment further compounds their economic instability [8], underscoring the need for a transformative system to help them fully benefit from emerging technologies and secure stable livelihoods. **As seen in Figure 1 our proposed project, FarmAssist, seeks to create meaningful telework opportunities for PWDs by harnessing cutting-edge technologies such as Natural Language Processing (NLP) and LLMs. These tools will empower PWDs to assist other farmers in their community in navigating the vast amount of unstructured information available online to successfully secure funding from programs such as EQIP and AgriAbility.**

Figure 2: **FarmAssist** aims to support rural farms in accessing critical funding resources by leveraging advanced technologies and creating meaningful telework opportunities for Persons with Disabilities (PWDs). (Left) Our proposed **Natural Language Processing (NLP) block** is intended to **collect and curate funding opportunities** for rural communities of farmers. This system processes **unstructured data**, simplifying it for easier analysis by a PWD. (Center) **FarmAssist** is the core technology designed to **empower PWDs** to bridge the gap between **rural farmers** and **funding opportunities**. (Right) Our proposed **Large Language Model (LLM) block** will serve as the **first point of contact** with the farmer, gathering information on the farmer's **needs and farm/crop characteristics**. The LLM will then **organize this information** and suggest a **funding path** based on programs identified by the NLP system. The **PWDs** will play a crucial role in **validating the selections** made by the LLM, serving as the essential **human connection** between farmers and funding program directors. By collaborating closely with farmers, they will help streamline access to financial resources, promoting the adoption of **sustainable agricultural practices**.

After interviewing 43 rural farmers in South Dakota (see Letter of Support from Mr. Jason Schoch), we identified three key technical challenges, as shown in **Figure 2**: (1) **Information Overload (Left)**: Farmers struggle to navigate the overwhelming amount of information related to government programs, such as program types, eligibility criteria, deadlines, application processes, and areas served. Our team proposes the use of NLP techniques to organize this information effectively which is crucial to ensuring farmers can access the programs they need. (2) **Farmer-Technology Interaction (Right)**: Our interviews revealed significant barriers to farmers' interaction with technology. While internet access is improving, many farmers lack reliable connectivity, appropriate devices, or the software needed to access government resources. However, all farmers have cell phones, and we found that text messaging was a far more effective communication tool than email. FarmAssist plans to leverage LLM technologies to gather essential information from farmers via text message, such as their goals, location, land-use, and management, and then organize this data to streamline the process. (3) **NLP and LLM have its own Limitations (Center)**: While promising NLP and LLM technologies are not yet capable of handling complex queries without human oversight. To solve this challenge we propose enabling **PWDs** to use their expertise to ensure that FarmAssist successfully connects farmers with the resources they need.

3 Background and Rationale

Federal programs that support rural farmers are receiving unprecedented investment, driven by major initiatives like the Inflation Reduction Act (IRA), the Farm Bill, and growing support from both public and private sectors. With the USDA projecting $4 billion only in conservation payments for 2024—a 10% increase due to the IRA—these efforts underscore the crucial role of nature-based solutions in tackling climate change [9]. Practices like cover cropping, reduced tillage, and agroforestry are helping mitigate environmental impacts while promoting sustainable farming [10, 11]. However, despite the availability of substantial funding resources, farmers still face several barriers to accessing these incentive programs, including: (1) **Lack of Awareness**: Many farmers lack awareness of available programs or are uncertain about how to initiate the application process for these incentives. A study by the American Farmland Trust revealed that in some regions, up to 40% of farmers were unaware of federal programs [12]. (2) **Complex Eligibility Requirements**: Programs often have specific eligibility criteria, making it difficult for farmers to determine whether they qualify [13]. (3) **Site-Specific Practices**: Farmers must align conservation practices with the unique conditions of their farms, such as conservation goals, soil type, climate, and cropping systems [14]. (4) **Administrative Burden**: Applying for and complying with conservation programs often requires significant time, paperwork, and ongoing monitoring, which can be overwhelming for smaller farms with limited resources [15].

FarmAssist's Technical Approach. A popular tool to help Illinois farmers access financial incentive programs for conservation practices is the **Illinois Sustainable Ag Partnership (ISAP) - Financial Incentives Database (FIND Tool) [16]**, a free online database with more than 60 programs. However, while the FIND Tool allows users to filter programs by state, county, current operation, and the practices the farmer is interested in, it is still **limited in curating programs tailored to the specific needs of an individual farmer**. For example, the tool does not account for **unique farm variables like soil type, climate conditions, or specific conservation goals**, which can significantly impact program eligibility and suitability. This is a significant limitation, as it **constrains the ability of farmers to use the tool as a comprehensive resource** to uncover programs that best align with their specific operations and needs. Consequently, many farmers may miss out on **opportunities for financial incentives** that could greatly benefit them.

Natural language processing techniques, particularly knowledge graphs, offer a powerful framework for interpreting vast amounts of agricultural data [17, 18, 19]. They can capture and represent the **relationships between farming practices, conservation programs, and environmental factors** in large databases, enabling users to explore and analyze this information in innovative ways. **Knowledge graphs provide the ability to visualize complex connections** between these entities within a network [20, 21]. Specifically, they help uncover **associations between farmers' needs, farm characteristics, and available funding or programs that best suit their operations—**associations that were previously difficult to identify using traditional text-based methods. However, a significant challenge in making knowledge graphs an effective tool for agriculture is **converting unstructured and often ambiguous textual data** from government websites and farmer records into a **structured format that accurately represents these relationships**. This process involves **abstracting, disambiguating, and linking information** to make it more accessible and usable for rural farmers. Additionally, to truly **bridge the gap between farmers and the resources they need**, a **user-centered development approach** is essential to understand how farmers interact with the data and how this impacts their decision-making.

FarmAssist is designed to address these challenges through a comprehensive, tech-enabled solution that combines Artificial Intelligence (AI) and the unique contributions of PWDs. By leveraging these technologies, FarmAssist aims to streamline the process of accessing financial support and implementing conservation practices while ensuring that each farmer receives personalized assistance. The following strategies illustrate how **FarmAssist** leverages emerging technologies and the expertise of **PWDs** to solve these challenges and provide targeted assistance to farmers:

1 Information Access through NLP: To overcome the barrier of information overload, NLP technologies have been successfully implemented to simplify and organize complex and often confusing government program information [22]. Our proposed system innovation will allow farmers to easily search for and understand available data while considering the rural farmer's particular needs, a problem that is still unsolved today.

2) Improved Farmer-Technology Interaction with LLMs: Our customer interviews revealed that farmers struggle with web-based tools to access funding [23]. **Farmers requested the ability to interact with FarmAssist via text messages** (see letters of support from Michele, Jason, and Joe). To meet this need, our team will evolve current LLMs technologies to solve the agriculture challenges presented here, by training them on relevant data to provide clear and useful recommendations. We expect that a text-message interface will reduce barriers, allowing farmers with limited internet or technical skills to easily access tools without needing complex software.

3) Personalized Assistance from PWDs: PWDs can bridge the gap between our language technologies and farmers by refining the outputs of NLP and LLMs within FarmAssist's adaptive prototyping platform. PWDs will ensure farmers receive relevant information while facilitating the smooth adoption of new technologies, all while creating valuable employment opportunities.

End-User Needs. The team wants to highlight that our system design is based on a series of interviews with rural farmers and program managers that are in charge of extension programs. **We want to point out a statement made by them (See letter of support from Jason:)** *"The FarmAssist proposal offers a much-needed solution for farmers, particularly those in rural and underserved communities. Drawing from my experience with the AgrAbility Beginning Farmer Project, I can attest to the importance of solutions like FarmAssist in helping farmers access new sources of funding, which is often one of the biggest hurdles they face"*

FarmAssit Impact. By facilitating greater access to conservation programs and reducing the administrative burden on farmers, FarmAssist will play a crucial role in promoting the adoption of nature-based solutions in agriculture. This project aligns with broader national efforts to combat climate change, improve soil health, and enhance the sustainability of food production systems [24]. **Notably, a farmer, who has agreed to be our first test user, has said (See letter of support from Joe):** *"In rural communities like ours, where guidance can be limited, this project can greatly assist farmers in adopting sustainable methods and ensuring long-term success. One key aspect that makes FarmAssist particularly accessible is its integration with text message platforms, which significantly lowers the barrier of entry for farmers. By allowing access to the platform via text, the technology can be more easily embraced by farmers who may not be familiar with complex digital systems, making it more practical and effective for a broader range of users"*.

4 Relationship with Research or Research and Development

We intend to complete this project in two stages, as seen in Figure 3. First, we aim to construct a rapid prototyping platform in Phase 1 based on a combination of design thinking and lean methodology to evaluate the use of LLM and NLP technologies for the workplace (**FarmAssist**) that incorporates a succession of prototypes. Phase 1 will also focus on assembling a well-connected interdisciplinary team of agricultural professionals, data scientists, occupational therapy professionals, and industry stakeholders, as presented in Fig 5, to ensure that FarmAssist v1 is highly impactful to PWDs and farmers in rural America.

Our vision for Phase 2 is to develop FarmAssist v2 into a minimum viable product (MVP) derived from the Phase 1 rapid-prototyping platform, improving the system into a fully functional AI-Agent-based platform [25, 26, 27, 28] that amplifies the work of PWDs to help address the

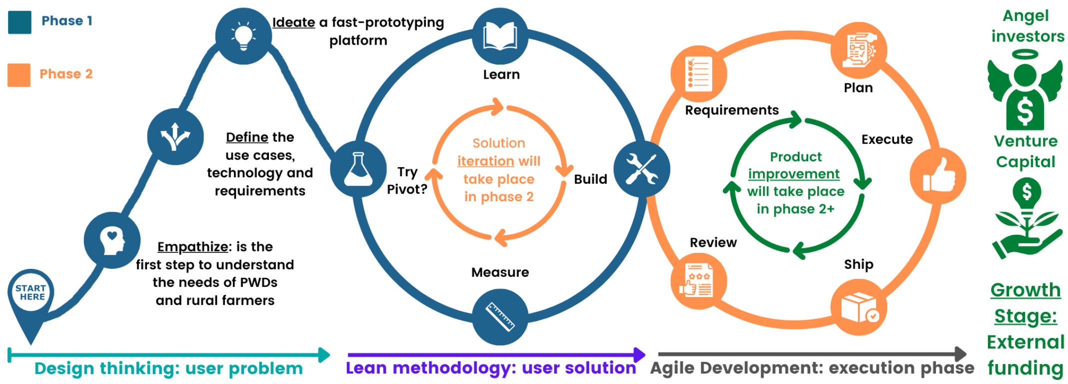

Figure 3: Our phase 1 work will use design thinking to empathize, define, and ideate while building a platform for rapid prototyping and testing a minimum viable **prototype**. Phase 1 will set the stage for phase 2 to create a higher-fidelity minimum viable **product** using an iterative lean approach based on a build-measure-learn iterative work to help this team get FarmAssist to the hands of rural farmers and PWDs faster. At the onset of Phase 2 (March 2027), the team will have used Agile development to mature FarmAssist as an automated AI-Agend based platform and de-risked the technology enough to reach out to external private investors, such as Kirchner Group.

needs of farmers. Both FarmAssist v1 and v2 will require continuous feedback from PWDs and rural farmers. If successful, the FarmAssist v2 platform will be extendable to a range of critical farming tasks, including funding application support, technical assistance for conservation practices, and financial management.

(a) The development of FarmAssist v2 will offer substantial technical, economic, and social benefits. **Technically**, the platform will be a reference of how **AI Agent-based** technologies can simplify and accelerate the funding process for farmers while providing PWDs with meaningful telework opportunities, thereby fostering workforce inclusion. **Economically**, FarmAssist can help rural farmers secure more funding for sustainable practices, contributing to a more resilient and sustainable agricultural sector. **Socially**, the platform empowers PWDs by facilitating their participation in the agricultural workforce, promoting equity and inclusivity in rural communities.

(b) The total cost of developing FarmAssist is modest compared to the long-term benefits it offers. The cost of Phase 1, which focuses on establishing a prototype, is projected to be below \$132,000, while Phase 2 is expected to cross the million-dollar mark and involve external investors. The economic benefits include improved access to funding for rural farmers, reduced administrative costs, and increased employment opportunities for PWDs through telework. The financial and social returns, including the sustainability impact, will far exceed the initial development costs.

(c) FarmAssist's success could influence policies related to workforce inclusion and rural development. By creating telework opportunities for PWDs and streamlining funding access, the platform may prompt government support for broader remote work initiatives and AI-driven funding solutions. Additionally, its impact on grant processes could drive policy reforms aimed at improving efficiency in agricultural subsidies and funding programs, aligning with national priorities on sustainability and social equity.

5 Technical Objectives

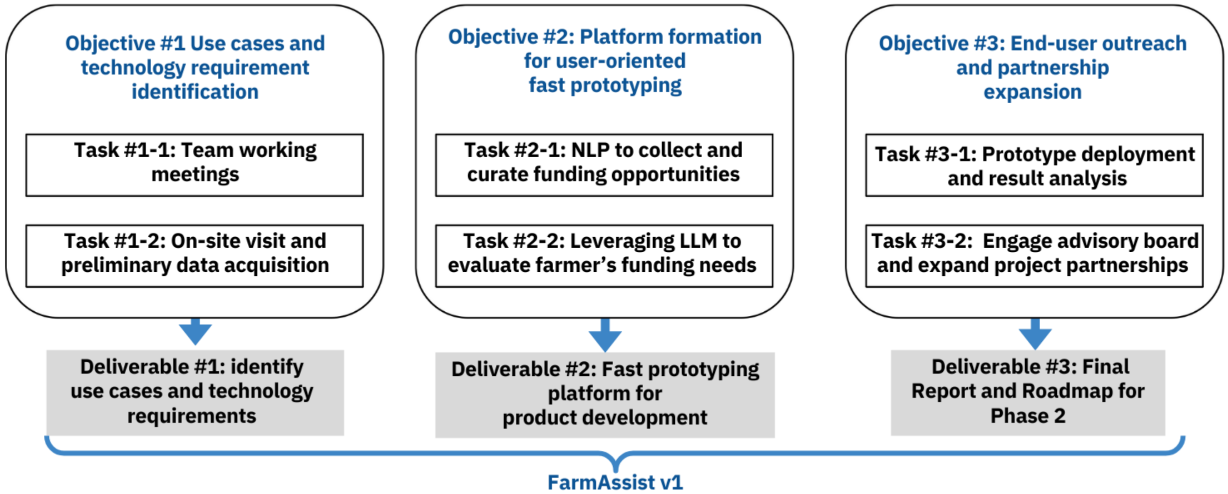

Figure 4: Overview of the Technical Objectives.

Overview of the Technical Objectives.

The primary goal of Phase 1 is to design and validate a digitized work platform tailored for the agricultural industry, integrating AI to enable telework opportunities for PWDs. To accomplish this, we will pursue three key objectives, as outlined in Figure 4. In **Objective #1**, we will define specific use cases and leverage our existing understanding of the industry and technological requirements to address the challenges rural farmers face in securing funding, and how our proposed AI-based technologies can enhance workflows for both farmers and PWDs supporting them remotely. In **Objective #2**, we will develop a fast prototyping platform to test and validate the effectiveness of our LLM and NLP-based solutions for both PWDs and rural farmers. In **Objective #3**, we will engage the farmers who have agreed to be early adopters for further testing and feedback, while also establishing key partnerships to complement our capabilities and ensure a user-centric solution that will be implemented in Phase 2. The following sections provide a detailed breakdown of each objective and the associated work plan.

6 Work Plan

FarmAssist is an ambitious project designed to transform how rural farmers access funding while creating meaningful telework opportunities for PWDs. Given the scale and complexity of this initiative, a multidisciplinary team is essential to ensuring its success. Figure 5 provides an overview of the diverse support network assembled for FarmAssist, which includes experts from various fields, as demonstrated by the letters of support for this proposal. Leading the project is **Dr. Alfredo Costilla Reyes**, an expert in machine learning, NLP, and LLMs, responsible for product development and business strategy. **Dr. Diana Zapata Rojas**, with her deep expertise in agricultural systems, nature-based solutions, and environmental sustainability, leads the development of decision-support tools for farmers. To ensure FarmAssist meets the needs of PWDs, the assistive technology our team has recruited an Assistive Technology Board represented by **Dr. Devva Kasnitz** and **Sharron Rush**, experts in accessibility and inclusion.

The team has also secured a collaboration with our rural farmer outreach facilitators led by **Jason Schoch** from SDSU Extension and **Dr. Nithya Rajan** from Texas A&M University, ensuring that the platform is highly relevant to rural farmers. **Blair Kirchner** contributes as our business advisor, bringing his expertise from the Kirchner Impact Foundation, while rural farmers **Michele Santangelo and Joe Turkovich** have agreed to provide critical feedback as early adopters of FarmAssist to ensure the platform is practical, user-friendly, and aligned with the specific needs of farmers in the field. **Their insights will be invaluable in refining FarmAssist's features to better serve its target audience and achieve widespread adoption in Phase 2 of this project.** Below, we elaborate on the tasks this multidisciplinary team will perform to ensure that FarmAssist is not only technically sound but also socially impactful and scalable tool.

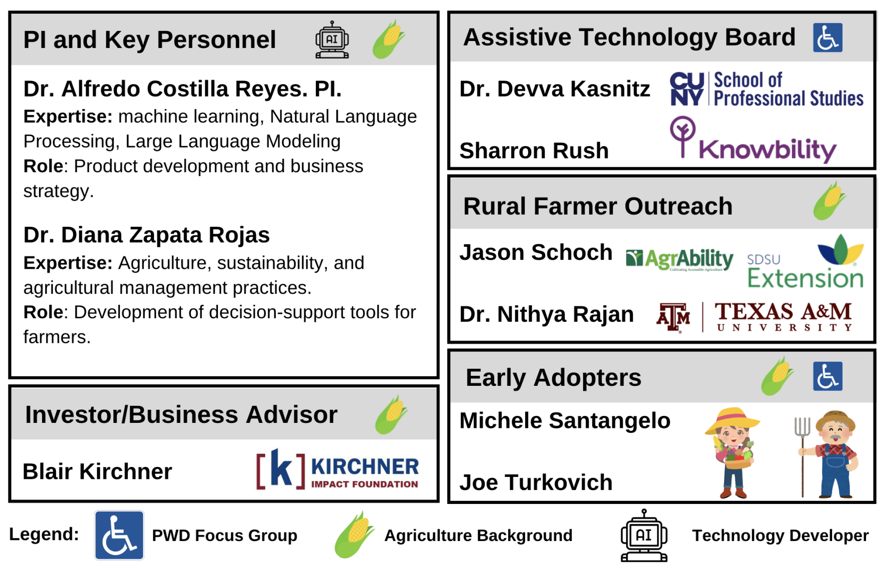

Figure 5: Overview of the multidisciplinary support network for FarmAssist, demonstrating the collaboration of experts across various fields, as highlighted in letters of support 1 through 7. These partnerships play a vital role in addressing the technological, agricultural, and workforce integration challenges of the project.

Objective #1: Use cases and technology requirement identification. In this objective, we will engage with scientists, farmers, and PWDs to gather information, better understand user and industry needs, and identify relevant technological requirements.

Task #1-1: Teamwork meetings. We will hold bi-weekly working sessions with our core team from March 2025 to November 2025 to facilitate collaboration, address key requirements, and draw on the team's expertise. The PI will organize a kickoff meeting with our advisory board in March 2025, followed by monthly meetings. The PI and Key personnel will engage farmers and PWDs through focus groups and formal interviews. We will organize breakout sessions and brainstorming activities to shape our approach.

Task #1-2: On-site visit and preliminary data acquisition. Building on our existing knowledge of farm operations, this three-day study will deepen our understanding of agricultural sites through four focused activities: 1) engaging with farm managers to gain further insight into their daily workflows, financial and sustainability challenges, crew size, and current practices; 2) reviewing historical farm records to pinpoint areas where technical support is most needed; 3) conducting interviews with farm workers, particularly those with disabilities, to assess their specific concerns and limitations; and 4) collecting targeted on-site data to facilitate the preparation of our minimum viable prototype for testing. The study will focus on rural farms in California, Texas, and South Dakota, with particular emphasis on those implementing conservation practices, ensuring our solution's broad applicability.

Objective #2: Plarform formation for user-oriented fast prototyping. For this objective, we will address the challenges, opportunities, and technical details in accessing public agricultural resources by using NLP techniques to collect, curate, and organize funding opportunities. The system will leverage **web crawlers, Named Entity Recognition (NER), part-of-speech (POS) tagging, and a Knowledge Graph (KG)** to structure unorganized agricultural data and provide tailored query recommendations to meet farmers' needs.

Task #2-1: NLP to collect and curate funding opportunities. When accessing public agricultural resources like AgrAbility, several challenges are encountered: 1) **Unstructured Data**—information is often presented in difficult-to-process formats like webpages and PDFs, limiting the extraction of useful insights; 2) **Implicit Relationships Among Resources**—farmers may miss relevant alternative solutions, such as funding that supports conservation practices, due to a lack of clear connections between needs and available resources; 3) **Domain-Specific Terminology**—the diverse agricultural terminology makes it difficult for PWDs to find relevant information using the correct terms; and 4) **Information Overload**—the abundance of data, including outdated or irrelevant information, complicates the search for up-to-date and relevant resources.

To address these challenges, we propose leveraging NLP techniques. To collect data, we plan to use web crawlers to gather information from webpages and documents on platforms like AgrAbility. Through **NER** and **POS tagging** [29], we can extract key terminologies from the data and build relationships for these resources. Additionally, we will integrate a **KG** [30] and a **query recommendation system** [29] to deliver more precise and tailored results that meet farmers' specific needs. PWDs will help validate the recommendation results and deliver them back to the farmers. Figure 6 is an example input and output of our NLP model. The input data includes relevant farmer information, and the model produces a list of matching funding programs based on the farmer's needs. Figure 7 illustrates the entire workflow of our proposed NLP model.

Technical Details

Web Crawler will be used to collect unstructured information from web pages and documents. The crawler automatically follows hyperlinks on the site to gather relevant content, such as PDFs and webpages. Our team will parse HTML and PDF formats to extract text data using tools **Scrapy** [31] and **BeautifulSoup** [32] for web pages, and **Tabula** [33] for PDFs. The extracted data will then be stored in a structured JSON format for further processing by NLP models.

Input		Output
Location	California	1. CUSP (California Underserved and Small Producers Program)
Land Size	80 acres	2. Organic Transition Pilot Program
Crop type	Corn	3. Conservation Agriculture Planning Grant Program
Size operation	Small farm	4. EQIP (Environmental Quality Incentives Program)
Income	Low income	5. Healthy Soils Program
Environmental focus	Soil erosion	

Figure 6: Example input and output of our NLP model.

KG Construction: NER and POS. The construction of a **KG** is crucial for linking agricultural terms and relationships systematically. We propose to use two techniques: 1) NER will be used to extract key entities from unstructured text, such as ”crop,” ”funding,” ”season,” and ”location,” and recognize them as nodes in the knowledge graph. To enhance accuracy in identifying agriculture-specific terms, we propose to implement pre-trained models like **BERN** [34] and do a fine-tune on agriculture-related data we have collected and other agricultural datasets; 2) POS tagging will help identify the semantic relationships between words in a sentence. This is essential for extracting relationships between entities, such as identifying which funding programs are explicitly or implicitly associated with specific requirements or conservation practices. To extract these relationships from raw text, we will use the pre-trained POS tagging model **NeuralCoref** [35]. Once the agricultural entities and relationships are identified, they will be organized as **triples (head, entity-relation-tail, entity)**, forming the foundation of the agriculture-specific knowledge graph. This structured representation will make information retrieval more efficient and accurate.

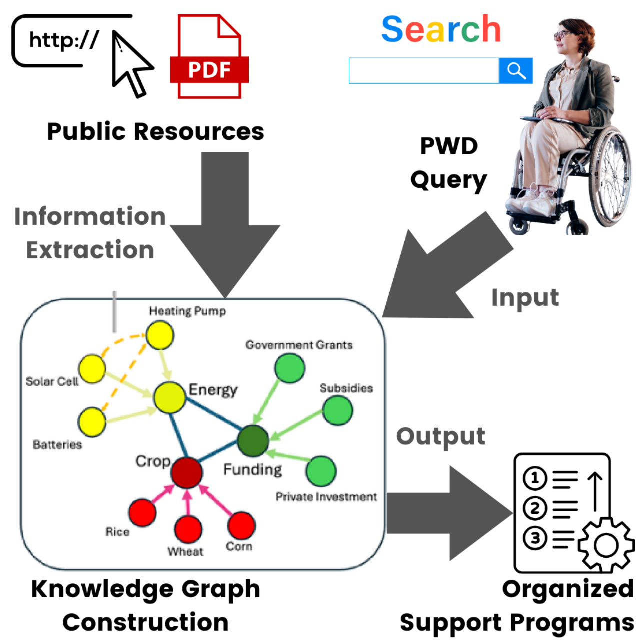

Figure 7: An overview of our NLP workflow.

Query Recommendation System. This system refines farmer queries using relationships within the agricultural knowledge graph through three main processes: 1) **Learning Program-**

Term Relationships—a recommendation model identifies implicit connections between agricultural support programs, terminologies, and resources (e.g., conservation practices, funding), refining farmers' queries; 2) **Program-to-Terminology Network**—an optimized embedding matrix maximizes relevant associations between agricultural terms and resources while minimizing irrelevant links; 3) **Optimized Query Suggestions**—based on these relationships, the system suggests refined queries. For example, if a farmer needs "corn conservation funding," the system might suggest related queries such as "regional agricultural grants" or "incentives for sustainable practices."

Evaluation and Validation. Our system's performance will be evaluated based on the following metrics: 1) **Precision and Recall Scores**—we aim for a precision score of 85% or higher, meaning that at least 85% of the suggested agricultural programs presented by FarmAssist are relevant to the farmer's needs. The recall score target is 80%, ensuring the system retrieves most of relevant programs; 2) **Retrieval Time Optimization**—the system will be optimized to achieve an average retrieval time of under 2 seconds, providing rapid responses to user queries without compromising accuracy.

Task #2-2: Leveraging LLM to evaluate farmer's funding needs. To assist PWDs in addressing farmers' needs, the proposed LLM must process natural language inputs, which are often vague or non-technical. Therefore, our LLM must handle these specific challenges: 1) **Ambiguity in Farmer Needs**—farmers may provide incomplete or ambiguous information, such as requesting "help with crops" without specifying the type of assistance (e.g., seeds, equipment); 2) **Domain Adaptation**—general LLMs may not perform optimally without being fine-tuned for agriculture. Domain-specific knowledge, such as understanding farming practices, crop types, and regional variability, is critical for accurately assessing the farmers' needs; 3) **Ensuring Accuracy**—to provide meaningful results, the LLM must ensure precise interpretation to match farmers with appropriate funding programs, minimizing the risk of miscommunication.

Our LLM will use a *transformer-based model* [36] to interpret farmers' requests, applying techniques like *tokenization*, *positional encoding*, and *attention mechanisms* to capture context and structure. By fine-tuning the LLM on agriculture-specific datasets and vocabulary, our model will undergo *transfer learning*, which can enhance its understanding of agricultural terms. Domain-specific vocabulary embeddings enable our LLM to better represent farmers' language, leading to greater accuracy in evaluating farmers' needs. The LLM will then convert unstructured data into JSON format and then tables, enabling PWDs to efficiently verify results and match farmers with relevant support opportunities. Figure 8 illustrates how our LLM is expected to interact with farmers vis text message to gather information and generate structured data. This data will then be provided to PWDs for further processing and validation.

Building Prototype Model:

1. Build a GPT Architecture. Our LLM will be based on the *Generative Pre-trained Trans-*

former (GPT) model [37]. Our GPT model will use a transformer architecture with multiple layers of attention mechanisms that process sequential text input, allowing the model to handle context and generate coherent responses. The architecture will have the following key components: 1) **Self-Attention Mechanism**—to capture long-range dependencies between words; 2) **Positional Encoding**—to ensure that the model understands the order of words; and 3) **Feedforward Neural Network**—to transform the output of the attention layer into meaningful representations.

2. Pre-Train on a Large Corpus of Text Data. The model will be pre-trained on a large, diverse corpus of text data to ensure it understands general language patterns. The dataset will include sources *Common Crawl* [38], *BooksCorpus* [39], and our own data collected in Objective #1, which provide millions of text samples covering various topics. The training objective will be next-word prediction, where the model learns to predict the next word in a sentence, enabling it to generate coherent text. The pre-training step will be conducted using *unsupervised learning* with backpropagation and *stochastic gradient descent (SGD)*. Once the general language patterns are learned, the model will be ready for task-specific fine-tuning.

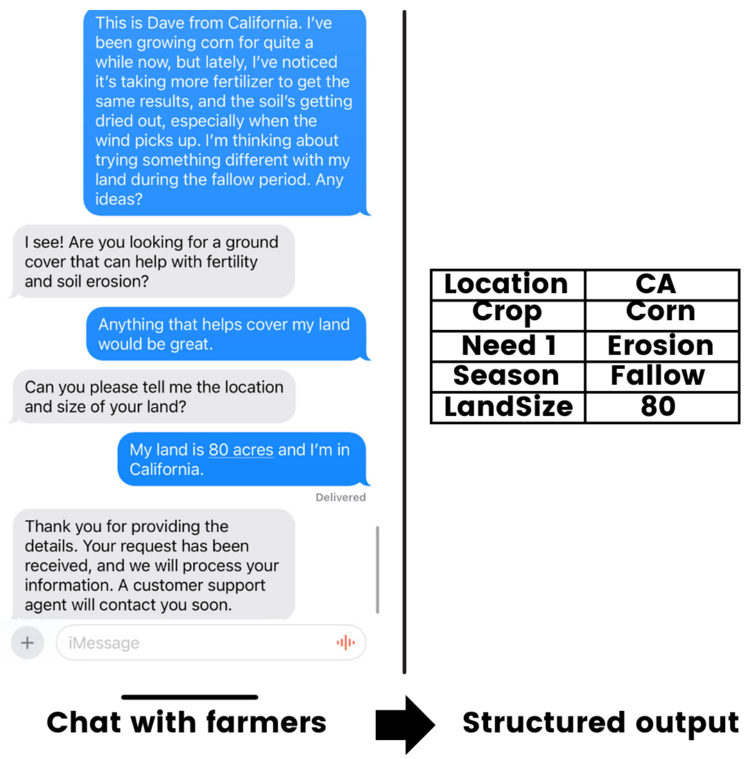

Figure 8: Text message to LLM example.

3. Fine-Tune on Dialogue-Specific Datasets. Our model will be fine-tuned on dialogue-specific datasets to enhance its conversational abilities. Our team plans to employ the dataset *DailyDialog* [40], which contains high-quality, multi-turn dialogues covering various conversational topics. If more data is required then our team will also include data from the *Persona-Chat* [41] dataset, where conversations include persona-based information. Fine-tuning will be done using our own data collection from users using supervised learning by training the model to predict the next conversational turn. This ensures that our model can handle back-and-forth interactions, maintain context, and provide natural-sounding responses during dialogue. The *Cross-Entropy Loss* function will be used to optimize the dialogue model.

4. Context Management. To handle multi-turn conversations effectively, we will implement context management to track the dialogue history across multiple turns. We will use a sliding window, where the model retains the last several turns (e.g., 5-10) of the conversation as part of its input to ensure continuity. This enables the model to understand references made earlier in the conversation. In addition, a long-term memory module will be implemented using *attention mechanisms* [42] to store key information from previous interactions. The use of GPT's transformer architecture allows for efficient context management by attending to all previous tokens in the input sequence. This ensures that important details from earlier turns are not lost.

Fine-Tuning

1. Fine-Tune for Agriculture-Specific Terms. We will fine-tune the GPT model on agriculture-specific datasets to make the model proficient in an agricultural context by following these steps: 1) **Dataset Selection—**we will curate datasets from sources like USDA publications, scientific papers (e.g., from *AGRICOLA* [43], *PubAg* [44], *Web of Science* [45]), and agricultural reports; 2) **Model Adaptation—**the fine-tuning process will involve additional training with supervised learning on these domain-specific datasets; 3) **Fine-Tuning for Specific Terminologies—**we will apply *transfer learning* [37] to ensure the model builds on its pre-existing language skills but adapts specifically to agriculture. As the model is fine-tuned, it will adjust its vocabulary embeddings to understand terms like "crop yield," "fertilizer subsidy," and "irrigation efficiency," to name a few.

2. Fine-Tune for Structured Output. We will fine-tune our LLM to generate structured outputs that PWDs can easily interpret, using training data that pairs unstructured inputs (e.g., free-text from farmers) with structured outputs in JSON format. By leveraging a sequence-to-sequence learning approach and attention mechanisms within the transformer model, the LLM will learn to map unstructured text into structured formats using custom templates. The model's performance will be evaluated using accuracy metrics, focusing on the correct structuring of output fields.

Objective #3: End-user outreach and partnership expansion. For this objective, our team will engage our rural farmer outreach facilitators to test our prototype. We will involve our assistive technology advisory board to assess progress and identify new partnerships required for full system development and implementation in Phase 2.

Task #3-1: Prototype deployment and result analysis. As part of our challenges and opportunities discovery process, we will implement a series of prototypes. These designs will be created using our fast prototyping platform developed in Objective 2. We aim to interact with the end users and test initial hypotheses to validate or refine them.

To better understand the activities occurring on rural farms and gather data from our first minimum viable prototype, our team will return to agricultural locations identified in Task #1-1 to evaluate and validate the prototype FarmAssist v1. We propose a one-day field study at each location, in which we will interview farm owners and managers to validate FarmAssist's features, assessing how they integrate into daily routines, funding application processes, workforce inclusion, and sustainability practices. We will test our AI-driven funding assistance model, using real-time data from farmers' operations and management records to identify key areas where funding is most needed. Through this activity will validate our proposed solutions with farmers and PWD teleworkers to ensure that FarmAssist addresses their most pressing needs.

To expand our outreach and showcase the FarmAssist v1 prototype, we will target events such as **field days**, **workshops**, **conferences**, and **training sessions** focused on small and mid-size farmers and PWDs in rural communities in California, South Dakota, and Texas. In California, we will participate in the Small Farm Conference and will utilize the UCANR Small Farms Network to attend related field days and workshops [46]. In Texas we will participate in the Southern Family Farmers and Food Systems Conference and South Dakota the Local Foods Conference and the South Dakota Soil Health Conference. These events will allow us to test our AI-driven funding model with real-time farm data and validate its effectiveness with farmers and PWD teleworkers.

Throughout Phase 1, and more critically in Phase 2, we aim to combine scientific R&D in AI with a build-measure-learn feedback loop to provide engineering solutions tailored to the agricultural sector. We will expand our findings into the system by refining and testing prototypes through a lean methodology approach. The lessons learned in each iteration of prototypes in Phase 1 will be crucial in advancing the end-to-end FarmAssist platform in Phase 2, focusing on features most requested by users during our customer discovery process.

Task #3-2: Engage advisory board and expand project partnerships The planned engagements with the advisory board, shown in Fig. 5, include: 1) Identifying product-market fit challenges and exchanging expertise on overcoming these difficulties; 2) Addressing issues related to non-disclosure and intellectual property agreements before the start of Phase 2; 3) Acting as a point of contact to connect with researchers, crop advisors, agricultural extension agents, farmers, and educators to find additional Phase 2 participants and partners; 4) Synthesizing discussions from our meetings into a "knowledge repository" to guide future project development and dissemination.

Finally, our Phase 1 project will take eight months to complete, and Table 1 provides a timeline.

Deliverables

Deliverable #1: Use-cases and technology requirements. This report will summarize findings from interviews and observations with farmers and PWDs.

Deliverable #2: Fast prototyping platform for product development. This platform will facilitate rapid development and testing of assistive technologies for use in the agricultural sector.

Deliverable #3: Final Report and roadmap for Phase 2. Our team will provide a detailed breakdown of Phase 1 activities and achievements and planned tasks for Phase 2, guiding the project's continuous development.

Research Objectives	Tasks	Team	M1	M2	M3	M4	M5	M6	M7	M8
O1: Use cases and technology requirement identification	Team working meetings	AC, DZ								
	On-site visit and preliminary data acquisition	AC			D1					
O2:: Platform formation for user-oriented fast prototyping	NLP to collect and curate funding opportunities	DZ, AC								
	Leveraging LLM to evaluate farmer's funding needs	AC						D2		
O3:: End-user outreach and partnership expansion	Prototype deployment and result analysis	AC								
	Engage advisory board and expand project partnerships	AC, DZ								D3

Legend:

- ☐ **AC**: PI Alfredo Costilla Reyes.
- ☐ **DZ**: Dr. Diana Zapata Rojas.

Deliverables:

- **D1** Identify use-cases and technology requirements
- **D2** Fast prototyping platform for product development
- **D3** Final Report and Roadmap for Phase 2

Table 1: FarmAssist Project Timeline (March 1, 2025 - October 31, 2025)

7 Related Research or Research and Development

Our women-owned and Hispanic-led Small Business Concern (SBC) leverages a solid foundation of prior research and development in AI, positioning us to bring meaningful innovation to the market. **PI Dr. Alfredo Costilla Reyes** has a proven track record in applying AI technologies to a variety of fields, including NLP for biomedical applications [29], which earned the *Best Demo Paper Honorable Mention at the 2023 Conference on Information and Knowledge Management (CIKM)*. His diverse portfolio also includes pioneering work in developing LLMs to combat cyberbullying [47], utilizing big data from Internet of Things (IoT) devices in agriculture [48], and an ML architecture to process large-scale amounts of video [49]. Moreover, his cutting-edge research on machine learning for edge devices was awarded the prestigious 2022 CIKM Best Paper Award, highlighting its significant contributions to the field [50].

Additionally, Dr. Costilla's successful completion of NSF SBIR Phase 1 [51] and Phase 2 awards [52] not only demonstrates the team's ability to manage complex research projects but also proves their capability to develop market-ready products that address real-world challenges.

Dr. Diana Zapata Rojas has an extensive expertise in soil science, biogeochemistry, sustainable agriculture, and nature-based solutions. She has conducted research on the effects of conservation practices such as cover cropping and tillage on greenhouse gas emissions and nutrient balance in cropping systems [53]. Her expertise also extends to developing decision support systems for predicting phenological stages in winegrapes, which has aided farmers in planning crop management and site selection [54, 55]. Dr. Zapata will lead the development of decision-support tools for the FarmAssist platform. Throughout her career, she has led efforts to implement demonstration projects and conduct outreach activities in both Texas and California, showcasing the benefits and trade-offs of conservation practices [56, 57].

These prior achievements underscore the team's ability to transform deep technology research into commercially viable and profitable innovations. Building on this success, the proposed research integrates advanced AI technologies into the agricultural sector, leveraging the knowledge and insights gained from our previous work. Figure 5 highlights the external resources and partnerships we have secured to support this ambitious project, demonstrating our capacity to attract the necessary expertise and collaborations. By applying proven AI techniques, FarmAssist aims to deliver a groundbreaking solution for farmers, ensuring both technical and commercial success.

8 The Market Opportunity

Although there are nearly 2 million rural farmers in the U.S. [58], many struggle to access government programs like EQIP, navigate complex grant applications, and implement conservation techniques [59]. This segment, which generated over $314 billion in agricultural outputs in 2022 [58], constitutes the **Total Addressable Market** for FarmAssist. In California alone, the USDA National Agricultural Statistics Service reports 43,265 small farms [60], representing our **Serviceable Addressable Market**. More specifically, in the Sacramento Valley, over 11,000 small family farms contribute nearly $4.5 billion annually to the state's economy [61], forming the **Serviceable Obtainable Market** for FarmAssist. Most notably, our team has secured our first **early adopters**, Michele Santangelo and Joe Turkovich, two rural farmers from the Sacramento Valley, CA; they have agreed to use FarmAssist's as beta testers during this Phase 1 work.

FarmAssist's **market opportunity has been validated** through extensive stakeholder engagement, including over 61 interviews with rural farmers across rural and tribal communities in South Carolina, Texas, and California. These discussions highlighted a consistent need for accessible funding and technical support. A particular testimonial from Mr. Schoch who is a Tribal Program Manager for AgrAbility & Federally Recognized Tribal Extension Program programs at South Dakota State University Extension, mentioned the following *"The FarmAssist proposal offers a much-needed solution for farmers, particularly those in rural and underserved communities. Drawing from my experience with the AgrAbility Beginning Farmer Project, I can attest to the importance of solutions like FarmAssist in helping farmers access new sources of funding, which is often one of the biggest hurdles they face."* (See letter of support 1, Jason Schoch).

Our customer base and our go-to-market strategy are built around leveraging significant social capital, a key challenge we identified during our customer discovery process. Direct outreach to rural communities can be difficult, requiring a strategic approach. By partnering with university extension departments, we significantly improved our ability to connect with rural farmers. Moving forward, we will continue to utilize this approach through key collaborations with **Jason Schoch**, Program Manager for a tribal extension program in South Dakota; **Dr. Nithya Rajan**, a professor at Texas A&M University, who engages farmers in Texas; and **Michele Santangelo and Joe Turkovich**, who are our point of contact with the farming community in Northern California. These partnerships have not only expanded our customer base but also strengthened the overall effectiveness of our go-to-market strategy.

The competitive landscape for FarmAssist includes a mix of traditional and tech-based solutions, as shown in Table 2. Platforms like Bushel Farm [62] focus on operational management and not directly help rural farmers benefit from local, state, and federal programs. On the other hand ISAP's FIND Tool [16], which provides conservation funding information in Illinois, and Extension Services, despite being a strong resource for regional assistance, often require significant resources, are geographically limited, and are not equipped to scale their solutions efficiently across different regions. FarmAssist, in contrast, offers an AI+human-driven platform that provides streamlined, direct funding assistance while promoting workforce inclusion for PWDs, addressing gaps in both efficiency and scalability that other solutions leave unaddressed. This combination makes FarmAssist a compelling alternative in the agricultural technology landscape.

Key Factors	FarmAssist	Bushel Farm	ISAP - FIND Tool	Extension Services
Type	AI-Driven Platform	Farm Management Software	Financial Incentive Database	Advisory Services
Focus	Funding access, PWD workforce inclusion	Operational management	Conservation funding for sustainable practices	Technical support, grant assistance
Target Audience	Rural farmers	Farmers of all sizes	Illinois farmers seeking financial incentives for conservation	Regional farmers
Key Strengths	Tailored funding support, PWD telework opportunities, AI-driven platform	Data-driven insights, large user base	Centralized financial incentive comparison, easy access to funding programs	Regionally tailored support, trusted institutions
Key Weaknesses	New entrant, needs market penetration	Does not offer funding assistance	Focuses only on Illinois, limited to conservation funding	Limited funding focus, resource-intensive

Table 2: Competition Analysis Table: FarmAssist vs. Alternatives

Business Model. We have engaged Blair Kirchner, an investor with over 20 years of experience in agriculture and food sectors, to join our business development advisory panel. Blair, a key figure at the Kirchner Group, has guided numerous companies through business discovery and market strategy. Together with Blair, we have identified two potential business models for FarmAssist: a **pay-per-successful application** model, where farmers pay only when they secure financial support, and an alternative **freemium model**, offering basic platform features for free with premium options available for a fee. Each model has its advantages, with the pay-per-application providing clear monetization and the freemium model offering broader access while requiring a critical user base for profitability.

Our **Commercialization Approach** will begin with pilot programs in California and will expand to Texas and South Dakota, our key agricultural regions. *Scaling the platform* will involve collaborations with agricultural technology providers and sustainability consultants, allowing us to expand into additional regions. The *economic benefits* of FarmAssist are considerable: farmers will gain greater access to funding, securing more financial support, which will drive the broader adoption of sustainable farming practices. Moreover, *PWDs* will benefit from meaningful telework opportunities, promoting social inclusion. We project FarmAssist's revenue to reach **$1 million annually** within five years, based on a *pay-per-financial-support-received* model, where 10,000 farmers pay an average of less than $9 a month to benefit from the platform.

References

[1] G. Deboy, Paul J. Jones, W. Field, J. M. Metcalf, and R. Tormoehlen. Estimating the prevalence of disability within the u.s. farm and ranch population. *Journal of Agromedicine*, 13:175 – 190, 2008.

[2] Cristina D Miller and Robert A Aherin. The prevalence of disabilities in the us farm population. *Journal of Agricultural Safety and Health*, 24(4):243, 2018.

[3] M. F. Strong and Vida Maralani. Farmworkers and disability: results of a national survey. *Journal of Vocational Rehabilitation*, 12:45–57, 1999.

[4] K. Martinez. Integrated employment, employmentfirst, and u.s. federal policy. *Journal of Vocational Rehabilitation*, 38:165–168, 2013.

[5] Sandy Wong, S. McLafferty, A. Planey, and V. Preston. Disability, wages, and commuting in new york. *Journal of Transport Geography*, 87:102818 – 102818, 2020.

[6] F. Michel, Martino Trevisan, Danilo Giordano, and O. Bonaventure. A first look at starlink performance. *Proceedings of the 22nd ACM Internet Measurement Conference*, 2022.

[7] Margaret N Friesen, Olga Krassikouva-Enns, Laurie Ringaert, and H. Isfeld. Farming with a disability: Literature from a canadian perspective. *Journal of Agromedicine*, 15:47 – 53, 2010.

[8] S. Whelan, D. Ruane, J. McNamara, A. Kinsella, and A. Mcnamara. Disability on irish farms—a real concern. *Journal of Agromedicine*, 14:157 – 163, 2009.

[9] Economic Research Service United States Department of Agriculture. Usda forecasts conservation program payments in 2024 to increase 10 percent, 2024. `https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=109005`. Accessed: 2024-09-03.

[10] Nicholas P Webb, Nadine A Marshall, Lindsay C Stringer, Mark S Reed, Adrian Chappell, and Jeffrey E Herrick. Land degradation and climate change: building climate resilience in agriculture. *Frontiers in Ecology and the Environment*, 15(8):450–459, 2017.

[11] Nathalie Seddon, Alexandre Chausson, Pam Berry, Cécile AJ Girardin, Alison Smith, and Beth Turner. Understanding the value and limits of nature-based solutions to climate change and other global challenges. *Philosophical Transactions of the Royal Society B*, 375(1794):20190120, 2020.

[12] American Farmland Trust. Transition 2021 recommendations, 2021. Accessed: 2024-09-08.

[13] Keith Campbell Foundation for the Environment. Expanding and increasing conservation technical assistance to america's working lands producers, 2023. https://www.agripulse.com/ext/resources/2023/02/28/1523Full-Assessmement.pdf Available at: https://perma.cc/VKY2-8FLT.

[14] Carsten Nesshöver, Timo Assmuth, Katherine N Irvine, Graciela M Rusch, Kerry A Waylen, Ben Delbaere, Dagmar Haase, Lawrence Jones-Walters, Hans Keune, Eszter Kovacs, et al. The science, policy and practice of nature-based solutions: An interdisciplinary perspective. *Science of the total environment*, 579:1215–1227, 2017.

[15] American Farmland Trust. Improving conservation programs: Insights for making programs work better for farmers and the land, 2023. https://farmland.org/wp-content/uploads/2023/06/Improving-Conservation-Programs-White-Paper.pdf. Accessed: 2024-09-02.

[16] Illinois Sustainable Ag Partnership. Tool website: Financial incentives database (find) tool, 2024. https://ilsustainableag.org/findtool/. Accessed: 2023-09-14.

[17] Saed Rezayi, Zheng Liu, Zihao Wu, Chandra Dhakal, Bao Ge, Haixing Dai, Gengchen Mai, Ninghao Liu, Chen Zhen, Tianming Liu, and Sheng Li. Exploring new frontiers in agricultural nlp: Investigating the potential of large language models for food applications. *ArXiv*, abs/2306.11892, 2023.

[18] Neha Kaushik and N. Chatterjee. A practical approach for term and relationship extraction for automatic ontology creation from agricultural text. *2016 International Conference on Information Technology (ICIT)*, pages 241–247, 2016.

[19] C. Parr, D. Lemay, Christopher L. Owen, M. Woodward-Greene, and Jiayang Sun. Multimodal ai to improve agriculture. *IT Professional*, 23:53–57, 2021.

[20] Hongchen Qin and Yiheng Yao. Agriculture knowledge graph construction and application. *Journal of Physics: Conference Series*, 1756, 2021.

[21] TianQing Yang, Fangju Ran, Mengyao Lu, and Jingzong Yang. Research on agricultural data mining model based on knowledge graph. In *International Conference on Electronic Information Engineering, Big Data, and Computer Technology (EIBDCT 2022)*, volume 12256, pages 787–791. SPIE, 2022.

[22] G. Vo and R. Lai. A method for simplifying the submission of an online request for an e-government service. *2014 International Symposium on Technology Management and Emerging Technologies*, pages 377–382, 2014.

[23] Bipartisan Policy Center. Technical issues white paper, 2021. https://bipartisanpolicy.org/download/?file=/wp-content/uploads/2021/11/BPC_Technical-Issues-White-Paper_Final-2.pdf. Accessed: 2024-08-31.

[24] U.S. Department of Agriculture. Usda fy 2022-2026 strategic plan, 2022. https://www.usda.gov/sites/default/files/documents/usda-fy-2022-2026-strategic-plan.pdf. Accessed: 2024-09-02.

[25] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. Autogen: Enabling next-gen llm applications via multi-agent conversation framework. In *COLM*, 2024.

[26] Shaokun Zhang, Jieyu Zhang, Jiale Liu, Linxin Song, Chi Wang, Ranjay Krishna, and Qingyun Wu. Training language model agents without modifying language models. *ICML'24*, 2024.

[27] Victor Dibia, Jingya Chen, Gagan Bansal, Suff Syed, Adam Fourney, Erkang (Eric) Zhu, Chi Wang, and Saleema Amershi. Autogen studio: A no-code developer tool for building and debugging multi-agent systems. In *Pre-Print*, 2024.

[28] Yiran Wu, Tianwei Yue, Shaokun Zhang, Chi Wang, and Qingyun Wu. Stateflow: Enhancing llm task-solving through state-driven workflows. *arXiv preprint arXiv:2403.11322*, 2024.

[29] Yi-Ning Chuang, Guangrun Wang, Chih-Yuan Chang, Kang-Hsiang Lai, Dongpu Zha, Rui Tang, and Xia Hu. Discoverpath: A knowledge refinement and retrieval system for interdisciplinarity on biomedical research. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 5021–5025. ACM, 2023.

[30] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'Amato, Gerard De Melo, Claudio Gutierrez, and Antoine Zimmermann. Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4):1–37, 2021.

[31] Scrapy. Scrapy, a fast high-level web crawling & scraping framework for python, n.d. Retrieved September 11, 2024, from https://github.com/scrapy/scrapy.

[32] Leonard Richardson. Beautifulsoup documentation, 2007.

[33] Tabula. Tabula: A tool for liberating data tables trapped inside pdf files, n.d. Retrieved September 11, 2024, from `https://github.com/tabulapdf/tabula`.

[34] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

[35] Bodhisattwa Srinivasa-Desikan. *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras*. Packt Publishing Ltd, 2018.

[36] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. 2018.

[37] Tom B. Brown et al. Language models are few-shot learners, 2020. arXiv preprint arXiv:2005.14165.

[38] Common Crawl. Common crawl, n.d. Retrieved September 11, 2024, from `https://github.com/commoncrawl`.

[39] Yukun Zhu, Ethan Park, Phillip Isola, and Alexei A Efros. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books, 2015. arXiv preprint arXiv:1506.06724.

[40] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. Dailydialog: A manually labelled multi-turn dialogue dataset, 2017. arXiv preprint arXiv:1710.03957.

[41] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. Personalizing dialogue agents: I have a dog, do you have pets too, 2018. arXiv preprint arXiv:1801.07243.

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

[43] AGRICOLA. Agricola database, n.d. Retrieved September 11, 2024, from `https://agricola.nal.usda.gov/`.

[44] PubAg. Pubag database, n.d. Retrieved September 11, 2024, from `https://pubag.nal.usda.gov/`.

[45] Clarivate Analytics. Web of science. https://www.webofscience.com.Accessed: 2024-09-04.

[46] University of California Agriculture and Natural Resources (UC ANR). Uc anr small farms network - small farms research and extension. https://sfp.ucanr.edu/. Accessed: 2024-09-08.

[47] Andrew T Lian, Alfredo Costilla Reyes, and Xia Hu. Captain: an ai-based chatbot for cyberbullying prevention and intervention. In *International Conference on Human-Computer Interaction*, pages 98–107. Springer, 2023.

[48] Alfredo Costilla Reyes. *Energy Harvesting Systems for the Internet of Things with Applications to Smart Agriculture*. PhD thesis, Texas A&M University, 2020. https://hdl.handle.net/1969.1/191606, Accessed: 2023-09-15.

[49] Daochen Zha, Zaid Pervaiz Bhat, Yi-Wei Chen, Yicheng Wang, Sirui Ding, Jiaben Chen, Kwei-Herng Lai, Anmoll Kumar Jain, Mohammad Qazim Bhat, Na Zou, et al. Autovideo: An automated video action recognition system. *arXiv preprint arXiv:2108.04212*, 2021.

[50] Guanchu Wang, Zaid Pervaiz Bhat, Zhimeng Jiang, Yi-Wei Chen, Daochen Zha, Alfredo Costilla Reyes, Afshin Niktash, Gorkem Ulkar, Erman Okman, and Xia Hu. Bed: A real-time object detection system for edge devices. *arXiv preprint arXiv:2202.07503*, 2022.

[51] National Science Foundation. SBIR Phase I: A Hardware-Aware AutoML Platform for Resource-Constrained Devices, 2021. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2136679&HistoricalAwards=false. Award Abstract #2136679.

[52] National Science Foundation. SBIR Phase II: A Hardware-Aware AutoML Platform for Resource-Constrained Devices, 2023. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2335642&HistoricalAwards=false. Award Abstract #2335642.

[53] Diana Zapata, Nithya Rajan, Jake Mowrer, Kenneth Casey, Ronnie Schnell, and Frank Hons. Long-term tillage effect on with-in season variations in soil conditions and respiration from dryland winter wheat and soybean cropping systems. *Scientific Reports*, 11(1):2344, 2021.

[54] D Zapata, M Salazar, B Chaves, Markus Keller, and G Hoogenboom. Estimation of the base temperature and growth phase duration in terms of thermal time for four grapevine cultivars. *International journal of biometeorology*, 59:1771–1781, 2015.

[55] Diana Zapata, Melba Salazar-Gutierrez, Bernardo Chaves, Markus Keller, and Gerrit Hoogenboom. Predicting key phenological stages for 17 grapevine cultivars (vitis vinifera l.). *American Journal of Enology and Viticulture*, 68(1):60–72, 2017.

[56] Community Alliance with Family Farmers. Bios field day at river garden farms. https://caff.org/bios-field-day-river-garden-farms/. Accessed: 2024-09-05.

[57] Texas A&M University Department of Soil and Crop Sciences. Conservation and sustainable agriculture highlighted at field tour. https://soilcrop.tamu.edu/department-updates/conservation-and-sustainable-agriculture-highlighted-at-field-tour/. Accessed: 2024-09-05.

[58] United States Department of Agriculture (USDA). 2022 census of agriculture - united states data, 2022. Accessed: 2024-10-14.

[59] A. Reimer, B. Gramig, and L. Prokopy. Farmers and conservation programs: Explaining differences in environmental quality incentives program applications between states. *Journal of Soil and Water Conservation*, 68:110 – 119, 2013.

[60] United States Department of Agriculture (USDA). 2022 census of agriculture - california state and country data, 2022. https://www.nass.usda.gov/Publications/AgCensus/2022/Full_Report/Volume_1,_Chapter_1_State_Level/California/cav1.pdf. Accessed: 2024-10-14.

[61] California Department of Food and Agriculture. California agricultural statistics review 2013 – 2014. https://www.cdfa.ca.gov/statistics/. Accessed: 2024-09-14.

[62] PitchBook. Company profile: Bushel farm, 2024. https://my.pitchbook.com/profile/56014-12/company/profile#insights. Accessed: 2023-09-14.