

Virtual Agent Chatbot

# Proof of Value (POV) Study

Last Updated: July 6, 2021

## Table of Contents

<b>Executive Summary</b>	<b>3</b>
<i>Process Highlights</i>	3
<i>An Agile Process</i>	3
<i>Proof of Value</i>	3
<i>Main Research Effort: Controlled Study of Proof of Value</i>	4
Top Project Limitations	6
<i>Top Challenge: Content</i>	6
<i>Study Challenges</i>	6
<i>Conclusion &amp; Recommendations</i>	6
<i>Next Steps: Authenticated Proof of Value</i>	7
<i>Ethics &amp; Data Collection</i>	7
<i>Expanded Inclusive Research</i>	7
<i>Explore Disambiguation</i>	7
<i>Potential connection with Live Agent</i>	8
<i>Collaboration with lines of businesses</i>	8
<i>Content</i>	8
<b>Market Fit</b>	<b>9</b>
<i>Value of a Chatbot</i>	9
<i>Knowledge within a Chatbot</i>	9
<i>Chatbot Proof of Value Study</i>	10
<i>Research</i>	11
<i>Branding: Voice &amp; Tone (March, June)</i>	11
<i>OCTO-DE Self-Service Feature Survey (June)</i>	12
<i>Unmoderated Controlled Study (June)</i>	13
<i>Moderated Inclusive Design Study (June)</i>	13
<i>Market Fit Recommendations and Next Steps</i>	15

<b>Product Fit - Microsoft Power Virtual Agent</b>	<b>16</b>
<i>Chatbot Features Developed Within PVA</i>	16
<i>PVA Findings</i>	17
<b>Product Fit - Other Virtual Agent Frameworks</b>	<b>17</b>
<i>Product Fit Recommendations and Next Steps</i>	19
<b>Creating &amp; Maintaining Content</b>	<b>20</b>
<i>Initial Assumptions &amp; Actions</i>	20
<i>Content &amp; Proof of Value</i>	21
<i>VA.gov &amp; Virtual Agent</i>	22
<i>Tone &amp; Voice</i>	22
<i>Emergency and Crisis</i>	23
<i>Content Recommendations and Next Steps</i>	24
<b>Challenges</b>	<b>25</b>
<i>Access to impactful reporting</i>	25
<i>Ability to fine-tune AI responses (e.g., confidence-score adjustments)</i>	25
<i>Level of debugging required</i>	26
<i>Scale to sophisticated responses (e.g., payments widgets, APIs)</i>	26
<i>Version control of mass content</i>	26
<i>PVA Limitations</i>	26
<i>Accessibility &amp; 508 Compliance</i>	26
<i>Accuracy</i>	28
<i>Feedback</i>	28
<i>Security &amp; Privacy</i>	28
<i>Usability</i>	29
<b>Next Steps</b>	<b>30</b>
<i>Authentication</i>	30
<i>Content</i>	30
<i>Ethics &amp; Data Decisions</i>	30
<i>Expanded Research</i>	31
<i>Government Community Cloud Implementation</i>	31
<i>Live Agent Connection</i>	31
<i>Mobile-First</i>	31
<i>Platform Placement</i>	31
<i>System Design</i>	32
<i>Veteran-First Process</i>	32
<b>Conclusion</b>	<b>32</b>

## Executive Summary

In Q4 of 2020, the Veteran's Experience Office (VEO) in collaboration with the Office of the CTO Digital Experience (OCTO-DE), began exploring how virtual agents can help Veterans self-serve. The virtual agent is one component of a Multichannel Technologies (MCT) Omnichannel modernization initiative to provide Veterans with seamless access to information and support.

## Process Highlights

### An Agile Process

The virtual agent must be produced using an agile framework. This channel requires constant evaluation, iteration, and maintenance. We cannot adequately train and validate chatbot features using waterfall methods and requirements-gathering processes.

### Proof of Value

VA lines of businesses (e.g. VBA) want virtual agents to reduce call volume. Chatbots are better able to *enhance the quality* of calls, by increasing how often call center agents answer complex calls. However, to do this, chatbots must effectively compete with human contact by providing value to Veterans. Therefore, to kickstart this project, we made the following two assumptions:

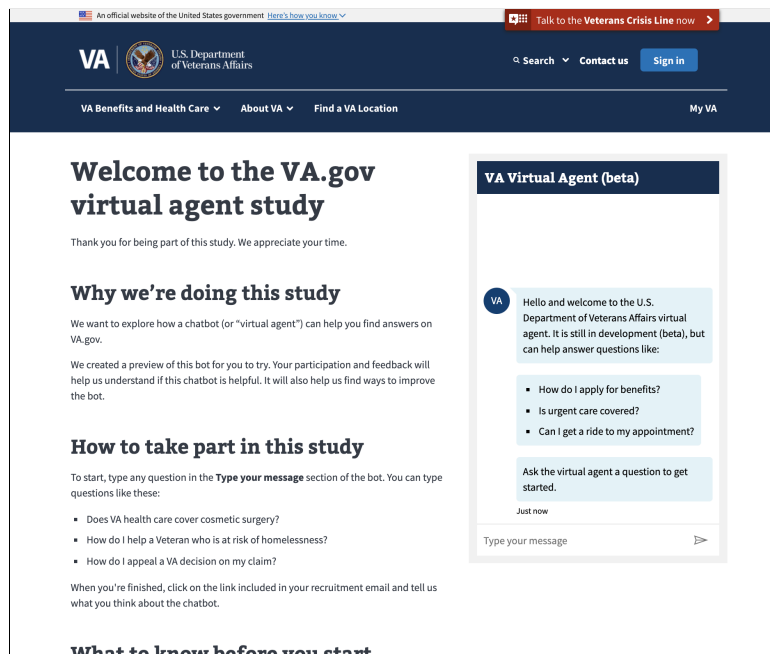
1. A virtual agent could provide value by:
  - a. Increasing awareness of existing VA self-service tools
  - b. Decreasing the time Veterans spend waiting for an outcome
  - c. Giving Veterans 24/7 access to either anonymous or secure support
2. For Veteran use, branding consistency, maintenance, funding and proper team allocation - there should only be **one centralized VA.gov virtual agent** (i.e. “superbot”) with features and capabilities that **scale over time according to what Veterans need**.

We built a proof of value virtual agent to validate these assumptions directly with Veterans, their families and caretakers. At minimum, the proof of value allowed our study participants to:

- See a welcome greeting with sample questions
- Get answers to *a sample* of health benefits and claims-related questions
- Find support if facing a crisis
- Be informed when the bot cannot answer a question
- Request to speak with a human (no direct connection to call centers)

We used two tools: Microsoft Power Virtual Agent (PVA) and QnA Maker, to build our proof of value chatbot. Both platforms belong under the Microsoft Dynamics 365 suite. All projects under the Omnichannel Modernization Initiative are using Dynamics 365 to explore individual projects.

The Virtual Agent team did not choose Microsoft at the start of this exploration. Instead, we evaluated how well Microsoft Power Virtual Agent performed as a viable platform for the virtual agent.



VA Virtual Agent POV hosted on <http://staging.va.gov/virtual-agent-study/>

## Main Research Effort: Controlled Study of Proof of Value

A total of 44 (out of 100 recruited) participants tested the proof of value. We recruited participants in segments of 10 with a focus on breadth and diversity - including women and marginalized populations (e.g., LGBTQ, rural, economic insecurity). Not all recruited segments produced participants.

Participants were given a direct link to the virtual agent and were allowed to engage with the bot unmoderated. This method gave participants the most realistic experience and provided us with more authentic feedback. All participants shared their feedback with us through an Optimal Sort survey.

The team analyzed conversational logs (what was typed to the bot) and survey responses, finding the following:

- Overall, participants indicated a willingness not only to try the chatbot but to use it again, and seemed excited about the product, how it could evolve and how it might help them and VA.
- Participants understood the chatbot wasn't human but still felt they could trust it.

### Veteran Voices

" I understand chatboxes have limited responses, as long as it points me in the right direction then that's fine. "

" I think it was quite excellent. The links were right on point and fit the answer perfectly! I was very impressed! "

" I liked the answers I was given - I feel like if I have a very focused question I would probably be more comfortable calling on the phone and talking to someone in person, but to find out quick and general info I would use the chatbot. "

" ...it can help if I have questions at hours that the VA is not available. "

- Participants reacted positively to the personality of the chatbot in terms of voice and tone.
- Chatbot responses were rated as fairly accurate, although the level of accuracy left some participants feeling frustrated.
- Participants experience challenges with current VA customer support channels, and a VA chatbot was viewed as an opportunity to avoid those challenges.
- Participants want personalized content (authenticated). Topics such as appointment management, debt management, general benefits eligibility, and GI Bill-related issues were popular reasons why participants call VA today.
- The conversational logs indicted a wide range of topic areas, with the top three being:
  1. VA programs and eligibility for them (36 separate engagements or instances someone asked this into the Virtual Agent)
  2. Misc. medical topics and conditions (26 engagements)
  3. Disability related (17 separate engagements)
- There were numerous other topics, including appointments, claim status, education related, VA touchpoints (locations, phone numbers), topics related to marginalized populations and others.
- The process of analyzing topics was a challenge. The team identified metrics in the product outline, and were unable to either find the data in conversational logs or the data itself required a large amount of data scrubbing and a manual process working outside of PVA to make sense of the data. We exported logs at the end of the study from PVA, imported them into Google Sheets, highlighting user content in red (within both Bot and User in the same cell in one large paragraph), then manually identify topics, counting and synthesizing in Mural.
- Additional metrics captured included:
  - # of Engagements - there were 379 total engagements (i.e. times a person entered something into the chatbot window, or an average of 8 engagements per session per user)
  - Bot Response Correctness. Did the bot surface the correct response based on data we have? 27 yeses, 26 varied (a combination of yes and no responses) 6 no
  - Response Quality. Can the user take action? 192 engagements had good responses from the chatbot (the content was correct and a user could take an action via a link); 121 had bad responses (i.e providing an incorrect link for a user)
  - Other metrics around escalating to an agent, variation of questions asked, multiple questions in one session etc. were also analyzed and are available in our Github in the raw data spreadsheet

[View the full research report](#) on Github

## Top Project Limitations

We experienced the following limitations that impacted our ability to scale the study to a wider audience (i.e., 5% of VA.gov visitors):

- No Government environments or Microsoft licenses. The proof of value was built on commercial licenses with possible security and privacy concerns. Due to licensing agreements, we could not “launch” the virtual agent in production environments without potentially incurring a cost-per-bot usage.
- Section 508 Accessibility concerns. Through participation with OCTO-DE 508 subject matter experts, we identified numerous 508-infringements. We validated some of these issues in a small study with visually impaired Veterans. We submitted a list of findings to Microsoft.

## Top Challenge: Content

We attempted to automate bot content by pulling content directly from VA.gov. However, this resulted in excessive manual edits. As a result, we drafted every piece of content. Chatbots depend on rapid development (and approval) of content. Existing content evaluation processes require evaluation in order to reduce bottlenecks.

## Study Challenges

The following challenges impacted the thoroughness of the study:

- DNS issues with mobile users. During the study we discovered that many mobile users experienced DNS issues that were out of our control. This reduced our participation numbers.
- Low to no-participation from marginalized populations. Additional research is needed to ensure that populations who are experiencing economic insecurity, low-internet access, LGBTQ+ have a chance to shape the virtual agent. We understand that some of these participants may have had issues accessing the bot (DNS issues).
- Time. The virtual agent project was under a finite budget. We moved fast to contain the study.

## Conclusion & Recommendations

Our study shows that Veterans see value in a virtual agent at VA. Therefore, we recommend building a "super bot," a centralized VA.gov virtual agent with scalable features and capabilities that we can improve iteratively and *with* Veterans.

We **do not** recommend Microsoft Power Virtual Agent (PVA) as the platform for the virtual agent because:

1. The platform lacks adequate content-management controls (e.g., version control, topic-classification, search) to handle the massive amount of content at VA. We used Google Sheets and Github to manage our content.

2. Inability to modify keyword phrases for Escalation, Greeting, and other critical bot responses
3. Inability to change the layout of horizontal buttons to vertical buttons for improved scannability
4. Confusing error messages made it difficult for the development team to troubleshoot.
5. Inability to create rich media responses such as payment widgets without code
6. Inability to easily connect APIs without code
7. Inability to modify existing conversational rules and behaviors. The chatbot bot often "locked" users in decision trees.
8. Unreliable metrics reporting resulted in a lot of manual labor to digest and no access to crucial information (e.g., confidence scores)
9. Poor 508 compliance
10. Inability to adjust responses based on AI confidence levels
11. Inability to assign a response to a topic based on confidence scores
12. Difficulty retaining functions while importing and exporting between instances

While we investigated alternative platforms, at this time, **we recommend a fully customizable solution using Microsoft Bot Framework.** We understand that issues with 508 compliance will continue to require Microsoft oversight and resolve. However, using a customizable platform will allow us to control the product's user experience and overall feature scalability.

## Next Steps: Authenticated Proof of Value

The Virtual Agent team is currently preparing for internal meetings to start formalizing a Phase 2 Product Plan. Our research identified that Veterans and VA.gov OCTO-DE staff recommend personalized authenticated experiences (see the research section of this report). Therefore, we will focus our efforts on creating an authenticated Proof of Value to test with a percentage of VA.gov visitors in early 2022. While we will continue to finalize details, the team will also pursue the following:

### Ethics & Data Collection

We will collaborate with Veteran Crisis Line and others to further establish best practices and decisions on the subject of crisis management and privacy.

### Expanded Inclusive Research

Perform additional research with population segments such as homeless, food/housing insecurity, Veterans based on conflict-era, LGBTQ+, disabilities, and mental health.

### Explore Disambiguation

We will *explore* (not implement) conversational repartee - or the chatbot's ability to ask clarifying questions.

## Potential connection with Live Agent

The Virtual Agent team will work closely with the Live Agent timeline in order to assess the best time and conditions for the virtual agent to connect to live chat.

## Collaboration with lines of businesses

We will work with lines of businesses to assess the best training data for the virtual agent based on feasibility and Veteran-need - and to help them incorporate the virtual agent into their existing business processes (i.e., content creation).

## Content

The Virtual Agent team will collaborate with VA.gov content team (and others) to assess how we can improve existing content management systems in a way that can help us:

- Reduce manual content creation for the virtual agent
- Ensure that the virtual agent has the latest information
- Add version-control to the virtual agent content process

Furthermore, we will also evaluate the focus of Phase 2 content in a way that allows us to avoid bottlenecks in content creation.



## Market Fit

To begin addressing the topic of market fit, let's look at the most popular *business-driven* question we had to answer:

*Will the virtual agent reduce call center calls?*

VA lines of businesses (e.g., VBA) want virtual agents to reduce call volume. Chatbots are better able to *enhance the quality* of calls, by increasing how often call center agents answer complex calls. Any impact to call volume is highly dependent on whether or not people can **find what they are looking for on their own (value to Veterans)**.

## Value of a Chatbot

Chatbots must provide **value** to compete with humans (call centers). To define "value," we began looking at what a well-developed chatbot could potentially do for Veterans:

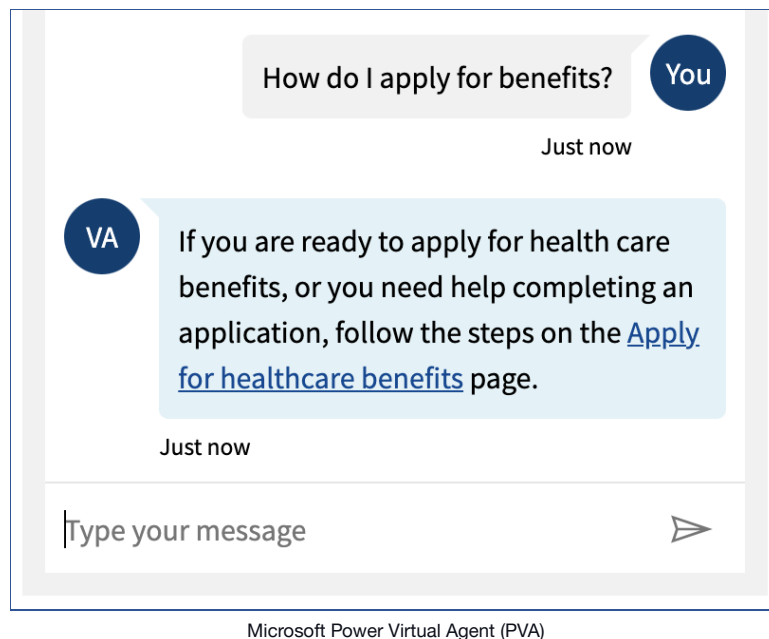
- Increase the awareness of existing self-service tools
- Decrease the time they spend waiting for an outcome
- Give them 24/7 access to either anonymous or secure support

## Knowledge within a Chatbot

Next, we considered what the bot *should* know. To start examining this, we built an unauthenticated Proof of Value (POV) bot to give study participants a preview of how the bot *could* behave--hopefully enticing them to ask questions that we would then log as future bot-training data.

Based on internal research (e.g., IRIS/Ask VA reports, VBA call center data), we chose health benefits and claims benefits content as a starting point for drafting chatbot dialogues. This allowed us to have focused conversations on the level of effort necessary to build (and maintain) a chatbot.

## Chatbot Proof of Value Study



The final chatbot proof of value (and the website it was located on) contained the following features:

- Open, free-text field to receive questions
- A greeting with suggested questions to start an interaction
- Approximately 100 unauthenticated bot-responses covering health benefits and claims
- Responses for questions the bot cannot answer
- Recognition and responses for Veterans experiencing a crisis (with some oversight from Dr. April Foreman)
- Ability to surface a phone number to reach a "human" (text-only)
- Study guidelines and information for those needing immediate help (with oversight from Patrick Bateman, Dr. April Foreman, and Danielle Thierry)

From an identification of value perspective, the proof of value had four main goals:

- Determine what information users were asking the bot by reviewing chatbot conversational logs and survey responses provided after chatbot usage
- Explore how our customers felt about voice, tone, and user experience of the bot
- Capture (directly through survey questions) if our customers thought a bot could be useful
- Understanding what features would make a bot valuable to them

From a team perspective, the proof of value had the following goals:

- Help us understand the limitations of Microsoft Power Virtual Agent (PVA) and associated tools (e.g., conversational sophistication, Section 508 compliance, reporting, scalability, version control)
- Define base level of effort managing a virtual agent (e.g., skills, team allocation, content management, legal, collaboration with legacy and future systems)
- Understand how it could impact other VA stakeholders (e.g., Veterans Crisis Line, LOBs, call centers)

[View the Product Outline](#) on Github

## Research

We conducted multiple research studies to validate our existing product vision and inform future product iterations, pivots, MVPs, and rollout plans.

### Branding: Voice & Tone (March, June)

We conducted a Branding Survey in March 2021 with OCTO-DE colleagues to discuss the chatbot's voice and tone. Key findings include:

- The OCTO-DE population we surveyed was split on whether or not to name the Virtual Agent. There were valid concerns and advantages for doing so mentioned in their comments.
- Those who wanted to name the Virtual Agent believed a recognizable, NATO alphabet, gender-neutral name would make the chatbot less cold and more personable.
- Those who didn't want to name the Virtual Agent expressed concerns that a name would not be diverse or inclusive enough, give the bot an unnecessary casual/friendly personality, and confuse users into thinking they were talking to a real human.

[View the OCTO-DE branding report](#) on Github

We also interviewed 16 Veterans and caregivers in April 2021 for their perspective on chatbot tone and personality. Key findings include:

- Veterans, care-takers and family members want communication between the VA and them to be personable, curious and empathetic.
- People have low expectations for a VA chatbot but are open to trying it and using it as their first point of contact.
- Naming may require a compromise. The VA can provide a name for context, like VA chatbot, so it is clear Veterans are not speaking with an actual person, but still be able to use a conversational tone that is not dry or bureaucratic.

Overall, we did not see significant reasons to build out chatbot branding and naming at this time. It would be beneficial to review this decision in the future when the chatbot is more sophisticated and able to answer questions about more complex topics on VA.gov.

[View the Veteran-facing branding research report](#) on Github

### OCTO-DE Self-Service Feature Survey (June)

We surveyed other OCTO-DE teams to understand what existing VA.gov infrastructure and teams could help improve the virtual agent's capabilities in the future, bringing the virtual agent closer to a true self-service virtual agent.

We received seven responses from various teams spanning Debt Resolution, Identity, Sitewide, Decision Tools and VAOS. The teams suggested the following:

Internal Teams Chatbot Ideas	Veteran Benefit
Appointment Scheduling options	Scheduling appointments for myself
Conversation History Logging	Having the VA understand all previous conversations I've had so I don't have to repeat myself
Chatbot-guided form completion	Pre-populating fields where the VA already has knowledge about me and presenting only the forms and instructions pertinent to me
Tier 1 benefit applications	
Self-service password reset	Lowering the turnaround time for passwords to be reset so users can have an authenticated experience
Sign in assistance	Content for helping users authenticate
Debt management/resolution topics	Content for debt assistance

The ideas generated by our internal consumers of a virtual agent all center around the idea of personalized experiences. These teams each wanted a way to connect their systems with a chatbot that knew who these users were in order to deliver content that was personalized to them. This strongly aligns with the expectations of participants in the Controlled Study. While we analyzed these options, for Phase 2, we suggest focusing on authenticated experiences (e.g. appointment management).

[View the OCTO-DE feature report](#) on Github

## Unmoderated Controlled Study (June)

On June 1st, we began recruitment toward a goal of 100 participants (split into population segments) to test the Proof of Value chatbot. We chose to run this test unmoderated so participants could more realistically experience the bot. Participants were recruited and given directions, a survey link to provide us with feedback, and a link to the chatbot hosted on the VA.gov staging environment.

We acquired conversational logs and survey responses that captured what Veterans want from a chatbot, frustrations, and general outlook through this method.

Highlights from the survey responses include 6 key findings:

- A: Accuracy:
  - The chatbot was rated as being fairly accurate, although the level of accuracy left some participants feeling frustrated.
- B: Trust:
  - Participants understood the chatbot wasn't a human but still felt they could trust it.
  - Participation from marginalized populations was low and understanding their needs will be key to ensuring trust is built into the heart of the product.
- C: Personality:
  - Participants reacted positively to the personality of the chatbot in terms of voice and tone.
- D: Product usage:
  - Participants experience challenges with current VA customer support channels, and a VA chatbot was viewed as an opportunity to avoid those challenges.
  - Overall users indicated a willingness to not only try the chatbot, but to use it again, and seemed excited about the product, how it could evolve and how it might help them and the VA.

[View the full research report](#) on Github

## Moderated Inclusive Design Study (June)

During development we encountered a number of Section 508 accessibility issues identified during Staging Review by the VA accessibility SME and by the Virtual Agent development team. We conducted interviews with blind Veterans to validate if the issues occurred and the impact on the user experience.

In order to prevent biases in our research, in June we conducted a Moderated Inclusive Design Study with three visually impaired Veterans to better understand:

- Background: Do partial vision or visually impaired/blind users have experience with chatbots
- Prevalence: Do they experience the challenges we've identified with the chatbot we built - see <https://github.com/departement-of-veterans-affairs/va.gov-team/issues/25113> for the list)
- Product use: If they wanted to use this chatbot, what would they use it for
- Satisfaction: Understand if participants would be deterred in using a chatbot because of the inaccessible aspect of the chatbot

Research study highlights included:

- Links in the chat log were not keyboard focused.
  - We validated that this issue occurred and that the links added to the VA chatbot's responses are not reachable by keyboard when navigating through the page with the TAB key.
  - This is a defect 1 status (i.e. the most severe, 'must fix' issue) according to Section 508, meaning it needs to be fixed before launching on va.gov
- Links were being announced as "messages" or just text and not links.
  - We validated that JAWS and VoiceOver are reading out links in the chatbot responses as "LINK TEXT, message" or just reading out the link text and never saying "link".
  - This is a defect 1 status (i.e. the most severe, 'must fix' issue) according to Section 508, meaning it needs to be fixed before launching on va.gov
- Many of the additional accessibility issues which were less urgent than defect 1 in severity were observed, including the challenge of finding chatbot on the page in the first place.
- Blind Veterans can differ in how they use technology, including assistive technology.
- Despite the accessibility challenges participants were eager to use the chatbot and indicated they would want the same feature set (such as authentication) as the non-blind population.

[View the inclusive design report](#) on Github

## Market Fit Recommendations and Next Steps

Our research has led us to these critical considerations.

### **Branding**

Veterans wanted a professional voice and tone but further research is required about naming the chatbot, especially when the virtual agent is able to provide more interactive experiences.

### **Desired Features**

The capabilities that OCTO-DE teams wanted to integrate with the chatbot strongly aligned with what Veterans expected a chatbot should be able to do. Namely, both groups were looking for personalized experiences--the VA teams to use their expertise to deliver them, and Veterans to use them to make their lives easier. Continue to engage with VA teams to look for integration touchpoints and talk to Veterans to see what they need.

### **Authentication**

- Include responses built with APIs, React components (e.g., payment widgets)
- Personalization - address the Veteran by name.
- Help with login/account creation.
- Securely transfer information from the virtual agent to the live agent.
- Securely store conversations for Veteran-retrieval.
- Secure login from unauthenticated experience to authenticated.

### **Disambiguation**

Pursue the ability for the virtual agent to ask clarifying questions to help the Veteran locate the information they need.

### **Additional Research**

Additional research into segments such as homeless, food/housing insecurity, Veterans based on conflict-era, LGBTQ+, disabilities, and mental health

# Product Fit - Microsoft Power Virtual Agent

## Chatbot Features Developed Within PVA

A critical requirement of our virtual agent research and experiment is assessing how well the proposed solution (i.e. Microsoft Power Virtual Agent) could help us build a virtual agent that helps Veterans (and their families) self-serve.

First, through conversations with VEO, we established that a centralized VA.gov-wide virtual agent (super bot) would provide a more robust and more reliable experience for Veterans instead of multiple bots. However, it also requires longer-term planning, a roadmap that captures iterative and scalable features over time (years), and thorough consideration of inter-system and agency impact (e.g., call centers).

Second, we considered two types of platform users: Content Managers & Engineers. At a high level, content managers need to quickly and easily create new content, run and assess reports for content quality, map existing content and maintain a system of content versions - *without a developer*. At a high level, engineers should identify and troubleshoot system issues, respond to 508 concerns, and quickly build rich features - without many platform constraints or excessive learning curves.

With these points in mind, we considered a couple of high-level assumptions on the topic of scalability of features and overall platform sophistication that include:

- Access to impactful, intuitive, and digestible reporting (e.g., conversational logs, response ranking by AI confidence)
- Ability to adjust bot responses based on conversational logs (e.g., dialogs, trigger words, response-confidence scores)
- Easily connect conversational topics without code
- Easily pull APIs without code
- Accessible and 508-compliant interactions
- Support for automated testing and deployments
- Easy user interface customizations (e.g., look and feel of buttons)
- Version control of mass content that could act as a *temporary* knowledge base\*
- Low-code ability to incorporate rich interactions (e.g., payment widgets)
- Easy to digest error-messages and the ability to update error-messages
- Ability to easily troubleshoot/debug (for devs)

Microsoft markets Power Virtual Agent (PVA) as a no-code solution. As a team, we considered the roles of Content Managers and Engineers when we set out to assess the capabilities of this tool.

Below are our findings.



## PVA Findings

Our goal is to build a centralized "super bot" that can surface knowledge from various VA sources, surface rich responses like payment widgets, and scale to sophisticated conversations. At this time, **we do not recommend** using Microsoft Power Virtual Agent (PVA) because:

1. VA.gov has a massive amount of content that requires version control, automated content updates, and involvement from multiple content creators. As a low-code tool, PVA cannot monitor/maintain large quantities of content. It makes it difficult to scan and map content topics and does not allow manual confidence-score adjustments.
2. Unreliable deployments. Importing and exporting is very error-prone and resulted in the team having to restructure the chatbot experience to contain the user experience. Export and import issues were not well-documented and gave us doubts about the strength of their import/export solution.
3. Actions such as changing Greeting phrases (System Topics) were locked to changes in the platform, making it difficult to impact the AI manually to improve user experience.
4. System error messages were unclear, making it difficult for us to track and troubleshoot an issue. It also does not allow us to change the phrasing of an error to make it easier for users to understand.
5. Inability to control how prompts behave, resulting in users getting stuck in decision trees and the bot being unable to surface other topics until a user leaves a prompt (i.e., clicks on a button)

## Product Fit - Other Virtual Agent Frameworks

During development, we encountered several limitations with the PVA framework that prevented us from developing features we had wanted to test with users. We were interested in understanding if these limitations were specific to PVA or characteristics of low-code/no-code frameworks.

We completed a high level competitive analysis of popular chatbot platforms to understand if and how they addressed the limitations in PVA. Our analysis was completed as 'technical spikes' which are specifically short, time-boxed Agile cards to answer a question or explore a functionality to understand how something works. We selected a number of dimensions on which to compare these frameworks, including accessibility, content management, and customization, as these were key requirements we needed in a chatbot.

At this time, we recommend Microsoft Bot Framework for a VA.gov virtual agent. Of all the platforms we evaluated, Microsoft Bot Framework (while not perfect) is the one best suited to meet our feature set. That being said, it is still important for us to complete a more comprehensive assessment of 508 requirements and issues in the tool as we prepare for Phase 2. The Microsoft Bot Framework development team has been responsive to 508 issues in the past, so we anticipate working with them closely to monitor the releases of more 508-compliant code in a timely manner.

Property	PVA	Microsoft Bot Framework	Rasa	Salesforce Einstein	Google Dialog Flow
Ease of hosting	✓	✓	✗	✓	✓
Frontend customizability	✓	✓	~	✗	✓
Frontend 508 compliance	~	~	✗	?	?
Content Management	✗	✓	✓	✗	✗
Frontend support & maintenance	✓	✓	✗	✓	✗
Testability	✗	✓	✓	~	~
Learning curve for developers	✓	✗	✓	~	✓
Understandable by non-developers	✓	✗	✓	~	✓
Salesforce Integrations	~	~	~	✓	~
Version Control	✗	✓	✓	✗	✗
Automatic Deployments	✗	✓	✓	✗	✓

Symbol	Explanation
✓	Framework provides this feature or is conducive to this cross-functional requirement (CFR)
✗	Framework does not support this functionality or CFR
~	Framework may support this feature or CFR but it is not a strong differentiating factor
?	Unknown if the framework supports

A more detailed assessment of frameworks could be done in the future if the business is interested. Please view supporting documentation on the [other frameworks](#) in our project repository.

## Product Fit Recommendations and Next Steps

### **PVA shortcomings**

The PVA product is still at an early stage in its product maturity lifecycle as evidenced by the lack of content management tools, version control capabilities, deployment mechanisms, expected user experience features like greetings, error messages, dialog flow, and lack of control of the frontend to address 508 accessibility issues.

### **Pursue a Custom Development Virtual Agent Framework**

Given PVA's limitations we are recommending a framework like Microsoft Bot Framework in order to have more control in developing a custom solution that will meet Veterans expectations when dealing with a modern interactive bot. This will also give VA.gov engineers the ability to address the gaps mentioned above for PVA and integrate with the many services in the VA ecosystem.

## Creating & Maintaining Content

Content is the user experience of virtual agents. It's the most time-consuming portion of the building and maintenance of a chatbot. There is no end to what a Veteran and their caretakers can ask of a virtual agent, and the amount of content at VA.gov is difficult enough for a search engine to comb through, let alone a chatbot that must "intuit" what a user is asking.

This section focuses on the content lift portion of this experiment, the lessons we learned while building a Proof of Value with a minimal data set of crafted unauthenticated dialogues, and ultimately our recommendations based on our experience.

### Initial Assumptions & Actions

Initial communications with VEO pushed for content to be pulled in from VA's Knowledge Management solution (known as KM or eGain), which is heavily utilized by call centers. However, content in KM is not frequently updated. It has no single point of contact for content updates, and it is not written with users in mind.

We chose the content within the modernized VA.gov site as the primary content source for the virtual agent proof of value for these reasons:

- Content is Veteran-facing and easy to read, and pre-approved by stakeholders
- Content is well-organized and linked, making it easier to find topic relationships to help draft relational dialogues
- Already established content team with established processes, making it easier to (at least discuss) the incorporation of the virtual agent into their content planning process
- Maintains content within Drupal, which we would use to pull content directly into the chatbot

As we progressed through the content creation process utilizing VA.gov, we discovered the following challenges & constraints:

- Automating content updates through Drupal resulted in non-conversational responses that required re-writing
  - The content at VA.gov, while well-written, did not directly translate into easily digestible information for chatbot readers.
  - Automated content displayed references to pages that would not have made sense within the context of questions posed to the bot.

health facilities, Vet Centers, regional offices, and other properties we own or lease.

To enter and remain in a VA facility, a service dog must meet all of these requirements:

- The dog must be trained to work or perform tasks for a person with a disability (including a physical, sensory, psychiatric, intellectual, or mental health disability). → Dogs that provide only emotional support, comfort, or companionship don't qualify as service dogs.
- The dog must be under the control of its owner or another handler at all times.
- The dog must not enter areas where an animal could cause problems with patient care, safety, or infection control (like operating rooms).

We don't allow any other animals in VA facilities. But we may make exceptions for certain needs like police dogs or animal-assisted therapy. →

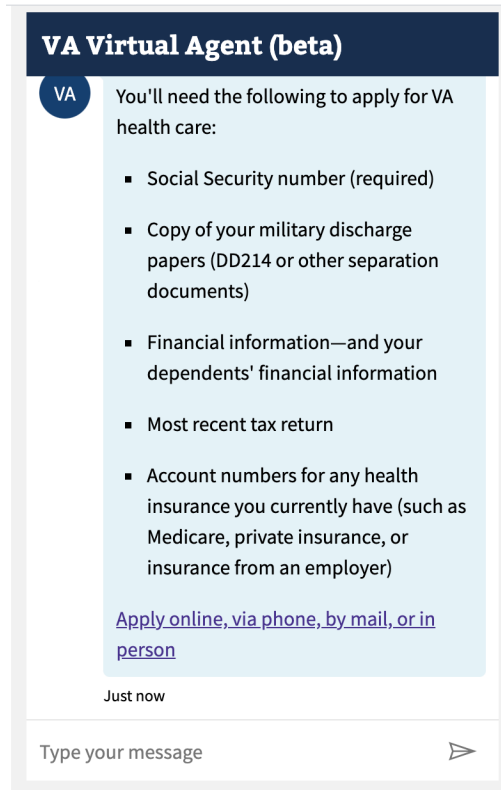
Type your message here ...

Example of a longer answer delivered in the chatbot

## Content & Proof of Value

In response to these findings, our team decided to draft a sample of responses and work with the content team on a "mini" content review process for the proof of value:

- We created approximately [100 health benefits and claims-related dialogues](#) and general bot responses (e.g., surface human, unknown responses, etc.)
  - Drafted responses for Veterans in crisis
- OCTO-DE VA.gov Content Team reviewed content on a weekly rolling basis.
  - We had to use multiple tools for collaboration between teams (including consultants) which made the process cumbersome and ineffective
- The content was uploaded using QnA maker in combination with Github so that three content managers were able to collaborate on content without impacting versions.
  - QnA maker did not allow us to break-up conversational responses, resulting in lengthy responses that may create cognitive overload for some (if not most) readers



## VA.gov & Virtual Agent

VA.gov has an existing set of design recommendations that include language form and style. As a conversational entity, virtual agents require a more *personal* touch to information presentation that differs slightly from how websites present information.

To create a dialogue that makes it easy for humans to interact with a virtual agent -- the point of a chatbot -- it's important to build on top of a website's voice and tone/content style guide in a way that's appropriate for a chat context.

This section looks over considerations in drafting dialogues within the confines of VA.gov:

### Tone & Voice

- Determining the use of “I” vs “We” or neutral voice for chatbot. Based on Veteran interviews, neutral is preferred because they know they are “talking” to a bot. This could change in the future when chatbot interactions and responses are more sophisticated
- Tone should be professional, service driven, and empathetic (especially for a future authenticated experience)

## Emergency and Crisis

Any open forum needs to address the possibility that a person may be experiencing a crisis while using that forum. **Artificial intelligence cannot adequately de-escalate a suicidal situation.** Consider the impact to the Veteran experiencing an emergency if the virtual agent surfaces an incorrect response or surfaces a dialogue that can trigger an event.

We recommend taking the following steps when handling crisis communications with virtual agents/chatbots:

First, decide **whether or not** the virtual agent should respond during a life-threatening crisis (e.g., danger to self/others, unrest, domestic abuse)

1. If the bot does not handle these issues at all, decide:
  - a. If the chatbot removes itself from responsibility via language "I cannot speak on this matter." (requires training to understand the user)
  - b. If the chatbot removes itself from responsibility via omission. The bot does not respond or surfaces an incorrect answer - helping to solidify that the bot is not a human.
  - c. Where else can they find help on the page (important if you are creating a link to a bot that opens on its own page)
2. If the bot will handle these situations, decide:
  - a. What sort of emergency the virtual agent should respond to (e.g., danger to self vs. hurricane support)
  - b. What information is enough information to help the person take immediate action
    - i. Text-only display of phone numbers
    - ii. Direct connection to a human (i.e., phone call, live chat)
  - a. Who responds to that action (e.g., Veterans Crisis Line, call centers, medical staff, off-hours help)
  - b. How often you review these responses and monitor conversations to identify people at-risk
  - c. How often you need to meet with legal or other lines of support handling these issues
  - d. The voice, tone, and language you will use to convey a helpful message
  - e. How much user experience testing (and with who) you will run to validate these dialogues
3. Work with Veterans Crisis Line (VCL) and crisis-resources early on.

Given the importance of this matter, we will say this again -- **do not attempt to train the virtual agent to *understand* how a human feels and de-escalate a crisis situation.** The current proof of value has the ability to understand some language that could be interpreted as threatening. We have seen instances where incorrect responses surfaced. The team did work briefly with Doctor April Foreman, resident Suicidologist to discuss this proof of value. Future iterations of this virtual agent will require thorough collaboration with VCL.

## Content Recommendations and Next Steps

Research interviews and unmoderated testing of the proof of value show that Veterans **want and expect to use the chatbot for self-service authenticated content.** Content will need to be carefully curated and/or drafted. To make the content-management experience efficient, we propose the following:

- Pursue a centralized knowledge point that has a shared content platform allowing teams to utilize the same updated content for a variety of products, including chatbot (VA.gov-wide effort).
- Explore creating a separate instance of Drupal database for virtual-agent only, and assess whether we can use markup to pull individual pieces of content (short-term solution).
- Hire dedicated writers and content strategists for the virtual agent. The VA.gov content team cannot support content creation for the virtual agent, or any other virtual agent
- Create a conversation-design style guide to implement within VA.gov existing Design System, that compromises VA.gov styling with the informational style of virtual agents.
- Pursue continued, iterative research into what content topics Veterans want to see
- Ensure that you have the adequate amount of engineers to implement API-focused responses to adhere to Veteran demands for authentication and personalized information.



## Challenges

For the proof of value, the Virtual Agent team utilized Microsoft PVA commercial trial licenses because we did not get access to Government Commercial Cloud production environments in time. Additionally, to subsidize costs for the proof of value, we also used QnA Maker, which shares space with the Coronavirus Chatbot. We had various challenges with environments and licenses that delayed our production by months and continuously interrupted our processes.

We "launched" the current proof of value under VA.gov staging environments. As of the creation of this document, the virtual agent has a Government trial license, with access pending for a full Government Community Cloud license (and production environments) in mid-June.

Environment & other challenges encountered prior to kick-off of production:

- Approximately 11 business days were spent troubleshooting PVA, unplanned outages, or poorly documented features.
- A span of 9 weeks passed waiting for PVA and QnA Maker access.
- [Time logged on Power Platform](#)

Some of the issues detailed below may be due to trial licenses. We have no way of verifying this at this time. However, below is a collection of the significant problems we encountered with PVA.

### Access to impactful reporting

- Conversational logs are missing confidence scores, which helps us understand what dialogues the virtual agent analyzed before choosing a response.
- Definition of "Engagement" is unclear. Some interactions with the virtual agent are logged as no-engagement.
- Conversational logs are not scannable. Conversations with the bot are logged in one cell, making it difficult to distinguish when the user ends and the bot begins.
- Inability to pull a content audit of existing dialogues

### Ability to fine-tune AI responses (e.g., confidence-score adjustments)

When a user's query is matched against a knowledge base, relevant answers are returned along with a confidence score. This score indicates the confidence that the answer is the right match.

- Within PVA, the ability to fine-tune AI responses (e.g., confidence-score adjustments) is not available ([#44](#))
- A chatbot framework should allow us to tweak the confidence score algorithm and provide us a mechanism to pull meaningful reporting

## Level of debugging required

System errors are not easy to track and troubleshoot. In April, we temporarily lost the bot and we spent two full days trying to figure out what happened. We are not sure if this was a trial license issue.

## Scale to sophisticated responses (e.g., payments widgets, APIs)

As indicated by the OCTO-DE Idea Gathering Questionnaire, other teams hope to integrate with a VA.gov chatbot. We briefly examined the ability for PVA to consume external APIs to allow for more sophisticated interactive experiences (claims widgets, appointment scheduling) but the Microsoft recommendation of Adaptive Cards at first investigation would not be robust enough to meet these tasks.

## Version control of mass content

No built-in version control exists in PVA, so we have to perform manual exports of the content in the bot. To do this, we are required to scrub the export of any security keys or sensitive information needed for deployments before pushing to GitHub. As a result of no version control, if the bot is lost during the import/export process, we would lose anything added in PVA since the last export.

## PVA Limitations

### Accessibility & 508 Compliance

To get a clear understanding of [Section 508](#) accessibility issues with our chatbot, we researched the accessibility of other chatbots and tested them with various screen readers including VoiceOver, the built-in screen reader on Macs.

We used [VoiceOver to test](#) BAH Bot, the National Suicide Hotline Chatbot, and the LA Business Virtual Network Chip Chatbot to determine how bot placement effects screen readers.

PVA requires custom development to make the user experience compliant with 508. We encountered several critical 508 [accessibility issues](#) identified during [Staging Review](#). The issue '[links in the chat log must be keyboard focusable](#)' and '[links must be announced to screen readers as links](#)' are both classified as [508 Defect 1](#), meaning the product would not launch on VA.gov because the issues severely disrupt the user experience especially with users who use assistive technology.

The Virtual Agent team [completed an assessment](#) to determine if the solution to these accessibility issues was possible through

1. Custom development within the Virtual Agent's control, or
2. If it was related to the VA Design System, or
3. If it required Microsoft's intervention

The results of this assessment are summarized in this table:

508 Issue	Responsibility / Sphere of Control		
	Microsoft	VA Design Library	Virtual Agent team
Accordions do not read properly on screen reader		✓	
aria-roledescriptions are not properly defined	✓		
Color contrast on message text boxes not sufficient	✓ (can be customized)		
Color contrast on timestamps not sufficient	✓ (can be customized)		
Screen reader focus and navigation	✓		
Skip to content		✓	
Smooth scroll in chatbot window introduces new issues for various cognitive abilities	✓		
Suggested Actions do not read properly on screen reader	✓		
Typing animation	✓ (can be customized, but not easily removed)		

We also identified that these issues belong to `botframework-webchat` (the user interface library that Microsoft provides) rather than PVA itself. As a result of our analysis, tickets have been filed with Microsoft for the issues that are only solvable by them.

While we worked within the Collaboration Cycle to quickly create our proof of value, the team recommends a full [508 audit](#) for future iterations of the virtual agent, and recommends continued evaluations through iterations utilizing testing with screen readers, for example, during development.

## Accuracy

We encountered multiple examples where we were unable to get the chatbot to answer questions accurately:

- Difficulty in accessing content while inside Entities. For example, we create a phone number Intent (parent), and place a list of phone numbers such as CHAMPVA inside Entities (child) for this intent. At times we were unable to surface CHAMPVA without first saying "phone number for." If a person said, "Call CHAMPVA" the bot was unable to respond.
- [Synonyms could only be one word and the synonyms file would be overwritten on each upload](#)

However, we understand that some of these issues have to do with the sample size of training data we used. This assessment is only for awareness and would require additional exploration.

## Feedback

We were interested in both quantitative and qualitative feedback from a user's experience with the chatbot. We tried asking the user for feedback after each question or interaction they had, but with the limitations of PVA, this conversation flow would get users stuck and unable to ask new questions. With no feedback mechanism built into PVA, we turned to the Contact Center team to add the Medallia customer satisfaction widget to our chatbot page ([#126](#)). However, ultimately for our controlled study, we chose to gather feedback from Veterans through a form after their interactions with the chatbot.

## Security & Privacy

We used commercial licenses to host the proof of value. To maintain privacy and security, we framed the proof of value as a study, and provided disclaimer language to prevent users from inputting PII data. However, Microsoft does have an encryption feature for content. **We did not test this feature.** Instead, we chose to mitigate issues before they could happen, and will rely on full Government Azure licenses for future deployments and releases to ensure privacy and security.

We did work with the collaboration cycle, and have or will implement the following:

- [Disable token retrieval API to prevent DOS attacks.](#)
- Turn off access to the study page after a period of study (June 30th disable date)

### Usability

PVA is limited when building questions and decision trees. Users can get stuck in conversations, and can't use buttons/prompts to escape that flow of question and answers and have to finish it before they can ask another question ([#115](#)).

## Next Steps

We have begun the process of assessing next steps for the virtual agent, which includes the following *high-level* considerations:

### Authentication

We have built a [high-level demo](#) demonstrating how PVA integration with ID.me and Lighthouse Claims API could help us authenticate a user:

- After logging in, a provided six-digit code must be pasted to the chatbot window to finish Authentication; this is the standard authentication behavior for PVA.
- We can call the Lighthouse Claims API by creating a custom flow and passing it in the AuthToken.

However, Phase 2 requires complete exploration and implementation involving Authentication, which includes evaluation of content and assessment of user experience between the Virtual Agent and Live Agent.

### Content

The Virtual Agent team will collaborate with VA.gov content team (and others) to assess how we can improve existing content management systems in a way that can help us:

- Reduce manual content creation for the virtual agent
- Ensure that the virtual agent has the latest information
- Add version-control to the virtual agent content process

### Ethics & Data Decisions

OCTO-DE & VEO should discuss if an “Ethics Group” is a necessity as the VA builds more open channels of communication. This committee would include subject matter experts in VCL, legal, health (including mental health), and public relations arena to:

- Address level of support for those experiencing suicidal ideations or other emergencies
- Decide what health and mental health data the VA can store & Veteran can have access to
- Create a system of response for political or media requests
- Create a system of protocols for emergency handling (e.g., hacks, improper bot responses etc.)

These decisions directly impact chatbot features such as records-storing and transferring (e.g., Live Agent/VCL).

## Expanded Research

There are populations that we could not reach during this short study. The Virtual Agent team recommends expansive research with the following populations:

1. Those experiencing economic insecurity (housing, food etc.)
2. LGBT+ communities
3. Rural populations
4. Low connectivity to internet or mobile-only data plans
5. Low visibility and other physical impairments

## Government Community Cloud Implementation

Full Azure licenses and production environments should be ready by mid to end of June. We will export all our current work over to the Government Community Cloud environment (GCC) when possible.

## Live Agent Connection

The Live Agent (or Live Chat) component of the Omnichannel Project has an estimated release date of September 15, 2021. It will support VBA/NCC calls only. **The virtual agent is not required to meet this deadline with VBA content.**

We will work with VBA to assess the best training data for the virtual agent based on feasibility and Veteran-need - and to help them incorporate the virtual agent into their existing business processes (i.e., content creation).

We will work with the Live Agent team to connect both components after user-testing and validation is complete.

## Mobile-First

By following a mobile first design approach, VA will make sure that the experience of using the virtual agent is successful regardless of what device a Veteran uses. The smaller screen/scale will also guide VA's Virtual Agent team in making the most essential choices in content and user experience design.

## Platform Placement

We have provided our review of Microsoft Power Virtual Agent (PVA). We *recommend* a platform (like Microsoft Bot Framework) that allows us to build a customized solution so that we can better control the user-experience, and promptly respond to technical issues. However, the final decision needs to be made at the business-level.

## System Design

Members of the Virtual Agent team are participating in discussions with VEO/OCTO-DE to explore how we can improve connections between disparate support channels. Connecting services makes it easier for the virtual agent to triage Veterans to the right support channels, and helps VA create and maintain consistency of information, branding and quality control - across all channels.

## Veteran-First Process

The virtual agent must be produced using an agile framework. This channel requires constant evaluation, iteration, and maintenance. We cannot adequately train and validate chatbot features using waterfall methods.

We understand that other components of the Omnichannel initiative may follow a waterfall method, or promote a system of requirements-gathering. The Virtual Agent team will take all requirements into *consideration*, but prioritize requirements that allow us to build and ship a product quickly that will help us understand how Veterans interact with the product. That knowledge will support cyclical iteration and continuous improvement efforts.

We also understand that we must work within a collaborative system (Jira) to promote knowledge-management. In Phase 2, the Virtual Agent team will start assessing the best ways to do this within an agile process.

## Conclusion

Veterans that used the proof of value chatbot during our study were successful in self-servicing for many of the use cases we executed. This shows evidence that a chatbot would be a great solution for VA.gov users looking for an additional VA customer support channel. Given the breadth of topics available to it, they responded well to its accuracy and professional tone. Their behaviors captured in chat transcripts demonstrated that the information the virtual agent provided was often useful and could be made more useful with refinement. Many Veterans indicated that they understood the chatbot was a work in progress and in spite of that, were willing to revisit it in the future. In order to deliver the content-rich, interactive, inclusive, experience owed to them, this report recommends we pursue a platform like Microsoft Bot Framework in order to more easily build, control, and maintain our interactions with Veterans.