VA | U.S. Department of Veterans Affairs ThoughtWorks®

Virtual Agent - Chatbot

Proof of Value Summary

Luciana Morais, Digital Services Expert (OCTO-DE)
Luciana.Morais@va.gov

July 7, 2021

Introduction

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 could 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.

A Proof of Value (POV) study was conducted to validate the assumption that a virtual agent/chatbot could provide value to Veterans by:

1. Increasing awareness of existing VA self-service tools
2. Decreasing the time Veterans spend waiting for an outcome
3. Giving them 24/7 access to either anonymous or secure support

This study concludes that Veterans want a chatbot for the reasons stated above. This report details the research and provides next steps for building a virtual agent that will exceed their expectations.

Contents

Proof of Value Study

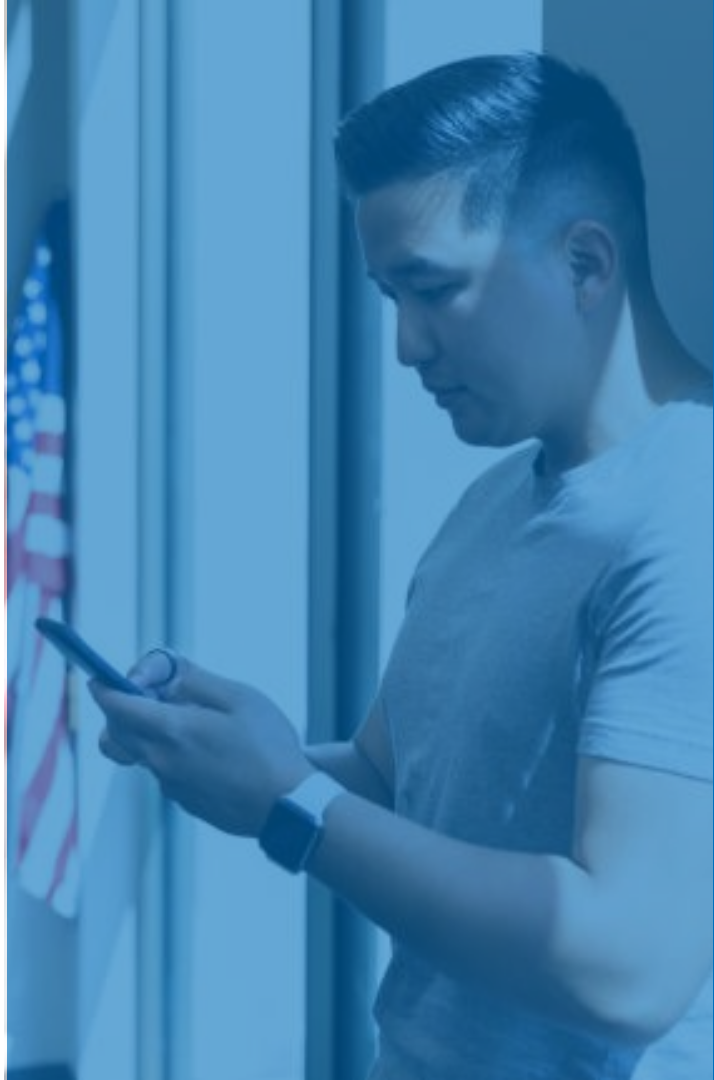
Chatbot Value Proposition

Value to Veterans

- Increase their awareness of self-service tools
- Decrease time spent waiting for an outcome
- Provide 24/7 access to anonymous or secure support

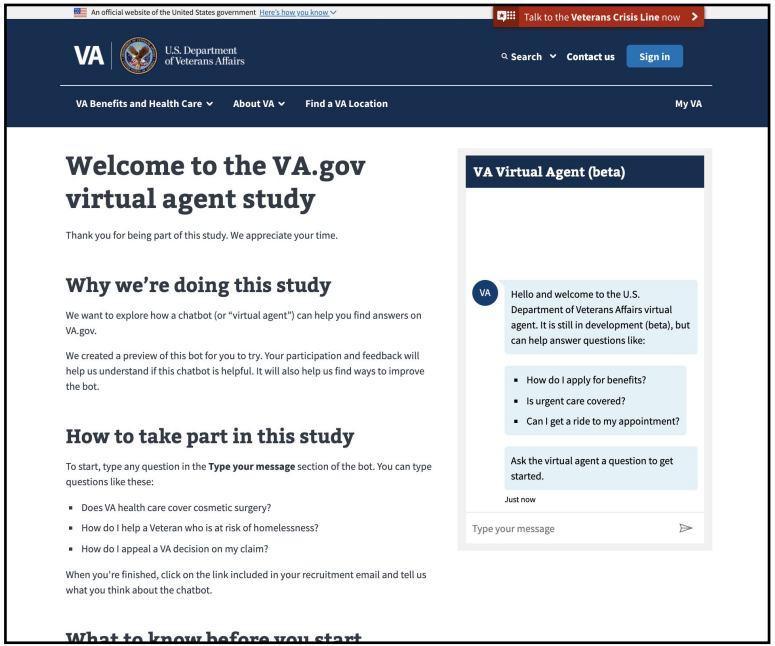
***Potential* Business Value**

- * Reduced call center volume



Proof of Value Study - Chatbot Features

- Based on a centralized VA.gov virtual agent mindset or “superbot”
- A greeting with suggested questions to start an open-text interaction
- Approximately 100 unauthenticated bot-responses covering health benefits and claims
- Responses for questions the bot cannot answer
- Ability to surface a phone number to reach a "human"
- Recognition and responses for Veterans experiencing a crisis
- Guidelines and information for those needing immediate help



Proof of Value Study - Goals

Do veterans want a bot?

Capture (directly through survey questions and interviews) if our customers think a bot could be useful

What should a bot know?

Use conversation logs and survey responses to gather a sample of topic areas that Veterans need a bot to know

What should a bot be able to do?

Understand what features would make a bot valuable to them

How should a bot act?

Explore how our customers feel about voice, tone, and overall communication with the bot

Proof of Value Research

- **OCTO-DE Branding Survey (March 2021)**
  - Distributed a survey to VA staff to discuss the chatbot's voice and tone
  - OCTO-DE Branding Report

- **Veterans & Caregivers Branding Interviews (April 2021)**
  - Likewise, interviewed 16 Veterans and Caregivers regarding naming the chatbot and how the chatbot should behave
  - Veterans & Caregivers Branding Report

- **Controlled Study (June 2021)**
  - A total of 44 (out of 100 recruited) participants tested the chatbot and shared feedback via a survey
  - Ten segments with a focus on breadth and diversity - including women and marginalized populations (e.g. LGBTQ, rural, economic insecurity). Not all recruited segments produced participants
  - Controlled Study Report

- **OCTO-DE Self-Service Feature Survey (June 2021)**
  - Questionnaire that collected ideas on how a VA.gov chatbot might help support increased engagement within their groups (seven respondents from Debt Resolution, Identity, Sitewide, Decision Tools and VAOS)
  - OCTO-DE Self-Service Feature Survey

- **Inclusive Design Study (June 2021)**
  - Interviewed two visually impaired veterans to observe any accessibility issues they faced
  - Inclusive Design Study Report

Research Findings

Do Veterans Want a Bot?

Overall, participants indicated a willingness not only to try the chatbot, but to use it again. They were excited about the product and how it could evolve to help them and VA.

> *I believe the chat box tool has great potential to answer my questions and reduce my time holding for a VA representative to answer my questions. (Casual, infrequent user of va.gov)*

> *Even though I said I prefer talking with a live person I would try using the chat box in the future if it was something I didn't feel was urgent to allow the persons to help others with urgent needs. (Male Veteran)*

Controlled Study - June 2021

Do Veterans Want a Bot?

Participants experience challenges with current VA customer support channels. A VA chatbot was viewed as an opportunity to avoid those challenges.

> *It can help if I have quick questions rather than going through the automated phone tool to get the answer I'm looking for. Plus, it can help if I have questions at hours that the VA is not available (Caretaker)*

> *I am satisfied with the bot. If it didn't directly answer my question, it provided a link to where I could look for the answer on another page. (Casual, infrequent user of va.gov)*

Controlled Study - June 2021

What Should a Bot Know?

The team analyzed the conversation logs and identified these as the most common topic areas, or topics participants asked about:

1. VA programs and eligibility for them (36 separate engagements - or users typing their question 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 (LGBTQ, disabled, etc.) and others.

Conversation Logs - June 2021

What Should a Bot Know?

For each of the top categories, examples within each category included:

- VA programs and eligibility for them
  - Many topics had names of individual programs (‘Veterans Choice Program’) and how to access them (‘qualify for community care’)
  - Wide range of additional topics (‘medical professionals’, ‘dependents’ related questions, ‘dental services and benefits to pay for them’)
- Misc. medical topics and conditions
  - Names of conditions (‘blood pressure’), ways to address them (‘new eyeglasses prescription’) and understanding coverage (‘eyelid surgery coverage’)
- Disability related
  - How conditions relate to disability (‘different types of disability compensation’) as well as managing the process (‘length of appeal for a disability decision’)

What Should a Bot Be Able to Do?

Additionally in June, we surveyed other OCTO-DE teams to understand how a chatbot might integrate with their system in order to provide Veterans with a true self-service agent. The teams suggested the following:

- Chatbot-led form completion
- Sign in assistance/password reset
- Conversation History Logging
- Appointment Scheduling options
- Benefits applications
- Debt management/resolution topics

OCTO-DE Self-Service Feature Survey - June 2021

How Should a Bot Act?

Using an automated chatbot can raise issues of trust especially around personal information, but participants indicated a desire to sign in and have a personalized experience, which signals a level of trust with the product.

> *I think I would've preferred a more personal approach in the beginning, especially if I'm signed in to VA.gov. For example, when I type my first question or open the chatbot, I would want some type of acknowledgement, "Hi, NAME. How can I help you today?" or "Thanks for your question, NAME. Let me get you some more information on that. (Caregiver)*

Controlled Study - June 2021

How Should a Bot Act?

Participants understood the chatbot wasn't human. Although some felt it was a little 'cold', many appreciated the voice and tone. Some mentioned that personality was less important - getting the answers was the primary goal.

> *I don't like overly friendly chatbots. It's just weird. This one just gets you answers, and I liked that. (Casual, infrequent user of va.gov)*

> *It was very professional/matter of fact but not rude at all. They know that the answers they are giving you could possibly save a life. (Female Veteran)*

> *I do not have an opinion about the voice; the content of the answers is all that matters. (Casual, infrequent user of va.gov)*

Controlled Study - June 2021

How Should a Bot Act?

VA staff (OCTO-DE) felt similarly regarding the chatbot's voice and tone:

- Our respondents were 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
- As the sophistication of the chatbot evolves to provide more nuanced responses and follow-up questions, we can revisit the topic of giving it a proper name

OCTO-DE Branding Report - March 2021

How Should a Bot Act?

We also interviewed 16 Veterans and caregivers in April 2021 for their perspective of tone and personality.

- Veterans, caregivers 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

Veterans & Caregivers Branding Report - April 2021

Study Challenges

1. Lack of Government environments and licenses prevented a larger study (i.e. 5% of visitors, A/B)
2. DNS issues negatively impacted mobile participants
3. Funding allocation impacted study length
4. 508 platform concerns require remediation before a larger study can be undertaken
5. Inadequate PVA metrics impacted discovery (i.e. confidence scores)

Platform Investigations

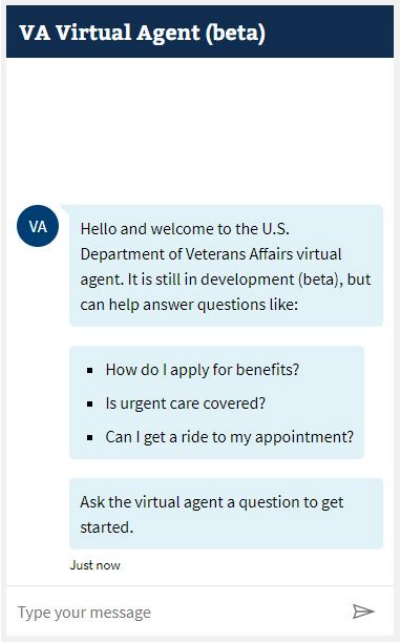
Platform Guiding Principles

- A centralized VA.gov-wide virtual agent (“superbot”) to provide a more robust, consistent, reliable, and maintainable experience for Veterans instead of multiple disjointed bots
- Longer-term planning - a roadmap that captures iterative and scalable features over time (years), and thorough consideration of inter-system and agency impact (i.e. call centers)
- Agile process that puts the Veteran *first*

Microsoft Power Virtual Agent Challenges

The Microsoft Power Virtual Agent (PVA) platform was used during the POV to build a chatbot that Veterans could interact with.

- We utilized a limited Microsoft PVA commercial trial licenses to build our POV while we awaited the fully-loaded PVA platform to be available in the Government production environment
- We had various challenges with environments and licenses that delayed our production by months and continuously interrupted development
- Some of the features gaps may be due to the trial licenses. We have no way of currently verifying this
- We used QnA Maker (a no-code question and answer content creation tool) which shares space with the Coronavirus Chatbot. It is not a fully-fledged content management system



The following slides summarizes other issues we encountered, and some of the gaps the PVA platform has in meeting Veterans' needs we previously outlined.

PVA's Gaps In Meeting Our Needs

- Access to impactful, intuitive, and digestible reporting (i.e. conversation logs, response ranking by AI confidence)
- Ability to adjust bot responses based on conversation logs (i.e. dialogs, trigger words, response-confidence scores)
- Ability to easily connect conversational topics without code
- Ability to easily pull APIs (preferably without code)
- **Accessible and 508-compliant interactions**
- Support for automated testing and deployments
- Easy UI customizations (i.e. look and feel of buttons)
- Version control of mass content that could act as a temporary knowledge base*
- 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)

PVA Has 508 Accessibility Gaps

To validate some 508 issues raised from Staging Review, we conducted a moderated study with three visually impaired Veterans. We observed:

- Two Section 508 Defect 1 status issues (i.e. the most severe, ‘must fix’ issues), that must be fixed before launching on va.gov
- Many less urgent but additional accessibility issues, like finding the chatbot on the page in the first place
- Blind Veterans can differ in how they use technology, including assistive technology
- Despite 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

> *My spelling is terrible - if i could dictate into it. Anything complicated for words. Typically if i want to use a chatbox, I type it up in MS Word first, spell check it and then copy it into the chatbox. (Participant #1)*

> *I’m pretty excited about these types of products. Being a blind consumer, I don’t deal very well with paper. A form that could be pre-populated with my information would be helpful. You have all my information already, so there’s no reason you couldn’t just populate it already. (Participant #1)*

Moderated Inclusive Design Study - June 2021

PVA Requires Custom Development

Days lost to technical challenges

- PVA is not a fully mature framework and the development team met many roadblocks during the course of standing up the chatbot for the study.
  - Approximately 11 business days were spent troubleshooting PVA, unplanned outages, or poorly documented features.

Inability to test decision-tree experience

- PVA has unpredictable decision-tree (i.e. prompts) logic and behaviors
  - Users got stuck in conversations, and were unable to use buttons/prompts to escape flow of questions and answers. Team was unable to troubleshoot, resulting in the team being unable to test prompts.

Why PVA does not support VA needs

- 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.
- 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.
- Actions such as changing **Greeting phrases were inaccessible**, making it difficult to impact the AI manually to improve user experience.
- **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.
- **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 (clicks on a button)

What We Need...

Content Managers

- Need to be able to quickly and easily create new content to meet Veterans demands. They need to be able to display this content on the chatbot the way they want it.
- ...run and assess reports for content quality
- ...map existing content and maintain content versions - without a developer

Engineers

- Need to be able to identify and troubleshoot system issues
- ...respond to 508 concerns by being able to customize the chatbot user interface
- ...build rich interactive features that Veterans want - without platform constraints or excessive learning curves

Recommendation:

Fully Customizable Virtual Agent Using Microsoft Bot Framework

Microsoft Bot Framework

Pros:

- ✔ Code-based, ability to do custom development
- ✔ Built-in hosting options with Microsoft Bot Service. Upload your bot, you're done
- ✔ Mature, highly-customizable, actively maintained frontend client
- ✔ Could reuse the frontend we already have
- ✔ Clear escalation path when you want something changed/fixed
- ✔ Microsoft would be responsible for data management / encryption, etc.

Cons:

- ✘ Framework is low-level, so big learning curve to make changes
- ✘ Frontend client does have some "advanced" accessibility (508 compliance) problems

Other Platforms

What Others Are Using

Rasa (FSA)

Pros:

- ✔ You design conversations, so more of a domain-specific language than coding. Readable by a non-techie.
- ✔ Can still use custom python code to do things like API calls
- ✔ Whole bot is in one application - don't need extras like LUIS. QnA Maker or something equivalent would still be useful
- ✔ Built in high-level framework so we could test that certain phrases triggered certain intents, etc

Cons:

- ✘ DIY hosting - need somewhere you can put a docker container. Would be responsible for support (protect against DDOS, performance test, figure out how and when to scale)
- ✘ Secret management to set up the JWT credentials
- ✘ Haven't found mature frontend client. Accessibility problems high in the ones we have seen
- ✘ The frontends not as customizable as the Microsoft provided frontend
- ✘ No escalation path when we need frontend changes. Its frontends are open-sourced, not actively maintained, and not backed by a major corporation

Google Dialogflow (USPS)

Can do many things but no clear differentiators compared to other chatbots

Pros:

- ✔ Comes with API which seems to allow some form of automation
- ✔ Virtual Agents come with support for many languages
  - Make the bot more accessible for someone whose first language is not English
- ✔ Tests can be written to test the bot responses
- ✔ Comes with different environments to make deploying from development to production easier

Cons:

- ✘ Dialogflow Messenger (Frontend library) is in preview and means it has limited support for the time being
- ✘ Manual Exports of the bot must be made
  - If the bot is deleted then all data is lost since time of last export
- ✘ Long term support is unknown. Google is known for discontinuing projects
- ✘ Usual low code problems

Other Platforms

Salesforce Einstein

Einstein would be a great tool to use if you need tight Salesforce integration, but it suffers from the same problems as PVA because it is also a low-code tool. We assume the deployments would work better than PVA but would require additional investigation.

Pros:

- ✔ Apex Actions allow developers to write code to handle bot functionality
- ✔ Access to detailed conversation logs
- ✔ Apex Actions can be managed through standard software development practices
- ✔ Apex Actions are testable

Cons:

- ✘ Very tightly coupled to Salesforce
- ✘ Deactivate bot required to edit
- ✘ Have not found any ways of doing UI customization yet
- ✘ Seems to need a lot more training utterances
- ✘ Unclear how to deploy bot
- ✘ Usual low code problems

Challenges

Content

Content *is* the user experience for virtual agents. It's the most time-consuming portion of building and maintaining 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.

Why We Chose VA.gov

- 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, potentially making it easier to discuss incorporating the virtual agent into their content planning process
- Maintains content within Drupal, which we *could* use, in the future, to pull content directly into the chatbot

Handling Bot Content at VA

- We drafted 100 dialogues for the POV (not including general responses)
  - Automated content from VA.gov, while well-written, did not directly translate into digestible information for chatbot readers
- Veterans know they are “talking” to a bot and they want it to know who they are. This will require a Conversation Style Guide that differs from VA.gov’s current web-focused voice & tone
- Collaboration between teams (including consultants) required multiple tools and processes to manage content, which made the process cumbersome and ineffective
  - Dialogues were hosted on Github, making it easier for multiple content owners to draft and manage version control
  - We used QnA Maker to upload content into the bot. This resulted in lengthy bot responses that increase cognitive overload
  - Weekly content reviews, while helpful, will not be effective long-term
- The existing content team (while amazing) would be overburdened by the demands of a bot. Dedicated writers are needed for this project

Emergency and Crisis

- Artificial intelligence cannot adequately de-escalate a suicidal situation.
- We cannot currently train the virtual agent to understand how a human feels and de-escalate a crisis situation.
- We have seen instances where incorrect responses surfaced.

The team worked briefly with Doctor April Foreman, resident Suicidologist to discuss this POV. **Future iterations of this virtual agent will require thorough collaboration with VCL.**

Should the Bot Answer a Crisis?

Yes

- What sort of emergency the virtual agent should respond to (i.e. danger to self vs. hurricane support)
- What information is enough information to help the person take immediate action
- Text-only display of phone numbers
- Direct connection to a human (i.e., phone call, live chat)
- Who responds to that action (i.e., VCL, call centers, medical staff, off-hours help)
- How often you review these responses and monitor conversations to identify people at-risk
- How often you need to meet with legal or other lines of support handling these issues
- The voice, tone, and language you will use to convey a helpful message
- How much user experience testing (and with who) you will run to validate these dialogues

No

- How should the chatbot remove itself from responsibility?
  - Via language - "I cannot speak on this matter." (requires training to understand the user)
  - Via omission - The bot does not respond or surfaces an incorrect answer - helping to solidify that the bot is not a human.
- 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)

Recommendations

General Recommendations

- One centralized VA (Virtual Agent) “superbot”
- For richer interactions like payment widgets and building conversational repartee, a framework that supports custom development and developers' involvement is required
- More branding (naming) research
- Expanded research with marginalized populations
- Implementation of PowerBI to pursue targeted metrics

Content Recommendations

1. Pursue a centralized Knowledge Base, allowing teams to utilize the same updated content for a variety of products, including chatbot (VA.gov-wide effort)
2. Create a separate Drupal instance for the virtual agent only, and assess whether we can use markup to pull individual pieces of content (short-term solution)
3. 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
4. Create a conversation-design style guide within VA.gov's existing Design System that marries VA.gov styling with the informational style of virtual agents
5. Continue iterative research into what content topics Veterans want to see
6. Collaborate with VA platform engineers to implement API-focused responses to adhere to Veteran demands for authentication and personalized information

Next steps

Phase 2: Authenticated POV & MVP

- Giving Veterans what they want
  - Authentication
  - Transfer to live agent (timeline pending)
- Building with Veterans
  - Continued inclusive research
- Better metrics, better decisions
  - PowerBI incorporation
- Protecting our Veterans
  - Ethics & Data (discussions)
  - VCL
- Strengthening our Foundations
  - Content at VA (discussions)
- Accessibility
  - 508 compliance
  - Mobile-first
- Working together
  - Incorporating VBA (pilot LOB)

To everyone...

Ahmed Ahmed
Evan Albert
Lauren Alexanderson
Jeff Barnes
Tuskee Barnes
Aaron Barth
Patrick Bateman
Booz Allen Hamilton HCD
Karl Brown
Mike Chelen
Allison Christman
Kevin Clawson
Dave Conlon
Trisha Dang
Anita Dewitt
Joanne Esteban
Clive Flory
Dr. April Foreman
Jacob Gacek
Peggy Gannon
Megan Gayle
Noah Gelman
Shira Goodman
Peter Hill
Katie Ho
Kevin Hoffman
John Hashimoto
Brian Hurst
Christopher Johnston
Ahmed Kochaji
Medha Kulkarni
Chante Lantos-Swett
Liz Lantz
Jennifer Lee
Brian Lloyd
John Lundy
Emily Mann
Abraham Marinez (ed.gov)
Dror Matalon
Ric Miller
Rachel Murray
Shefali Nayak
Mikki Northuis
Mark Nunemaker
Mary O'Toole
Meg Peters
Dennis Peterson
Paul Phillips
Trevor Pierce
Beth Potts
Ramya Ramesh
Rebecca Rebstock
Rachael Roueche
Ian Santarinala
Shahrukh Shamshair
Helen Shin
Shane Strassberg
Samara Strauss
Danielle Thierry
Ryan Thurlwell
Alyson Watson
Steven Weinberger
Christine Wilkes (ed.gov)
Martha Wilkes
Jason Wolf
Mary Woodruff
Charles Worthington
Kaeli Yuen

Thank you!

(If we forgot to include you, sorry!)

Q & A