AnswerLab®

# Elements of a Successful Digital Assistant and How to Apply Them

# Table of Contents

## Introduction



## Dimensions



## About the Research

## What is a digital assistant?

'an AI assistant that understands natural language by text or voice commands and helps users get things done'

Digital assistants go by many names - chatbot, virtual assistant, conversation agent—to name a few. They live in different devices (phones, smart speakers), platforms (apps, web), and environments (cars, homes). And, they can be activated by typing, talking, or both.

They might be designed with human-like features, perhaps eyes and a mouth, like in a social robot: the consumer robot Pepper and Honda's Asimo are good examples. Or instead, they can be a disembodied voice like the search-engine based assistant Google Now (2012) that evolved into what we know today as the Google Assistant.

## What does this mean for brands?

Regardless of what we call them, digital assistants are becoming more common. Since Apple's launch of Siri in 2011, we've witnessed a surge in market commitment to digital assistants from both tech giants and businesses across industries. Amazon (Alexa), Google (Google Assistant), Microsoft (Cortana), and Samsung (Bixby) each have their own proprietary assistant. According to recent data, 300,000 bots are automating communication between people and businesses via Facebook Messenger each month.

Brands are using digital assistants to provide customer support, deliver services, recommend products, and entertain users. Like a human assistant, a digital assistant's purpose is to make your life easier and more pleasant by efficiently completing the tasks of daily life and freeing up time for deeper work or more enjoyable activities. Digital assistants that do this successfully can improve a brand's customer experience, increase brand loyalty, and decrease customer support costs.

# What makes a digital assistant successful?

The technology behind Siri, Alexa, and their peers is evolving rapidly but the user experience is lagging behind. One year on from our first wave of smart speaker research, brands still struggle to meet user expectations when it comes to conversational interactions. To understand what elements make up a digital assistant users need and want, we conducted research with a variety of digital assistants including 9 smart speaker applications and 12 phone-based chatbots, in a nation-wide exploratory study. Ultimately, we hope to help brands create assistants that do not frustrate but instead add value as part of a broader customer experience.

## A note on chatbot history

Chatbots originated in a primitive yet influential computer program called ELIZA in 1966. The early goal of such chatbots was to pass as human. As shown in Figure 1, it was easy to guess that ELIZA was a computer. Although ELIZA clearly failed to mimic a human, the challenges it faced in identifying contexts and keeping a conversation going are the same challenges we face today. Fast-forward 60 years after ELIZA, the tidal wave of AI's history and development now points to a new goal.

As the advisor to many tech giants' founders and the pioneer of digital assistant, Terry Winograd would say, the proper goal of artificial intelligence is to help humans live more fully human lives, not to replace them.

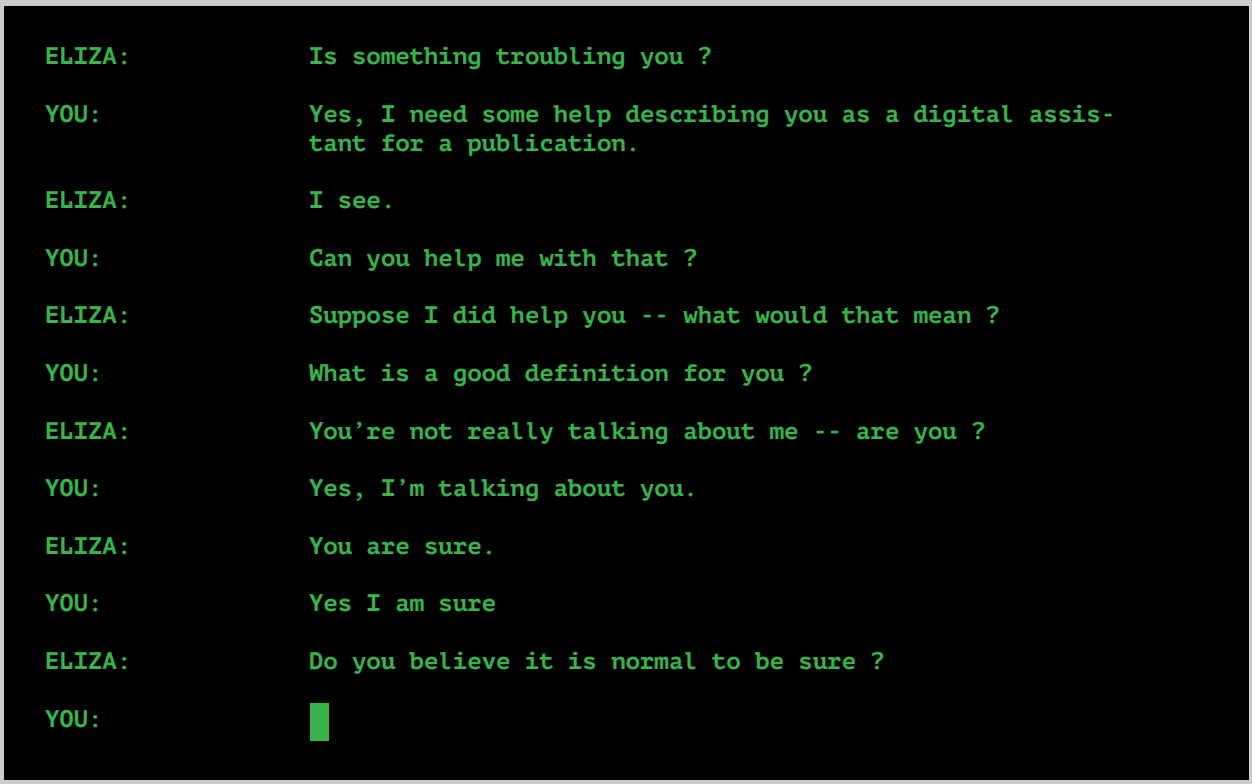



*Figure 1. A sample dialog with ELIZA (retrieved from http://www.masswerk.at/elizabot/eliza.html)*

# Key Findings

We conducted a 2-week diary study with 37 smart speaker owners across the United States to understand what makes a digital assistant successful, followed by in-person in-depth interviews with a selected group of participants from the diary. Participants completed daily tasks and reported their impressions and unmet needs from using a variety of chatbots and voice applications. Tasks included shopping, banking, investing, diagnosing health problems, and general information-gathering.

### 01
### Expectations change over time.

- While many early adopters were patient with smart speaker capabilities, over time expectations have risen. It's much harder to impress new users, and early adopters are getting tired of hearing "Sorry, I don't understand that." repetitively from their smart speakers.
- Users wanted more fun and engaging conversations with the digital assistants over time.

### 02
### The ability to get things done is more important to users than the how.

- While clear input modality preferences exist, users see chatbots as the 'written Alexa' and expected both to deliver added value, effectiveness, and engagement, whether they type or speak to the assistant. Users judge a digital assistant purely based on its ability to complete tasks, not the level of technology present in the backend.

### 03
### Hand-off experiences are crucial.

- Handoffs between the digital assistant and other platforms are becoming more common, e.g., from a smart speaker without a screen to a screen-based device such as a phone. Users are often disappointed by not knowing where to go next or simply being dropped when the digital assistant doesn't seamlessly transition them to another platform. Users expect more hand-holding from the digital assistant, especially over voice.

### 04
### Brands must build trust, especially when stakes are high.

- Trust is the connective tissue that ties the users to the product. Users will not return to a digital assistant after one use if it fails at tasks that have high social, financial, and health risks.
- When stakes are high, build trust by setting expectations, communicating values and limitations.

### 05
### Design is critical to building effective conversations between digital assistants and users.

- Technology alone won't solve experience problems - AI is inherently context-naïve. Teams must consider and design for the dynamic and complex user contexts.

### 06
### When digital assistants fail, brands do too.

- In today's experience, brand absence from digital assistants is not yet a problem, but a mediocre experience is. And, for established brands, the bar is higher.

### 07
### Brands must get three experience dimensions right for a digital assistant to be successful:

- Value
- Effectiveness
- Engagement

AnswerLab®

# Three experience dimensions

# Three experience dimensions brands need to get right for a successful digital assistant

Many decisions go into building and launching a digital assistant. When starting from scratch, brands need to evaluate whether a digital assistant is the right solution for the users' needs. Then they need to consider technology approaches, data sources, back-end systems, user scenarios, and existing brand elements like tone and voice. The process is daunting to say the least.

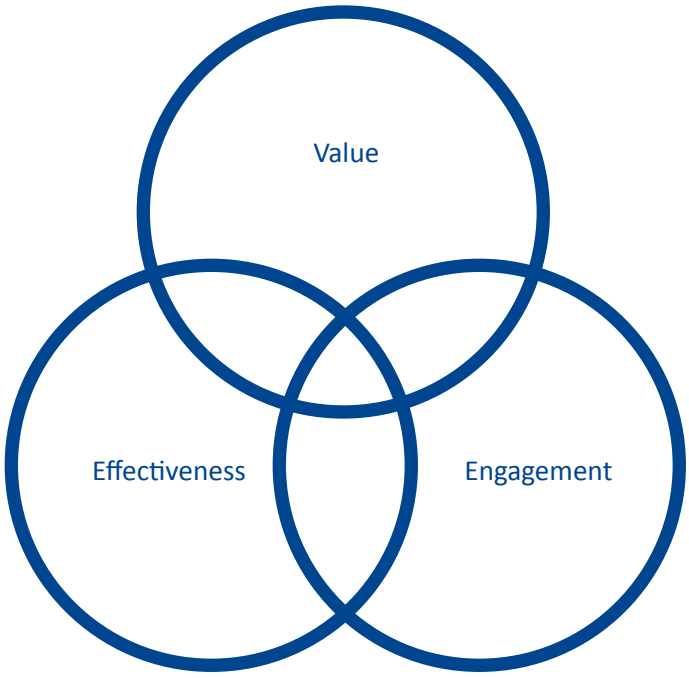
Brands have a number of technical levers to pull when designing digital assistants. And often, the level of AI involved depends on the job to be done. But building effective conversations between digital assistants and end users is increasingly more of a design problem than a technical one.

Regardless of your approach, all successful digital assistants need to balance these three dimensions:

**Value**:
Provide value beyond existing brand interactions

**Effectiveness:**
Accurately understand and answer inquiries, be knowledgeable and aware of the contexts and flexibilities in human interactions

**Engagement:**
Show signs of personality and brand voice in the appropriate contexts




It's not realistic for brands to tackle everything when they're just starting to build their assistant. Prioritize value and effectiveness, and then the sky's the limit for engagement.

AnswerLab®

# Dimension 1:
# Value

# Dimension 1: Value

To be adopted by users, a digital assistant must deliver on the value dimension by adding value to users' existing ways of interacting with a brand.

From a user's point of view, delivering value is the first requirement to adopting any new technology. This was the case across both conversational input modalities (text and voice) and across task types (from making payments to managing accounts and information seeking) in our study. When it comes to adopting a digital assistant, emphasis is placed on 'adding' value to how participants normally complete tasks.

"Convenience is top priority. I want an assistant that's easier than using an app."
- Albert, Participant

## Digital assistants add value by either:

**01 Enabling users to get tasks done in less time**
**OR**
**02 Providing a richer interaction in the same amount of time**

### 01
### Enabling users to get the same task done in less time

Capital One's chatbot Eno reduced users' need to click around an app or web. Unlike other phone-based chatbots that live on a bot platform or a stand-alone app, Eno answers entirely by text messages once the user links their account. Users liked this because they only needed to login once to do routine tasks, eliminating the need to remember and enter passwords.

Similarly, users liked that Bank of America's chatbot Erica allowed an extensive list of frequently used banking tasks in one place. In fact, many mentioned that interacting with Erica could be a new habit-forming interaction.

"It (Bank of America's Erica) made my mobile banking experience much easier; if I got used to it, I probably wouldn't browse around the app as much anymore."
- Luis, Participant

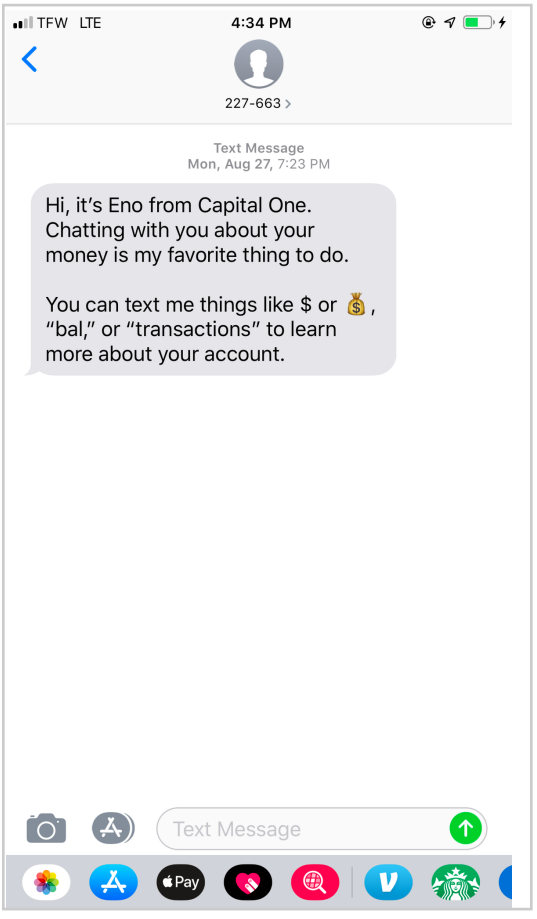



*Figure 2. Capital One's Eno introductory message*

## 02
## Providing a richer interaction in the same amount of time

Ada, an AI-powered health assistant app for symptom assessment and health information exceeded participants' expectations. Ada lets one type in symptoms and then asks a list of questions to narrow down to probable causes. The causes came with their statistical likelihood based on an ever-growing medical database matched to the user's profile.

Ada narrowed down results and guided participants to a diagnosis faster than they could otherwise reach via a web search or a wait at the doctor's office.

"Ada was very informative compounding both personal medical information, medical history, medications. [It] gave information following a quick set of questions compared to the traditional paperwork. Having the app makes it convenient and accessible."

- David, Participant

# Digital assistants that didn't demonstrate value failed due to:

**01 Unmet user expectations**
**OR**
**02 Interactions not improving on existing means of interaction**

## 01
## Unmet user expectations

A digital assistant may be quite effective at a few use cases, but fail to deliver on the ones that are most needed and expected.

Participants could successfully check balances and send and receive money from contacts using the PayPal voice app. Participants wanted to use voice to search past transactions and send money to new contacts, as opposed to just existing contacts. Many expected the existing PayPal voice app to satisfy these needs and felt disappointed that it did not.

"I think the PayPal skill's functionality is limited. I can do the same things plus more on either my phone or laptop."

- Max, Participant

"I didn't find any value at all in PayPal. Typically when I use PayPal I'm sending money to a new contact. I couldn't actually do this using the skill and I actually would've had to have sent money to the person already in order to even have a new contact?!"

- Scott, Participant

Perhaps due to the more valuable digital assistants that users were exposed to in the study, most did not "forgive" American Express's Facebook Messenger bot for being a menu-based bot.

This means that despite being told by the assistant about the limits of its capabilities, e.g., the menu options, users expected more from the bot and felt frustrated when the bot could only do a few things (see Figure 3).

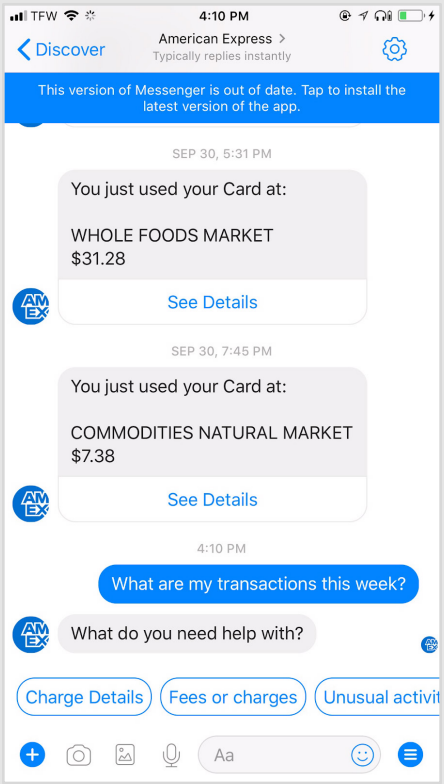
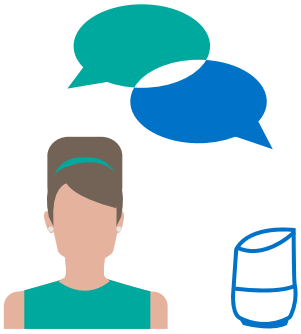



*Figure 3. Sample American Express Messenger bot interaction*

## 02
## Interactions not improving on existing means of interaction

When a digital assistant failed to be a faster way of interacting with a brand, it was abandoned. Participants commented that they would “just Google” or “just use the app” instead when using a chatbot or voice experience was taking too long.

In this example, a participant asked Progressive about home insurance. The assistant responded with information about auto insurance. He not only felt annoyed, but perceived this as an attempt to upsell irrelevant information, which degraded his trust in the brand.



| | |
|---|---|
| *Morteza:* | *Hey Google, ask Progressive about home insurance.* |
| *Device:* | *Okay, let’s get Progressive...Once you have your car picked out, you’ll want to revisit your insurance policy. Progressive can help save people over $600 on car insurance. Would you like me to send you an email to get a quote?* |
| *Morteza:* | *No.* |
| *Device:* | *Alright, no problem. Also, if you’re buying a used car, there are 5 things you really need to know. Would you like me to tell them to you?* |
| *Morteza:* | *No.* |
| *Device:* | *Okay, is there anything else I can help you with?* |
| *Morteza:* | *Obviously not.* |
| *Device:* | *<no response>* |
| *Morteza:* | *No, go away.* |
| *Device:* | *Alright. Thanks for visiting Progressive. Talk to you later.* |

“I thought the Progressive (voice app) was a real waste of time. I asked for home insurance and it gave me overt advertising...this app is designed just for advertising, and did not accomplish anything for me.”

- Morteza, Participant

# Brand **absence** is not a problem. Brand **mediocrity** is.

With the advent of any new technology comes a fear of missing out. While brands worry about not being part of a cool, new technology, we observed that users don't judge brands by their absence or presence in the digital assistant race, but rather the quality of the experience. When digital assistant experiences didn't deliver value, participants were not willing to forgive the brand. At best, they blamed the platform on which the assistant is built, if different from the brand. In most cases, a poor experience hurt brand impression.

> "I thought American Express is a premium brand but after using their chatbot, I changed my mind. The chatbot was very poorly designed so it makes me wonder if American Express put any effort into it."
>
> - Vadim, Participant

When an experience delivered exceptional value, brand impression was elevated significantly.

> "Bank of America's Erica definitely improved my perception of the Bank of America brand. It is great to see my bank staying ahead of the curve."
>
> - Luis, Participant

Many participants have the perception that the goal of automation via digital assistants is to help businesses cut costs and human jobs. Poor experiences were frequently interpreted as brands putting business needs before their own. Interestingly, this often led to exceptional experiences exceeding their expectations.

> "The chatbot Eno made me think that Capital One is trying out a new way to have basic financial questions answered for their consumers at the consumers' convenience. That's great. **Sometimes I feel like when companies do something similar, they are just trying to find ways to cut jobs,** however, I think capital one designed this with customer convenience in mind."
>
> - Marieh, Participant

Recommendations

# How brands should approach providing added value

For a digital assistant to become part of people's everyday lives, it has to be useful. It must actively solve users' problems on a scale that exceeds their current level of value.

For brands to ensure they're building a digital assistant that provides added value, we recommend:

01

**Ensure you're solving for a problem that exists,** not just jumping on the digital assistant and AI bandwagon. Conduct exploratory user research to understand the opportunities.

02

Once you've identified a user need, **assess whether a conversational interface is the right tool for the job.** This quiz by Google Assistant's Design Guide is a great first step to see whether you should even invest in a conversational assistant.

03

**Compare your digital assistant interaction to users' normal means of interaction**, e.g., web or app; for the same task to ensure it takes less time to complete a task in the former than the latter.

04

Once you demonstrate value, the most **successful digital assistants remain sticky by progressively disclosing useful features.**

05

**Check in frequently to monitor whether users think your digital assistant continues to provide value.** New technologies suffer from the novelty effect - users give the experience high ratings at the beginning due to novelty but interest quickly declines over time. User longitudinal research can reveal changes in user behaviors and attitudes over time.

AnswerLab®

# Dimension 2:
# Effectiveness

# Dimension 2: Effectiveness

Providing added value is only relevant if a digital assistant can effectively execute on those expectations. An effective chatbot or voice assistant is flexible in interpreting the questions, able to keep track of conversation, and gives accurate, on-topic answers in a convincing, natural manner.

As with value, there's a bar of functional effectiveness that must be met for user adoption. Users want to save time and free attention by using an assistant. A digital assistant is only as efficient as it is effective, and it is most effective if it delivers results without friction. Users want services that make complicated tasks simpler and faster to complete, without jeopardizing accuracy. Voice input using the state-of-the-art speech recognition software is three times faster and 20.4% more accurate in English, than typing on a smartphone keyboard. [1]

Yet, participants in our study frequently reported ineffective interactions. Many found getting responses like "Sorry, I don't understand that." tiresome. Early adopters of smart speakers, in particular, expressed growing frustration that their speakers weren't "catching up" quickly enough to reliably understand and respond, especially with routine commands.

The implications of task failure further impacted users' perception and tolerance of effectiveness. When menial tasks or trivial questions failed, like asking for baseball stats or directions, participants were less annoyed.

> "I've had the speaker for 2 years now. The assistant says 'I don't understand' to something I ask every week. That's not good. I'm not sure if it is the hardware, software, or if it is me. It's better to say 'I'm trying' than just 'I don't know.'"
>
> - Ashish, Participant

When stakes are high - like financial, social, or health risks - a failure means users immediately abandon the assistant and revert to more reliable ways of getting tasks done by themselves. For example, many commented that they expected a banking and credit management digital assistant to give accurate account information on the first try; they would never return to it if mistakes occur.

What does effectiveness look like? Participants in our study rated chatbots and voice assistants as highly effective when they:

- efficiently and sufficiently onboard the user, explaining what the assistant can do and suggesting what to ask them
- allow users to speak naturally and ask multiple questions at once
- can interpret a wide range of user requests
- provide quick and accurate responses
- allow users to have continuous interactions without having to invoke the wake word every time
- understand a range of dialects
- are able to keep track of linguistic context

These are tall orders for technology that's still evolving.

*1. https://hci.stanford.edu/research/speech/paper/speech_paper.pdf*

## Designing for effectiveness means designing for timeless human interaction challenges

Low effectiveness ratings are result of the gap between natural human interaction and human-machine interaction. In spite of changing technological circumstances, the ways humans naturally interact with each other in their environment presents timeless challenges to digital assistants. Brands that keep the following challenges top of mind while designing digital assistants will not only gain sympathy for the problems at hand, but also an appreciation of the vast opportunities ahead.

Let's explore these challenges and their implications for digital assistants.

### Challenge 01
## Humans don't express meaning in just one way

In user-centered design, it never gets old to be reminded that humans primarily think in terms of *what* they want to do, not *how* they want to do it. Humans express the same meaning in countless ways. In the example above, Ellen wanted ideas for dinner from the voice assistant. She played with different wordings to say what she perceived to be the same thing, and felt surprised that one worked and others did not.

| | |
|---|---|
| *Ellen:* | *Hey Google, what should I have for dinner tonight?* |
| *Device:* | *Sorry I don't know how to do that yet.* |
| *Ellen:* | *Hey Google, what do you think would be good for dinner tonight?* |
| *Device:* | *Sorry I don't know how to help with that.* |
| *Ellen:* | *Hey Google, what is for dinner tonight?* |
| *Device:* | *I was thinking noodles. I can recommend a good restaurant near you.* |

"You can see just the slight wording difference means I get an answer versus I don't. I have to speak in a specific way"
- Ellen, Participant

Depending on the choices of technology and design, platforms can vary in their ability to extract meaning, but users are not always made aware of that. As a result of not knowing the limits of an assistant, users may feel quite disappointed, use it once and not return to it.

Ellen was willing to try various ways of saying the voice commands until the smart speaker understood her. But novice users often won't do this work for the speaker. David tried once and gave up; as a result, the voice assistant missed the opportunity to show its abilities and so value.

| | |
|---|---|
| *David:* | *Hey Google, can you tell me the available flights from SFO to John Wayne Airport tomorrow, Wednesday September 12th?* |
| *Device:* | *Sorry, I don't know how to do that yet.* |

"It does not do travel."
- David, Participant

Challenge 02

## Humans are always context-aware and context-dependent

Humans are really good at knowing what information is useful and relevant in a given situation. We continuously update how we speak and act as contexts change. People want the same from their digital assistants. The more context-aware a digital assistants is, the more it saves us the effort to actively (re)enter contextual information to the assistant, and the more effective it is.

What does a context-aware digital assistant look like? Imagine browsing text or graphic heavy stock market charts on a website; you see a peak in price, move the mouse over it, and want to say 'Alexa, if the price goes below here (cursor location on the screen), sell.' For this to work, Alexa needs to see or minimally utilize the cursor location to know where 'here' is, and give a fully context-aware response.

Designing for the right mode of interaction is significant. For example, a user should have the options to input information to an assistant by text, voice, touch, or all of them, and receive information by either or all, as it befits the task. As digital assistants evolve, it is even more important to go beyond the mode, and design for contexts by utilizing data that the user provides and permits.

## Brands need to consider a number of contexts when designing digital assistants:

**Physical Context**

We use digital assistants differently in different physical environments - cars, homes, hotel rooms. This context is dynamic and complex. The boundary of private and public space is blurred even within one's home, a living room changes from private to semi-public if it has large windows on street level, whether we're alone, with family members, or hosting guests.

Physical location dramatically impacts how people communicate with a digital assistant - especially if they need to speak aloud. Add to this the complexity of moving from one space to another within the same task.

## Linguistic Context

Take a sentence like 'I put the raincoat in the bathtub because it was wet,' humans understand it immediately without even noticing the ambiguous reference of 'it.' Such linguistic context is still challenging to most digital assistants. In the example below, the voice assistant mistook Delia's command 'not right now' literally.

*Delia:* *Alexa, open translated.*

*Device:* *Here's Translated from the Education and Reference category. Welcome to Translated skill…….*

*Delia:* *Can you say 'Happy Birthday' in Hebrew?*

*Device:* *'Happy Birthday' in Hebrew is 'חמש תדלוה סוי.' You can say repeat, slow down, or just ask for another translation. What would you like to do?*

*Delia:* *Say 'my favorite food is' in French.*

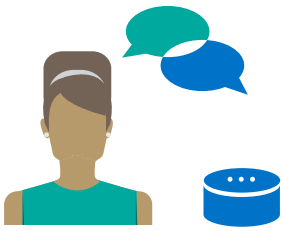
*Device:* *'My favorite food is' in French is 'mon plat prefere est.' Do you want to translate anything else?*

*Delia:* *Not right now.*

*Device:* *'Not right now' in French is 'pas tout de suite.' What else?*

## Social Context

Human interactions are inherently social, but interactions with machines are not. Cindy asked Alexa to set a reminder. Upon hearing the alarm, Cindy said 'Thank you' to Alexa, intending for the alarm to stop. Cindy was surprised that the alarm continued to ring.



*Cindy:* *Alexa, set a reminder at 2 pm.*

*Device:* *What's the reminder for?*

*Cindy:* *The interview.*

*Device:* *I'll remind you at 2 pm.*

*[At 2:00 pm, the alarm came up.]*

*Cindy:* *Thank you, Alexa.*

*[The alarm continued to ring.]*

What happened here? It is part of the western social and cultural convention for a human assistant to take 'Thank you' as an implicit signal to stop. Imagine that we set up a morning wake up call with the hotel concierge and the phone rings, we pick up the call, say thank you to the concierge, and the call ends. In contrast, Alexa is only set up to receive a command about the alarm, so it is not otherwise situated in the broader sociocultural context the user is in.

Challenge 03

## Humans don’t rely solely on words and sound in conversations

In natural human-human conversations, speakers and listeners use a variety of non-verbal cues like gestures, eye gazes, pointing, body and head movements, and facial expressions to communicate. In fact, research has shown you can tell two people are having a conversation just by looking at their movements - they synchronize even if they are not looking at each other!

What does this mean for conversation design? Brands should be aware that natural human interactions are multimodal. So reducing the multimodal sources of information to just words (in the case of talking to chatbots) or auditory information (in the case of talking to a voice assistant) is not only not a natural way to interact, but also can be difficult.

Timing is another non-verbal quality of natural human interactions. Conversations take turns. Turns are sometimes short, sometimes long. When two people are conversing, they often overlap with each other to show understanding, anticipate response, or to repair misunderstandings, all in the service of helping the other person feel connected[2].

Talking to a machine is different. Human-machine speech interactions consist of non-overlapping turns where a person says something first, followed by the assistant responding. And people find saying a wake word before each command awkward. It interrupts the flow and timing, making a conversation feel stilted. In response to this, Alexa enabled the ‘Follow-Up’ mode earlier this year. It allows the user to not have to say the wake word ‘Alexa,’ for up to 5 seconds after an exchange. A small technical advance, ‘Follow-Up’ represents a big UX improvement towards making device conversations more human.

Digital assistants in the future can perhaps tune in to our non-verbal and non-auditory behaviors by responding to qualities beyond sound or words, such as the tone, cadence, and even our facial expressions.

*2. https://www.theatlantic.com/science/archive/2016/01/the-incredible-thing-we-do-during-conversations/422439/*

# What does it look like when brands effectively meet these challenges?

User descriptions such as ‘quick and accurate responses,’ ‘understands me,’ and ‘I only need to ask it once’ were frequent responses to effective digital assistant interactions.

> “I’ve had many assistants in the past not being able to understand my voice or my dialect. Hound had no issues. I watched as it transcribed what I was saying in real time. It (Hound) was able to give me the information I was looking for the very first time.”
>
> - Karla, Participant

| | |
|---|---|
| *Ellen:* | *OK Hound, Are there any Japanese restaurants near me?* |
| *Device:* | *(Showing Japanese restaurants near me from Yelp)* |
| *Ellen:* | *OK, show me the ones that are more than 4 stars.* |
| *Device:* | *(Showing filtered results by 4 stars)* |
| *Ellen:* | *Navigate me to the first one.* |
| *Device:* | *(Navigating to the first restaurant by opening a dialog box for Google Maps)* |

The contextual knowledge of Hound isn’t perfect, of course. In the example above, when the user changed the wording of the second command to ‘OK, show me the highest rated restaurants on that list,’ Hound failed to keep track of the continuous context, took it literally, and pulled separate search results instead.

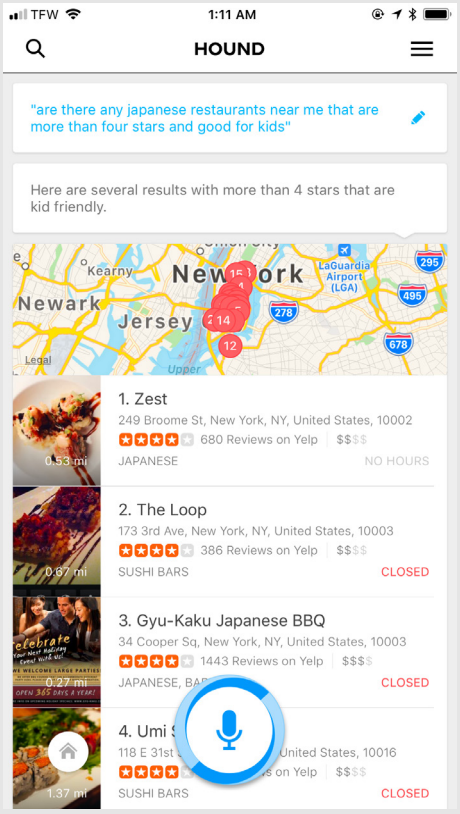
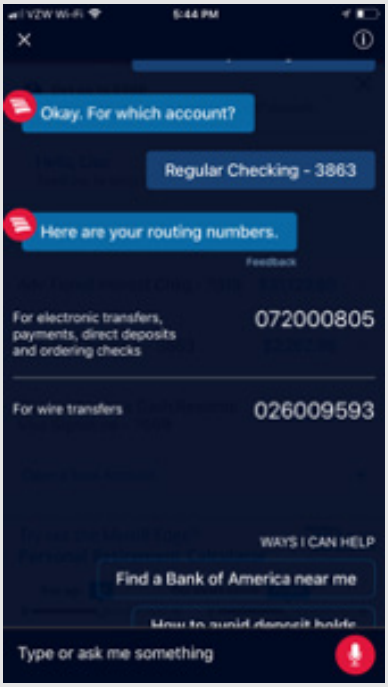


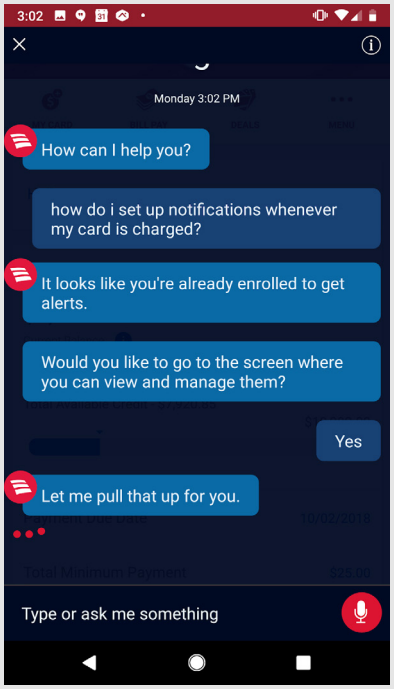

*Figure 4. Sample interaction with Hound*

Bank of America's Erica assistant interpreted a wide range of user requests. It asks relevant follow-up questions (Figure 5), looks up past user information to give proactive responses (Figure 6), and parses data based on customized requests (Figure 7). When Erica could not answer a user's question, it explained why and offered an alternative solution (Figure 8.)

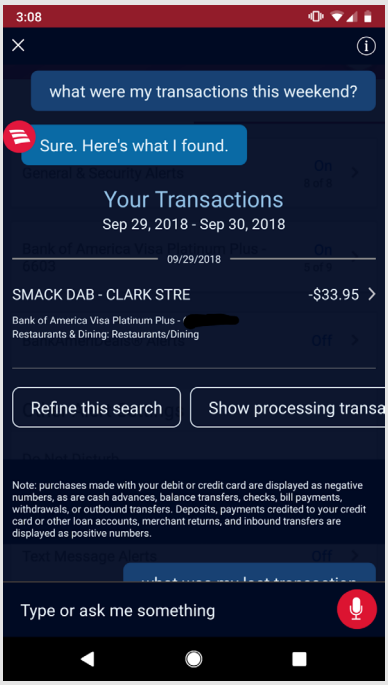
"Erica was able to answer almost every question I asked. The only ones it couldn't were because it had to send me to another page or for security reasons."

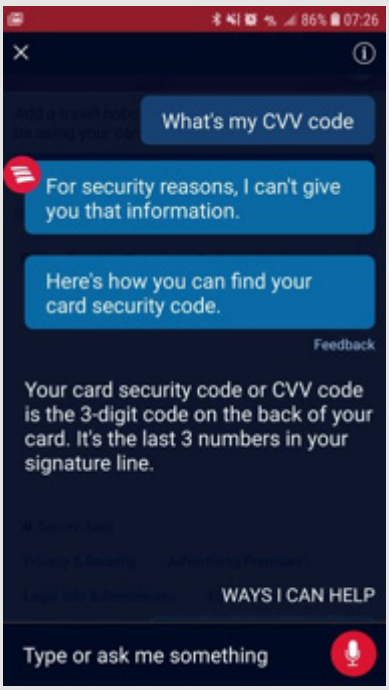
- Lisa, Participant




*From top left to bottom right: Figure 5, Figure 6, Figure 7, and Figure 8*



# What happens when a digital assistant fails to deliver on effectiveness?

Effectiveness failures may include not understanding the question, providing incorrect answers, or delivering a poor hand-off experience when the chatbot can no longer help.

In this example from American Express's bot on Messenger, the participant's questions weren't understood and they didn't get helpful answers. User perception of value and effectiveness are interlocked in this case.

While eBay's digital assistant received the highest rating for being fun and engaging, effectiveness ratings were low for both its voice and chatbot experiences. The complexity and ambiguity of shopping inquiries placed higher demand on the assistant's efficacy to not only get it right, but also be faster than existing, familiar functions, like filter.

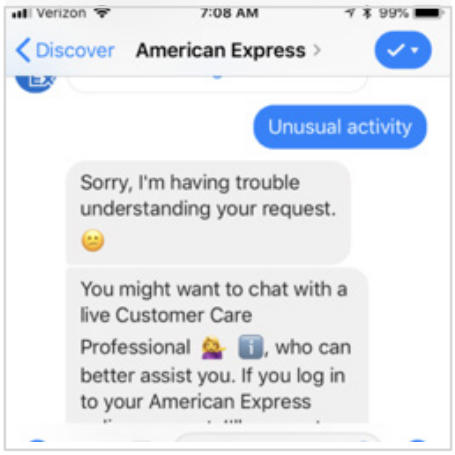



*Figure 9. American Express Messgener bot fails to understand the user's request*

> "While the Shopbot gave me options such as type of shirt (long sleeve, 3/4 sleeve, size, budget options), they were just filters rather than helping me determine or find relevant interests or brands faster. It also went back to women's clothing after I had mentioned men's in size small, and reminded it once already."
>
> - David, Participant

Participants also weren't impressed that eBay's Shopbot understood the meaning behind emojis like 'thumbs down,' but misunderstood their intent to 'sell an iPhone X' (see Figure 10, 11).

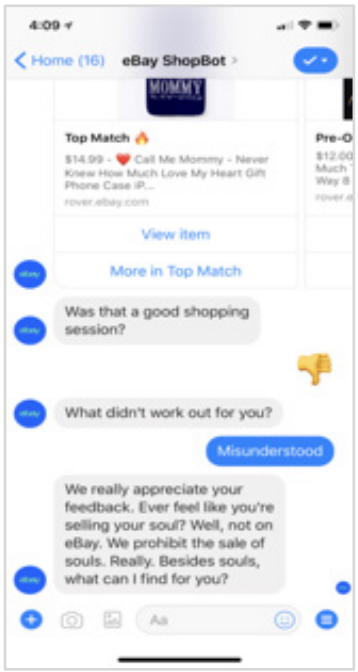



*Figure 10. eBay ShopBot's response to feedback*

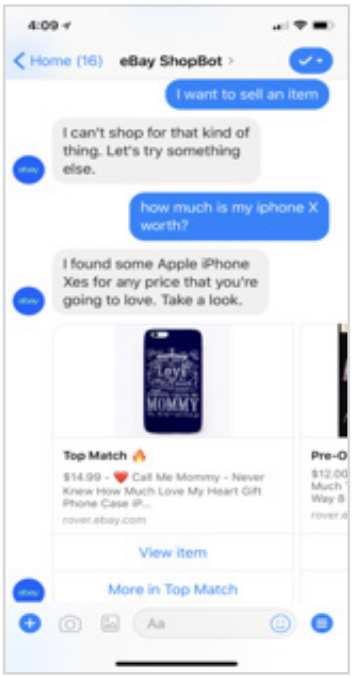



*Figure 11. eBay Shopbot fails to understand the user's question*

Poor hand-off experiences across platforms also figured heavily into whether participants found a digital assistant to be effective. Getting hand-offs right is especially critical along the shopping journey for retail. In the case of retail, digital assistants can help users browse, find, and occasionally, even do comparisons quicker (see the dialog with eBay's voice app on Google Home below). But in the evaluation stage, users often need devices with a screen to see product details.

*Ashish:* *A pair of Nike sneakers.*

*Device:* *Great, do you want in a shoe size men's that's 8, 8.5, or something else?*

*Ashish:* *Yes, men's.*

*Device:* *Oh, are we moving on?*

*Ashish:* *No. Men's Size 11.*

*Device:* *Got it. Looking for shoe in men's sizes 8, 8.5, or something else?*

*Ashish:* *Something else.*

*Device:* *Is it blue, red, grey, or something else?*

*Ashish:* *Blue.*

*Device:* *Got it, I looked through 2,421 options on eBay and the best deal on this item is $129.99. Need help with anything else?*

*Ashish:* *Yes please email that to me.*

*Device:* *What can I help you find today?*

*Ashish:* *That's it. Thank you.*

"The Google Assistant action wasn't effective because it found what I was looking for, and then "throw the fish back in the water". What I mean by this is that after finding the item, there was no option to add it to my cart or save it to my wishlist so that I can view it when I login back in on my phone or computer."

- Krystal, Participant

Recommendations

# How brands should approach designing for effectiveness

Effectiveness is crucial for user adoption. If a digital assistant cannot interpret questions correctly, keep track of conversation, and give accurate answers, it will not be successful. Users want a service that simplifies complex tasks, not one that makes their interactions more difficult.

For brands to ensure they're building a digital assistant that provides added value, we recommend:

01

**Know your competition.** Industry standards drive user expectations and the effectiveness of competitor experiences will impact how users perceive your digital assistant. Conduct competitive research to understand user expectations.

02

**Ensure have the right tool for the job.** Before you build a digital assistant, assess the available tools for their ability to meet your experience goals. Whether to use off-the-shelf chatbot tools (e.g., Microsoft's Azure bot service) or build from scratch, evaluate the tool against effectiveness.

03

**Get your users to talk back in order to learn their language.** When your digital assistant misunderstands the user intent or fails to respond in a way that keeps the conversation going, the solution often starts with word choice. Use methods like exploratory research and role-playing to uncover the users' word choices and what they expect in response.

AnswerLab®

# Dimension 3:
# Engagement

## Dimension 3: Engagement

> "If we design our AI to simply function well, our society may progress with increased speed in efficiency and convenience. But if we are also designing them to have thoughtful personalities and belief systems, our society may advance in areas where we have ostensibly made less progress—enhancing joy, delight, compassion, and deeper relationships."
>
> - Mariana Lin, former Creative Director at Apple and Writer for Siri

Engagement is a more nuanced dimension of digital assistants - and its use and success vary depending on a digital assistant's purpose. It is important that brands consider how to apply this dimension to be engaging, fun, original, and show signs of personality.

While less critical than value and effectiveness, engagement plays a key role in user experience over time. Participant feedback showed **engaging qualities currently function as 'buffers' for bad experiences or 'enhancers' of good experiences,** as opposed to being an essential element. This is only temporary, however, as it's influenced by the inadequacy of effectiveness and value in most digital assistants today. As we've shown, most digital assistants today only work well for simple commands and limited use cases. Keeping users engaged then becomes key to retaining users; or as many have argued, engaging qualities allows AI to 'save face' when it inevitably makes mistakes[34].

*3. Thies, I. M., et al. (2017). How do you want your chatbot? An exploratory Wizard-of-Oz study with young, urban Indians.*
*4. Pearl, C. (2016). Designing Voice User Interfaces: Principles of Conversational Experiences. " O'Reilly Media, Inc*

Moreover, users' changing expectations suggest engagement is rapidly becoming more important over time. In the first day of our study, participants listed three top qualities that make up a 'good' digital assistant in an open-ended question. At the end of our study, we asked them the same question again. Interestingly, personality traits that delight and engage users emerged as one of the new top three qualities. Participants didn't just want to go to a digital assistant for information; they wanted engaging qualities to enhance their experience.

In fact, Emma Coats and Danielle Kemmick, Principal and Founder of the Google Empathy Lab, said that their personality team explicitly uses the 'Yes, and' concept from theater improvisation to engage with the user. This means Google Assistant uses specific design tactics to respond in a way that keeps the conversation going, as opposed to shutting it down, and ultimately, fosters the user's desire for connection.[5]

The importance of designing for engagement goes beyond it being a stimulant or an incentive for interacting with digital assistants. As Mariana Lin pointed out, digital assistants, if designed with personality beyond function, can be a conduit for a more meaningful future of AI that inspires us for connection, reflection, and joy, beyond just faster task completion time.

"I felt like the Hound's strengths were that it was able to understand me and provide me the information I was seeking as quickly and efficiently as possible. However...it did not feel very 'personable'. This reduces my comfort level with the app, and makes me think of it more as a robot than my friend. This leads to lower trust in the app, which in turn leads to concerns about privacy."

- Darpa, Participant

*5. Danielescu, A., & Christian, G. (2018). A Bot is Not a Polyglot: Designing Personalities for Multi-Lingual Conversational Agents.*

# Should brands design for anthropomorphism?

Anthropomorphism - the word doesn't roll off the tongue, but the idea behind it is simple. We anthropomorphize when we ascribe human form or personality traits to nonhuman things. All humans anthropomorphize, though the things we anthropomorphize differ from culture to culture. We see faces in the clouds and treat pets as people with motivations and personalities. People anthropomorphize their digital assistants - especially voice assistants.[6]

Humans anthropomorphize by projecting our own stories, feelings, and thoughts onto things that are not us. We do so with nonhuman things and other people, including AI. There is nothing odd about this. Technology, like many things, is an extension of its user; it gives expression to a user's pre-existing experiences. On the flipside, the makers of AI technology have pre-existing values, stereotypes, and expectations about gender, race, and capital just as the users do, and so can inevitably embed these values while building a digital assistant. For the reasons above, the feelings we derive from our interactions with digital assistants can vary a lot.

Designers have explicitly utilized our anthropomorphizing tendencies to facilitate social interaction between robots and people, e.g., by adding features like eyes or a mouth to a robot head. Conversely, many brands intentionally avoid designing for human-like features. Is there a best practice approach?

*6. Barrett, L. (2011). Beyond the brain: How body and environment shape animal and human minds. Princeton University Press.*

Given users will inevitably anthropomorphize one way or another even with no degree of human features in the digital assistant, anthropomorphism itself isn't something designers can control. That said, there are clear **shoulds** and **shouldn'ts** when it comes to anthropomorphism:

- When designing interactions that are human, technology **should** behave as humanly as possible. This means mimicking how humans naturally interact with each other.

- We **should** pay attention to how anthropomorphism affects users' expectations of a digital assistant's capabilities and the emotional impact.[7] In our study, we saw that giving digital assistants human-like features increased users' expectations of a system's overall competence, which led to higher likelihood of user frustrations than if the digital assistant was perceived as less human-like, e.g., by disclosing its identity as a bot upfront. Such features could be as simple as the choice of wake word for a voice assistant.

- We **should never** imply that the technology is human. Deceit is not only unethical but also a key eroder of trust. We saw this with strong negative reactions people had when Google Duplex did not disclose it wasn't a person - and Google's subsequent decision to clarify that a bot is a bot.

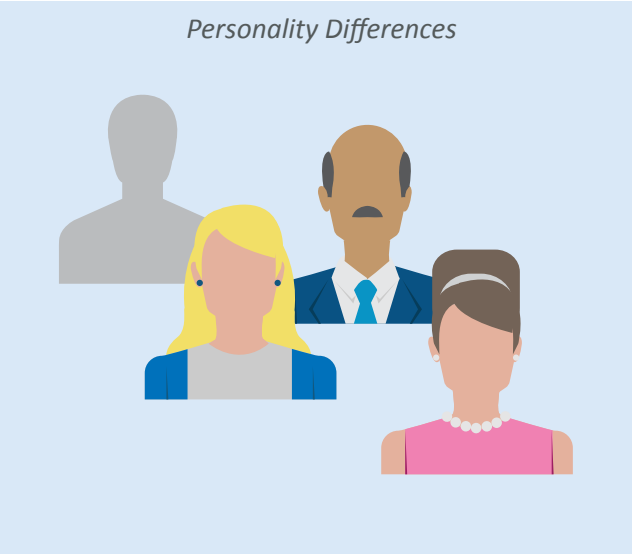
In this quote, when the participant called Alexa by 'Alexa' instead of a more neutral term 'Computer,' he reported a higher expectation for Alexa's abilities and so a higher chance to feel disappointed.

> "[The wake word] Alexa makes it sound like a person, so when it makes a mistake, I feel mad and I yell at it. I feel like I'm being rude to a child. When I changed its name from 'Alexa' to 'Computer,' I'm not mad at it when it makes mistakes."
>
> - Morteza, Participant

**01**
**Users notice personality differences of digital assistants**

One question is whether users can reach consensus for the differences in digital assistants' personalities. The answer is yes. Participants were keen to notice and often commented on signs of personality and the lack thereof in the digital assistants in our study.

Assistants that were designed with a specific personality easily stood out, partially because participants felt that most automated response systems that they experienced in the past did not show personality.




*7. Duffy, B. R. (2003). Anthropomorphism and the social robot. Robotics and autonomous systems, 42(3-4), 177-190.*

## 02
## Brands should design around components of personality

Participants honed in on the following components of personality for digital assistants: personality traits, the choice of voice, and gender. Here, we explore the component that's applicable to both chatbots and voice assistants, namely, personality traits and gender. In addition to traits and gender, the acoustic characteristics and gender are both key for forming impressions of a digital assistant, especially when voice is the primary mode of interaction. These are key choices that brands need to make when creating an audio persona of their digital assistants. For example, studies show that when we hear a new voice we automatically form a 'first impression' of the voice owner's personality.

In the coming months, we will release a full report on what brands need to know about personality and digital assistants. Stay tuned to our Insights blog to hear about its release.

## 03
## Personality Traits

Take the intended personalities of Alexa, Microsoft's Cortana, Google Assistant, and Siri for example. As expected, they were all created to be friendly and helpful, as these traits are basic requirements for an assistant that needs to work for a large audience. A few other universal basic requirements of an assistant included being impartial and always enthusiastic.

In addition to the universal traits, subtle differences exist. For example:

Users experience **Alexa's** personality through its telling jokes and Easter Eggs.[8]

Siri was perceived as friendly, but also with an edge and has a sassy side.[9]

**Google Assistant's** personality team has publicly emphasized the need for personality, in that "having no personality was actually a personality in itself."[10] They described Google Assistant as a "cool librarian."[11] Users often perceived Google Assistant as being conversational, approachable, and inviting.

Microsoft devoted serious efforts to crafting **Cortana's** personality. The choice to make Cortana not sound too human was deliberate, so as to not invite unrealistic expectations, as we discussed earlier. In the design process, Microsoft talked to users about their ideal personal assistant worldwide, including conducting cultural testing.

When we talk about brand personality, brands are often treated as people. We may evaluate a person's personality traits around different domains such as competence, sophistication, sincerity, and maturity. Users may perceive a brand similarly, using adjectives such as sophisticated, creative, offbeat, elegant, luxury, down-to-earth, etc. Just as consistency is key to evaluating the stability and trustworthiness of a person's personality, establishing a consistent brand experience across channels - including that of digital assistant - is important for building trust and customer loyalty.

*8. https://www.businessinsider.com/amazons-alexa-has-a-different-personality-in-europe-here-are-some-of-her-traits-2017-6*
*9. Fowler, G. A. "Are Smartphones Becoming Smart Alecks?" WSJ. October 15, 2011.*
*10. http://time.com/4979262/google-wants-to-give-computer-personality/*
*11. https://www.cnbc.com/2018/09/01/ryan-germick-leads-google-assistant-personality-or-home-smart-speaker.html*

## 04
## Word Choice and Tone

There are limited tools for expressing digital assistant personality traits:

- word choice and the voice itself if voice-activated
- word choice alone if text-activated

Word choice is intimately bound up with the tone it conveys. Tone refers to the feel of the response: Is it informal or formal, does it feel approachable or cold? The factors that influence our perception of tone are varied and subtle, yet were quite noticeable to participants. For example, in the greeting message received from Capital One's chatbot Eno, the participant marked on the paper printout what exactly conveyed the tone that she liked in an interaction she had with Eno.

"First things first"…is conversational...If it says 'Here is the first step,' it would have a transactional, stiff feeling."

- Laura, Participant

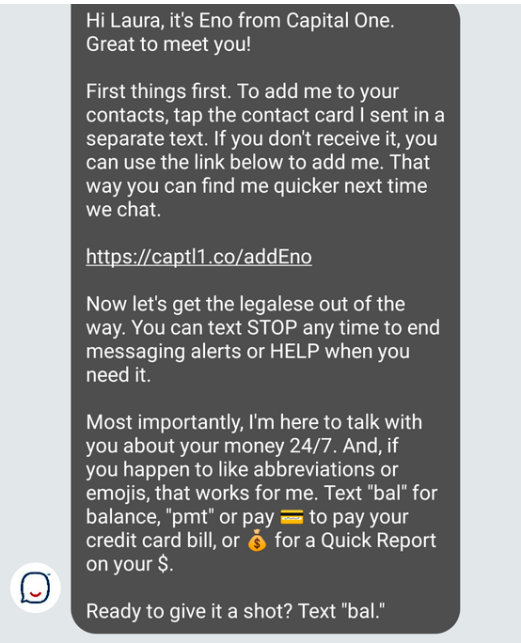



*Figure 12. Example of Capital One's Eno's conversational interactions*

## 05
## Gender

Whether to give a digital assistant the female or male gender is a popular topic this year. As we discussed previously about our ingrained need to anthropomorphize, the choice of gender for a digital assistant invites questions and concerns. For example, in the top search result for '[Why is the digital assistant's voice always female]', the author urged us to challenge the stereotypical gender roles as represented by the predominantly female-voiced digital assistants. Indeed, there is a rich history of the role of women in domestic labor, where jobs such as personal assistants were frequently relegated to women.

Female-voiced digital assistants trigger that association. Yet, consumer research in digital assistants showed that cross-culturally, it is much easier to find a female voice that has mass appeal than a male voice[12]. To reconcile the public's needs for a gendered voice with our responsibility to not encourage gender binary expressions that are ever-changing and disappearing in the western society, a simpler solution in the near term is to provide options. Siri and Google Assistant have both offered the option to switch to male voice options, in addition to having female voice only.

*12. Clifford Nass. 2005. Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship. MIT Press.*

# When it comes to personality, trust is a must.

In in-depth interviews following our diary study, we asked participants to consider, describe, and rank digital assistant personality traits. Participants identified the top, must-have qualities they would want a digital assistant to have, for banking, retail, and healthcare tasks respectively, as well as the qualities they did not want.

The results revealed rich insights about the traits that users want their digital assistant to display, and why.
Here are the highlights:

Key industry differences exist. For both **retail** chatbots and voice assistants, participants **expected more engagement** than they did from financial service assistants - those that conduct digital banking, credit management, or investment tasks. For the latter, effectiveness without displaying any signs of personality didn't damage the experience. However, effectiveness combined with engagement exhilarated the participants.

Across industries, **everyone chose trustworthiness as a must-have quality for digital assistants**, whether it be retail, banking, or healthcare. It is important to note that trust is certainly, more than just a personality trait. Trust summarizes the entire ideal experience between a person and her digital assistant. In a sense, everything we've discussed thus far - improving value, effectiveness, and playing with the elements of engagement - are all in service of building trust between the user and the assistant. So, removing the barriers of trust with the user is the crux of being a successful digital assistant.

Many principles for building trust in human-human relationships apply to human-machine relationships. But differences exist between the kind of tasks participants would trust AI with versus a human. When asked if there were tasks that were inappropriate for a digital assistant to do, participants frequently noted that they did not trust an AI assistant to handle high-stakes tasks that they would otherwise trust a human assistant with, e.g., composing an email to a superior at work, diagnosing mental health issues, or assessing FICO score. This is not surprising. As we've discussed, participants were especially intolerant of failures of digital assistants at tasks that have high social, financial, and health risks.

"I would never trust a suicide prevention bot. That's too risky."

- Veronica, Participant

Imagine trust as the connective tissue that ties the user to the product. These high-stakes tasks are the fragile spots that require more trust development to prevent user drop-off. The following example illustrates an interaction with Ada the healthcare chatbot app that led to loss of trust. The user didn't believe the accuracy of the answer and as a result, abandoned the task.

"I happen to have a skin condition I can describe plainly and know the cause. After a lot of questioning, the bot gave a vague description it has in its database, which is incorrect. It also says it typically affects people over 40. But I'm not yet over 40. I'm wary of using it and will go to other sources instead."

- Jeanene, Participant

# How are brands meeting expectations for engagement?

**01**
**Creating fun, but relevant interactions**

Brands who are successfully creating engaging digital assistants are meeting the bar for value and effectiveness first. Engagement must be added once you've hit the threshold for the first two dimensions. Successful examples of this engagement vary depending on the types of tasks users are trying to complete and the brand's industry.

Both eBay's chatbot and its voice app on Google Home received the highest rating for being fun. A user described her experience as the following:

> "It was fun to interact with eBay's chatbot. It had huge amounts of personality. When I asked for a watch to buy, it told jokes such as 'A watch is sold every 3 seconds on eBay. About time! Get it?' And then it would go on to ask me what exactly I was looking for. When I asked it to look for laptop stickers it said, 'Ask and you shall receive.' It just made interacting with the chatbot more fun and made me want to keep spending time with it."
>
> - Karla, Participant

| | |
|---|---|
| *Morteza:* | *Hey Google, ask eBay to search rainbow pillows for kids.* |
| *Device:* | *Something you already own or something you want to purchase?* |
| *Morteza:* | *Rainbow pillows for kids?* |
| *Device:* | *Do you want black/blue or something else?* |
| *Morteza:* | *Something else.* |
| *Device:* | *Help me improve. What went wrong?* |
| *Morteza:* | *Nothing.* |
| *Device:* | *Great. Do you want shape square, rectangular or something else?* |
| *Morteza:* | *Something else.* |
| *Device:* | *I looked through 145 items on eBay. The best is 9 dollars and eighty cents. I've sent it to your phone. Go ahead and check and I will talk to you there.* |

In addition to using witty one-liners throughout the user's shopping experience, eBay's voice assistant established trust in an important moment of the shopping journey. In the following example, the assistant not only narrowed down the search results over voice by asking relevant questions, but also provided sufficient criteria for that process, i.e., 'I looked through 145 items on eBay.'

> "The question about which room the pillow is at is brilliant. I would have never thought about it. Also, the last line where it (eBay) told me that it looked through 145 items on eBay helped build confidence. That's what made it trustworthy."
>
> - Morteza, Participant

## 02
## Setting expectations to build trust

In this example, the participant cited words such as ‘Okay, last question’ as evidence for setting expectations and ‘Don’t forget that this is not a medical diagnosis’ as a sign of conscientiousness; both functioned as trust boosters for Ada, the healthcare chatbot.

Interestingly, Ada scored high on the engagement dimension but the majority did not interpret engagement for Ada the same way they did for a retail assistant, i.e., fun, snarky, and unique. Accuracy, privacy, thoroughness, and care make up the definition of engagement for a healthcare digital assistant. Many liked that Ada reminded them that it’s not a doctor and urged the user to seek medical advice as well. This generated more interest in using Ada as a secondary source of self-diagnosis for immediate insight on their symptoms.

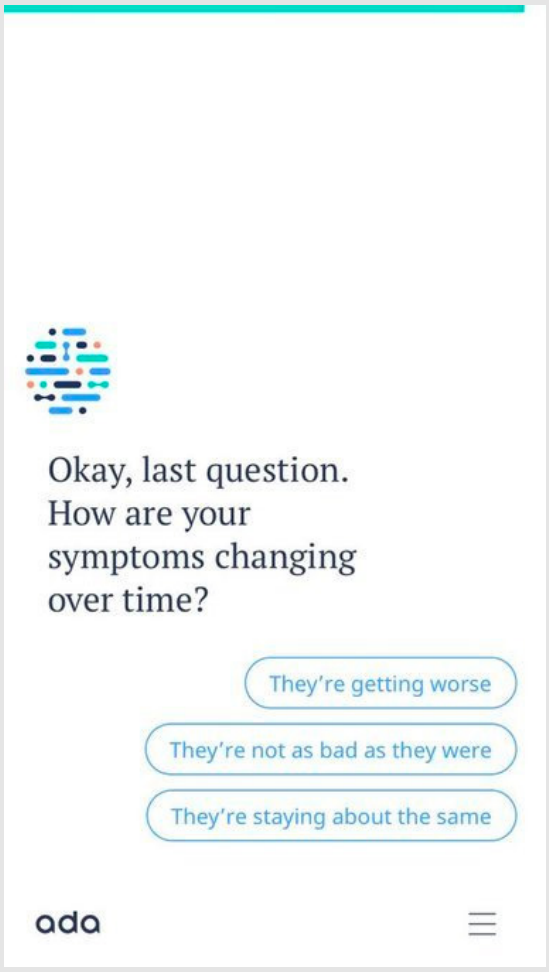



*Figure 13. Ada builds trust by setting expectations*

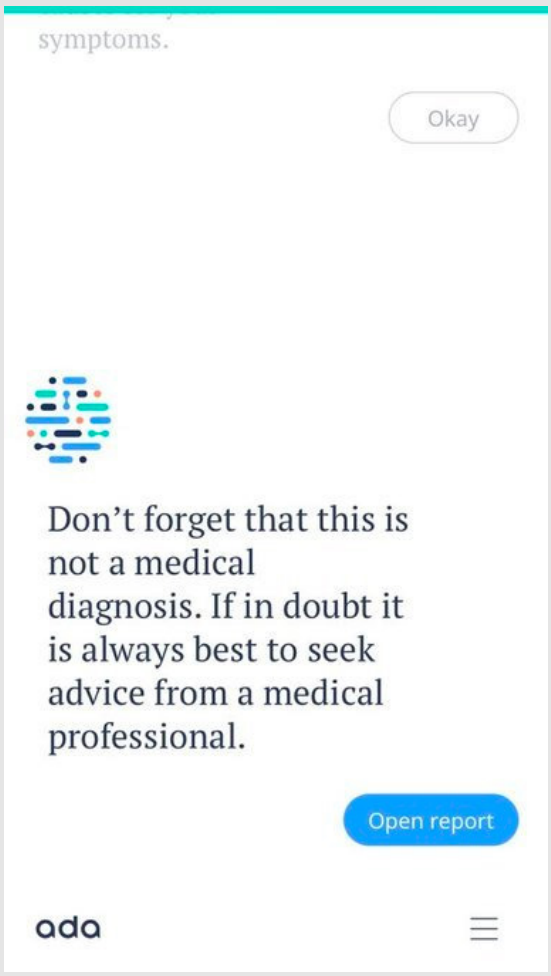



*Figure 14. Ada reminds users of the technology’s limitations*

# What does it look like when brands poorly execute engagement elements?

**01**

**Failing to balance effectiveness and personality**

A successful deployment of engagement in today's experience boils down to the balance between effectiveness and personality, in addition to the context of use, such as industry differences. When the digital assistant does not meet user expectation in terms of functionality and added value, being fun and engaging can hurt the experience and cause mistrust. Many participants pointed out moments where engagement in the absence of delivering added value rang hollow. For example:

> "There was quite a bit of banter for the eBay chatbot - almost to a fault. The extra 'filler conversation' was a little distracting - 'too long, didn't read' is a thought I had as I literally scanned the replies for actual information relevant to my search."
>
> - Karla, Participant

As another example, the participant was searching for a couch that costs less than $2000 in the screenshot above. He was eager to narrow down the search results. When the Shopbot failed to understand the 'smaller than' sign in the user command, it delivered a hand-coded joke, i.e., 'Hmm I don't know what that means. I only speak English (and a bunch of coding language),' before repeating the question again. The participant was angry and the joke fell flat.

Naturally, when an assistant fails functionally, the participants wants to know why it did not work out and how they can ask it differently in the future. It may be too costly and impossible for AI to self-diagnose and serve up the precise feedback whenever a mistake occurs. But there are opportunities for writing better error messages that can placate the situation and invite the user to try again.

> "When there is a mistake, I love that Google Home says 'I can't do that right now.' That's better than Alexa saying 'I'm not sure if I get that.' **The former is more informational and humble."**
>
> - Jeff, Participant

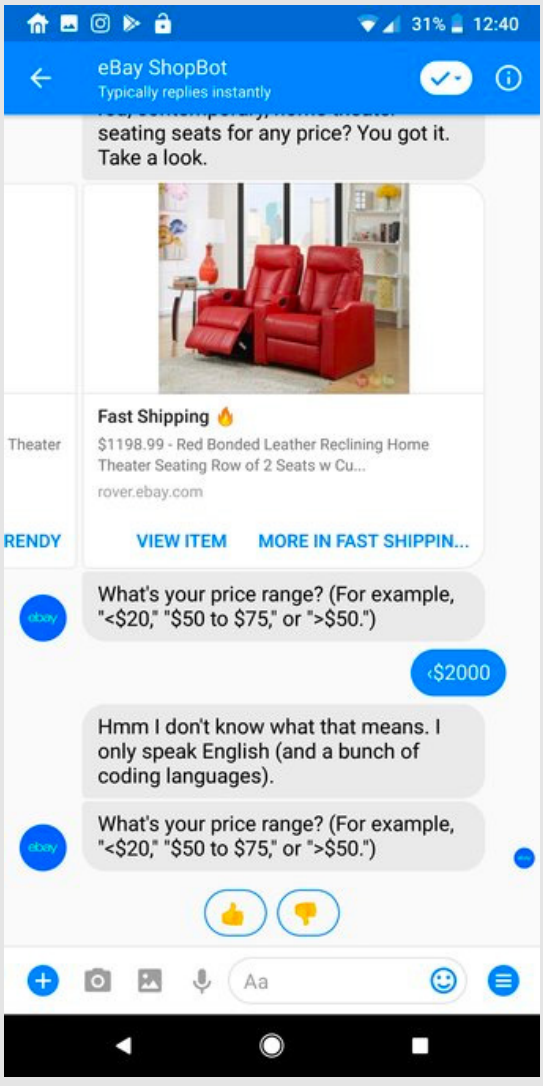



*Figure 15. Sample failed interaction with eBay ShopBot*

**02**

**Privacy concerns and low risk tolerance**

Participants showed lower risk tolerance when a financial services assistant is built on a non-branded third-party platform. Like many such assistants, American Express's Facebook Messenger chatbot required the users to login to their banking account first. Participants commented that this did not make their existing means of getting tasks done easier; they said that they may as well login on American Express's own website or app, as opposed to handing over sensitive financial information to Facebook.

In the following example, the participant connected his American Express account to its Facebook Messenger bot for the first time. Upon doing so, he saw the message stating 'Only send me information you're comfortable sharing with Facebook or anyone else who could see this conversation.'

The participant called out the verbiage 'anyone else who could see this conversation' as vague, which diminished trust in the bot and his willingness to use the bot for financial tasks.

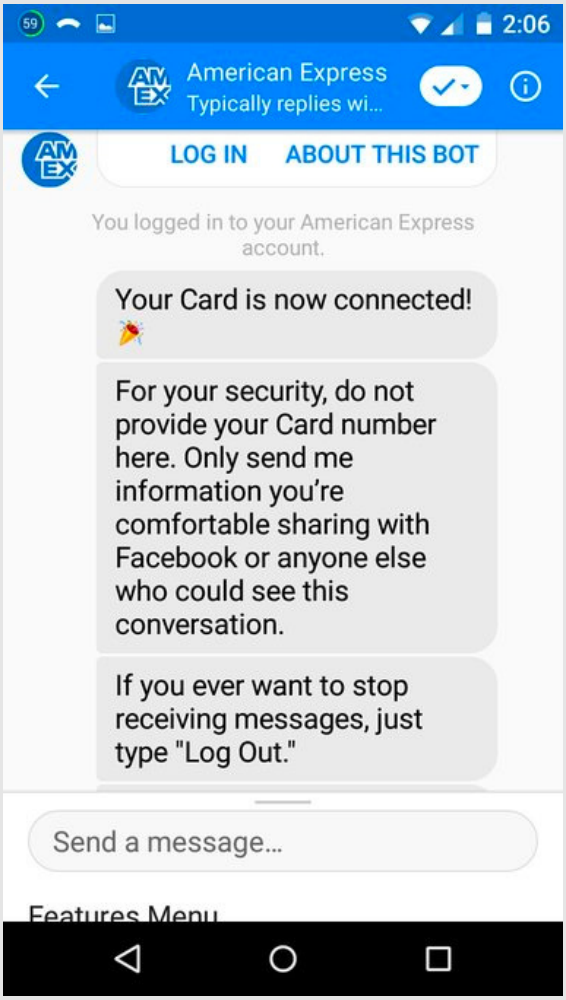



*Figure 16. Interaction with American Express' Messenger bot that decreased trust and raised privacy concerns for the user*

Recommendations

# Design for engagement by understanding how users perceive personalities

To implement engaging elements effectively, ensure you're meeting expectations for value and effectiveness first. Then, consider how users perceive personality traits based on the job to be done, industry, and audience.

To design for engagement, we recommend:

01

**Remember anthropomorphism is unavoidable and that's OK.** Designers can't control anthropomorphism, but should balance the human-like features to either meet or surpass user expectations of how capable the digital assistant is.

02

**Respect the audience.** Tailor the personality of the digital assistant to its audience. It's a bit like making a studio film vs. an independent film. There is more flexibility and freedom to express personalities with a specific audience than a broader audience.

03

**Strike a balance between personality and effectiveness.** Balance the amount of personality with the level of effectiveness of the digital assistant. Currently, a capable assistant without personality is preferable to an incapable assistant with personality. As the technology advances, adapt the personality design to new possibilities.

04

**Build trust and maintain it.** Trustworthiness is the universal trait users want. Boost trust by sharing what the digital assistant needs to know about the user and why, providing a sufficient but not unnecessary amount of confirmation, and using error messages as an opportunity to provide useful information and redirect the conversation.

05

**Do cultural research.** Before launching in an international location, conduct user research in the cultural context of the audience to understand the cross-cultural differences in user preferences for its personality.

06

**Do research on the perceived personality of your digital assistant.** As with all social interactions, many factors such as the distribution of participants' personalities, gender, work background, expectations will influence their judgment of the personality of the digital assistants. For example, an introvert user might find it easier to interact with a low pitch voice assistant than a high pitch one, compared with an extrovert user. Make sure to know your participants as much as possible, and aim for a large sample size to minimize such effects.

AnswerLab®

# Appendix

# About the Research

A comprehensive review of domain-specific research and industry research showed a lack of and desire for qualitative, ethnographic studies to uncover the attributes of a digital assistant, based on user's needs, motivations, and perceptions. We used this literature review to frame the hypothesized elements while asking participants to add elements by providing open responses.

We conducted searches of Google Scholar, EBSCO Host, and Google from 1997 to 2017, using keywords such as 'conversational agent' 'chatbot,' 'voice assistant,' 'social robots,' and 'multi-modal interaction.'

The chatbot brands tested included American Express, Bank of America, Capital One, AskTrim, Ada, Domino's, eBay, Sephora, and the Loebner Prize winner chatbot Mitsuku.

The voice app brands tested included Domino's, eBay, Starbucks, PayPal, American Express, Capital One, Progressive.

Other brands tested: Hound, Cortana, Google Assistant, and Amazon's Alexa.

## Nationwide Diary Methodology

- N = 37
- All own and actively use a smart speaker device, such as Amazon Echo (N=18) and Google Home (N=19).
- All have set up at least two voice apps (skills for Amazon/actions for Google) on their speaker.
- All were open to installing and using new voice apps in healthcare, financial services (digital banking, credit management, digital payment, investment, insurance), and retail industries.
- All were open to installing and using new chatbots[12].

## Recruitment Criteria

- Age: 21-54
- Even gender split
- Mix of household income
- Mix of education levels (HS or higher)

*12. The user-facing definition we used in recruit email for chatbot is that it is a computer program that one can text or speak to, frequently encountered in customer service, but not limited to that.*

# About the Team

### Lin Nie
User Experience Researcher and 2018 Research Fellow

As the 2018 Research Fellow, Lin Nie led AnswerLab's original research efforts exploring topics of voice and artificial intelligence and sharing insights with our clients. Lin is a UX researcher at AnswerLab where she leads research to help fortune 500 brands create experiences people love. Prior to AnswerLab, she consulted for Amazon and led foundational user research for startups. Lin has a PhD in Experimental Psychology and spent nearly a decade studying what AI can and cannot do in the context of human perception and cognition, along with notable figures in the field including the American Psychologist James J. Gibson, who invented the 'affordance' concept. Her research in cognitive science and artificial intelligence has appeared in Wired and Slate.

### Una Casey
Research Project Coordinator

Una manages logistics of user research projects for Fortune 500 companies across a wide variety of industries, including e-commerce, retail, finance, and technology. As project coordinator for AnswerLab's 2018 research fellowship, she fielded all participants, programmed and monitored the diary study, and coordinated all logistics for in-home research visits. Both an analytical and creative problem solver, Una has 3 years of consumer marketing experience.

# About AnswerLab

AnswerLab delivers trusted customer insights that help the world's leading brands create experiences people love. We focus exclusively on user experience research to understand what people see, do, think, and feel when using voice interfaces, AR/VR, websites, mobile applications, wearables, and other digital products.

Through thousands of research engagements, we build strategic partnerships that help clients scale their UX impact year over year. To answer your toughest research questions, we employ a mix of techniques, methodologies, and best practices tailored to provide clear, actionable recommendations that generate strategic business results.