



Openstream.ai[®]
Delivering Intelligence



EXCLUSIVE
CHECKLIST
REPORT

The Rise of Chatbots and Interactive Virtual Assistants

Let's Talk: Building Conversational AI Systems That Work

By Marydee Ojala, Editor-in-Chief, *KMWorld*

Chatbots and IVRs (interactive virtual assistants) have become essential tools for customer service operations. Paralleling their increasing ability to deliver accurate, even insightful, answers to customers' questions and concerns is people's acceptance of interacting with a computer. As we move toward a self-service world in many aspects of modern life, such as pumping our own gas, ordering restaurant food via a touchscreen, and online purchasing, it's no surprise that chatbots are now considered a normal addition to customer service and customer support.

At one point in the early incarnations of chatbot technology, developers worried about how people would react to talking with a computer or interacting with a robot. Would they reject the concept? Would they opt for human-only communication? Was creating chatbots a waste of time? Those initial fears have proved unfounded. In fact, we talk to computers all the time now. It's considered normal. We've entirely accepted hearing computerized voices from our GPSs. Cute little robots are popping up in museums, libraries, and businesses to answer visitors' questions. They are particularly popular with children.

Plus, as our phones become more about computer-driven activities than actually placing a phone call, the line between speaking to a person and speaking to a computer has become very blurred; most of us don't see a huge distinction. We can easily check our bank balance, pay a utility bill, or find a discount coupon on our phones. As long as we get the correct answer to a question, as long as we obtain a solution to a problem, as long as we can place an order, we really don't care about the underlying mechanism. If it's an individual, that's fine. If it's a chatbot, that's fine too. Many times, for simple requests, the using the phone as computer is faster than using the phone for calling and talking to a person.

Not only are we fearless about chatting with computers and robots, but we also actively invite conversations with devices in our homes and offices, asking our virtual assistants what the weather is going to be, how to travel from one place to another, how much something costs, what the lyrics are to a particular song, how to cook what we just found in our refrigerator, and a host of other both trivial and nontrivial topics. We say "I talked to John" when we actually texted, emailed, or read and responded to a social media post. John might not even be a person, but a bot.

GENERATIVE AI

The recent frenzied attention that ChatGPT, a generative AI chatbot created by OpenAI using a large language model (LLM) based on GPT 3.5, has garnered raises the bar for chatbots in general. Suddenly, the perceived potential for having a conversation with a chatbot increased exponentially. Experiments with asking ChatGPT to write poems, school papers, resumes, course syllabuses, instruction manuals, computer code, emails, social media posts, short stories, and explanations of complex topics abound. You can engage in a conversation with ChatGPT, ask it for advice, and have

it refine its original responses. Some of these experiments and conversations are truly astounding, with actual utility in the real world. They can spark new ideas, suggest alternative ways of phrasing written documents, and debug code.

What is ChatGPT not good at? Turns out, many things. Its training set ended in 2021 so don't bother asking it to comment on current events. More importantly, it has a tenuous relationship with truth. It can compile information from many sources and present them to humans in a human-like way, sounding as if a human consulted all those sources, and wrote a report based on human learning. But ChatGPT doesn't actually consult sources, and it can't tell the difference between what is true and what is false. This is largely due to its training set being extremely broad, consisting of many varied sources of information that's freely available on the open web. This is very different from a chatbot trained on internal information, which gives it the domain knowledge lacking in ChatGPT. When it is deployed to help employees and customers in a conversational manner, it performs well.

"You can engage in a conversation with ChatGPT, ask it for advice, and have it refine its original responses. Some of these experiments and conversations are truly astounding, with actual utility in the real world. They can spark new ideas, suggest alternative ways of phrasing written documents, and debug code."

ChatGPT also can't show empathy. It doesn't understand emotions, so it can't utter phrases such as "I'm sorry." The notion of being sorry is beyond its skill set. Ask it to explain the concept of being sorry and it can do that, but it can't actually feel sorry for anything. Suppose a customer calls to order a new phone. The bot first wants to know why a new phone is needed. Maybe the customer simply wants the latest and greatest model. The chatbot could say, "Great, that's exciting. Here are some options for you." But what if the previous phone was lost or stolen? Responding with "Great, that's exciting" would be inappropriate. An empathetic response would be "Sorry to hear that. Do you want to replace it with what you had or order a newer version?" Humans are good at this. Computers need to be programmed to respond appropriately and with empathy.

“As chatbots become smarter, however, their design becomes more dynamic. The ability to hold a conversation, to engage in the back and forth, and to build on one comment to intelligently move the conversation forward was beyond the capabilities of early chatbots. Now it’s becoming ever more possible.”

CONVERSATIONAL PITFALLS

Conversations can be full of other pitfalls that computers might fail to understand. Traditional chatbots have worked well in answering straightforward questions with factual answers. Ask them about what hours a store is open or location addresses and the answer is immediate and accurate. It gets a bit more complex when the question is about a product. The chatbot might know in general about products offered, such as dresses are on the second floor, but if a specific dress comes in purple in a size 5 is probably beyond its knowledge. In that case, good chatbot design escalates the question to a human.

What does having a conversation entail? It could be personal, as in “How’s the family?” It could be making small talk, as in “How about them [insert sports team name here]?” Business questions are different. They are purposeful. Chatbots and IVRs deal with customer inquiries and complaints quickly and efficiently, which is what customers want. However, some queries are more complex than others, resulting in longer interactions. Add ambiguity to the mix. While a term in a query may be perfectly clear to one person, that doesn’t mean it’s understood by the other person—or bot—in the conversation. The classic example is a soft drink that could be a soda, a pop, or a sodapop, depending on what part of the country you’re from. Even more specific, a milkshake in some parts of New England lacks ice cream. If you want a traditional milkshake, it’s often called a frappe in Boston and a cabinet in Rhode Island.

Conversations involve some back and forth. A customer might ask about a lawnmower part. The conversation could delve into make, model, and age of the lawnmower. Depending on the customer’s answer, the next step in the conversation would be about the nature of the problem for which the part is necessary, availability of parts, and options for delivery and payment. Conversations are rarely of the “ask a question, get the answer, end of transaction” type.

QUESTIONS IN CONVERSATIONS

Computers are literal; humans are not. A simple question about opening hours could be phrased as “What hours are you open?” “When do you close?” “Are you open on Sundays?” “Has yesterday’s hurricane shut you down?” These are all the same question, so the chatbot needs to recognize the various permutations of human questions. Just as there are variations in how people ask questions, the rationale for why a chatbot asks questions is another conversational pitfall. Take age, for example. A chatbot might ask “How old are you?” Not the most polite of questions and likely to cause a negative reaction. “This site is age-restricted. Are you over 18?” is a better way to phrase it. A medical site might ask if you are over or under a certain age because some vaccines require different dosages based on age. Again, the explanation of why the bot (or a person) is asking the questions removes any reluctance about answering it.

Human conversations also involve nonverbal aspects. A shrug of the shoulders or an eye roll can be just as telling as the actual words being used. Intonation is another clue toward meaning. Shakespeare’s six word utterance by Hamlet, “to be, or not to be,” takes on different interpretations depending on which word is emphasized. Even a simple “oh really” could express curiosity or disbelief, depending on tone of voice.

CONVERSATIONAL AI

A basic premise of conversational AI is interactivity rather than a static approach (for instance, FAQs). An FAQ was the starting point for traditional chatbots and still has value for situations that call for simple responses. However, as chatbots become smarter, their design becomes more dynamic. The ability to hold a conversation, to engage in the back and forth, and to build on one comment to intelligently move the conversation forward was beyond the capabilities of early chatbots. Now it’s becoming ever more possible.

This all begs the question of whether customers prefer chatting with a bot to having a conversation with another human being. It depends. In a grocery store, for example, sometimes you’re in a hurry. You want to buy three items quickly and leave. You go to the self-service area, scan your items, pay the bill, and leave. Other times, it’s a more leisurely trip. You have time to chat with the cashier and the human interaction makes you happy. There’s a grocery store chain in The Netherlands, Jumbo, that has introduced KletsKassa (chat checkout), which a lane for customers who are not in a hurry and prefer spending some time talking with the cashier.

Looking ahead, the power of chatbots and IVRs as integral elements of customer service and support will only increase. Generative AI holds great promise for chatbots that can hold meaningful conversations with customers, provided they are trained on a set of relevant documents. Using a plan-based collaborative multimodal dialog system increases conversational coherence. Particularly when chatbots are designed to follow a predetermined plan to achieve a specific goal—which would be developed internally and be unique to each individual company using chatbots in conversational mode—they are more accurate and flexible than a more general application such as ChatGPT. It appears we will be effortlessly talking with our devices for a long time to come. ■



Marydee Ojala is editor-in-chief of *KMWorld* magazine and is Conference Program Director for Information Today, Inc. She works on conferences such as Enterprise Search & Discovery, *KMWorld*, Data Summit, and AI & ML summit. Ojala is a frequent speaker at U.S. and international information professional events. In addition, she moderates the popular *KMWorld* webinar series.

The Rise of Chatbots and Interactive Virtual Assistants

By Raj Tumuluri, CEO, Openstream.ai®

INTRODUCTION

As we're getting ready to go to print, ChatGPT from Open.ai was opened to public and has impressed with its uncanny ability to generate text when given a prompt. It also claims to be able to engage in conversation with humans. We decided to assess some of the limits of its ability to converse, comparing it to another approach to dialogue, specifically Openstream's Eva, a plan-based collaborative multimodal dialogue system. A more detailed analysis is available in our blog at: <https://openstream.ai/chatGPT-and-dialogue>.

We have seen tremendous progress recently employing Conversational AI technology across a variety of industries and related use cases. Use cases vary from simple text or voice-based chatbots answering FAQs to virtual assistants helping users accomplish their goals to embodied virtual agents which reflect human empathy.

However, until recently, endowing conversations with virtual agents that are more natural, going beyond classically limited intent+slot models, has been elusive. Many of these bots/virtual assistants only allow for atomic values as inputs. True human conversations vary well beyond the fixed sets of answers to questions and instead are typically more complex, introducing one or more constraints on those values.

For a system to be able to hold a multi-turn conversation, it must track context, adapt to changing directions, and understand the multiple intents that could be present in a single turn. Current dialog systems are very limited in their capacity to do so and only in select scenarios.

However, in real life conversations, the speaker never truly knows what the listener expects as input. So then, why must developers assume that in conversational AI design? The need of the hour is to build a system that can track and maintain context during the course of a conversation that contains multiple speech acts and sometimes extends that understanding throughout multiple conversations.

CURRENT STATE OF CONVERSATIONAL AI

Most alternative systems classify user's intent and thereafter analyze keywords/entities and use them to slot fill. Thus, in the past, use cases were kept simple owing to the technological limitations for the practitioner to build solutions.

But this approach is stymied by natural speech since humans use multiple speech acts and don't always supply atomic values to describe the arguments in the conversation. Most conventional conversational AI (including ChatGPT) systems do not have the reasoning capabilities to understand constraints or interpret obstacles that users might articulate throughout the conversation. Those that do have a plan-based model and are based on a cognitive theory, incorporating mental states for both the human and the system in the conversation.

CHALLENGES OF CONVERSATIONAL AI BEING SOLVED

Recently, Amazon research revealed that end-users today prefer automated digital humans since they outperform live human agents on speed and availability in customer interactions. This has led to

additional research and investments by large organizations to facilitate the building of Large Language Models.

Such initiatives are expensive, prohibitive, and beyond the scope of organizations with more limited resources. However, the growth of open-source communities has, to some extent, tried addressing this need in the market and is attempting to democratize the large-scale adoption of such fundamental research.

This brings us to the second problem at hand, most of the large language models which are being built are trained in English. The diversity of languages spoken around the world, its variations, dialects, the prosodies, and the nuances make it a difficult task to train such large language models in multiple languages as it is time consuming and resource heavy. Therefore, an intelligent system needs to have reasoning abilities independent of the language that can be adapted easily.

Thirdly, in most natural human-to-human interactions, non-verbal expression plays a major role in communicating intentions. It is crucial for systems to be able process and analyse such behavior. However, training machines to understand intention was not possible until very recently.

“Recent studies have shown for many use cases, the user-engagement is much better when the output of the bot/virtual assistant is rendered using a human-like avatar. Now, that brings an additional set of expectations from the users, that the (human like) avatar to show empathy in facial/bodily expression and tone.”

For practitioners like Openstream.ai® it is encouraging to see the depth of research being conducted on transformers going multimodal. The advancement in neural networks and computer vision technology allows Eva™, the embodied virtual assistant from Openstream.ai, to solve more complex use cases using multiple modalities. In addition to providing a faster way to communicate the intent using natural interaction, multimodal input/output processing helps in mutual disambiguation due to noise in one input modality/channel.

It is easier and more natural for a user to gesture at and ask, “how much does *it* cost?” rather than spelling the proper name of what the *it* refers to.

Similarly, it becomes more convenient to imply a need while shopping online such as “show shirts like this in red from Ralph Lauren in XL size” ... where part of input to the bot/virtual assistant is provided through a non-verbal mode such as through touch or a pointing gesture.

Recent studies have shown that for many use cases, the user-engagement is much better when the output of the bot/virtual assistant is rendered using a humanlike avatar. This realization carries with it an additional set of implied expectations from users. Users expect that the

more humanlike an avatar is, that it should reflect empathy in its facial and bodily expressions as well as the tone of the responses.

The empathy goes a step further in that it should be in response to user behavior as well. Whether that is expressed through verbal or non-verbal behaviors. It's perhaps easier to see that someone is distracted by *observing* them. Thus, the bot/virtual assistant must also have the ability to observe the user for engagement, mood, and sentiment for better outcomes from the conversation.

PROMPT: WHAT ARE THE PROS AND CONS OF CHATGPT VS PLAN-BASED DIALOGUE SYSTEMS

ChatGPT and plan-based dialogue systems are two different approaches to building natural language processing (NLP) systems that are capable of engaging in conversation with humans. Here are some pros and cons of each approach:

ChatGPT:

Pros:

- ✓ ChatGPT is a powerful and flexible approach to building NLP systems. It uses a large language model trained on a vast amount of text data to generate human-like responses to user inputs.
- ✓ ChatGPT can handle a wide range of conversational topics and can adapt to new ones quickly.
- ✓ ChatGPT can generate responses that are highly coherent and coherent with the previous conversation.

Cons:

- ✓ ChatGPT systems can be difficult to fine-tune and customize for specific use cases.
- ✓ ChatGPT systems can generate responses that are not relevant or appropriate for the conversation.
- ✓ ChatGPT systems may require *a large amount of computing resources to run effectively.*

This is the trend that the most advanced conversational assistants are already heading toward.

WHAT ARE THE POSSIBILITIES WITH CONVERSATIONAL AI?

Openstream.ai® one of the leading conversational AI platform vendors offering Eva™, its advanced end-to-end multimodal, context-aware virtual assistant is leading the pack in paving the way toward that very goal.

Based on an unscripted, proprietary plan-based dialogue, the system brings empathy and explainability to conversations.

With the goal of increasing understanding beyond simple intent classification and named entity recognition, Eva produces a logical form for each utterance via a semantic parse giving Eva an advantage over other text or voice-based agents in sentence splitting and multiple intent-based scenarios.

This logical form is consumed by Eva's reasoning service in an attempt to understand what the user said beyond atomic values as designed by other products in the market. Eva operates by

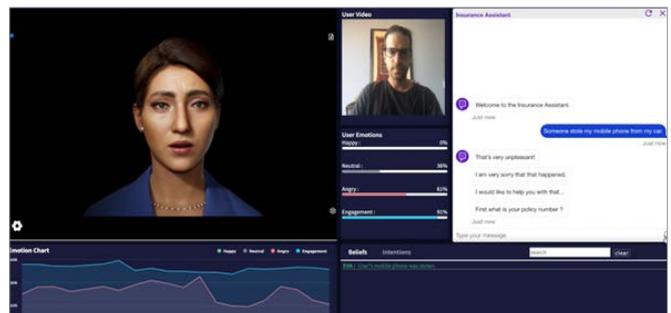
modelling mental states exposed in these logical forms for both Eva herself and the user. For example, their desires and the objects of that desire. This allows the agent to reason and to engage in interactions with emotions leveraging a “neuro-symbolic” architecture which cannot be scripted and is purely data-driven.

The system reasons through its understanding of the end-user's plan, clarifying ambiguous answers and overcoming obstacles with-in it by asking further questions. It also establishes trust through a step-by-step explanation, proactively offering rationale for its utterances. The dialogue planner establishes what preconditions are necessary to achieve the user's goals and confirms mutual beliefs about them seamlessly for a frictionless and satisfying experience.

The robust goal-based dialogue system powering Eva not only analyses and synthesizes the avatar's performance to align with the dialogue state expressions but also analyses the human expressions and emotions using computer vision. This allows the digital human to appropriately reciprocate an appropriate emotion in real-time to gain trust of the end-user. This multimodal analyses and the corresponding cognition of intent, natural emotion, and empathy generation by the embodied virtual agent is unique to this system.

ON THE HORIZON

There are challenges in conversational AI from a practitioner's perspective that need to be sifted through in order to separate the current market hype from reality. In fact, there are promising approaches to those challenges just on the horizon. However, despite recent advances in natural language understanding (NLU), sentiment analysis,



and context-awareness, there is still a gap between the capability of CAIs today, and the potential benefits which can be capitalized on by dedicated research and product teams in the short term.

Large scale adoption of conversational AIs remains a few years away. However, agents like Eva have begun to demonstrate reasoning, context awareness, and broader understanding of natural language and dialog. And they will continue to build momentum and success by gaining trust and respect within the industries that they are revolutionizing and the end-users which they serve and support.

About

Openstream.ai® is a leading provider of embodied virtual assistant platform (Eva™) employing open standards based multi-modal conversational AI for unparalleled user experience and intelligent process automation across industries around the world with documented productivity enhancement and ROI.