Mindtree
A Larsen & Toubro Group Company

*Welcome to possible*


## CONVERSATIONAL IVR

# Gateway for Customer Success

### Chapter 3 – Looking Beyond the Abyss

## Abstract

Understanding the pain points in the conversational AI ecosystem and overcoming the chasm leading to contact center modernization.

# Content

## Conversational IVR

# Rise of chatbots

The concept of chatbots dates back to the 1950s when Alan Turing first proposed the idea to label a system as sentient if it could act, react and interact like a sentient being. The advent of messaging apps and improvement in technology led to chatbots becoming more robust and efficient.

Post 2015, chatbots evolved from a niche technology piece to key technological capability that every business wanted to try. Many chatbots were being rolled out and shut down during between 2015 and 2018, and by 2019, they had already secured a place in major communication channels such as web chat, Interactive Voice Response (IVR) and Facebook messenger. It is estimated that in 2020 that value of chatbot marked was 17.17 billion USD, which will reach 102.30 billion USD by 2026.

**A key chatbot that was developed during the initial days of COVID-19 was the Government of India's MyGov Saathi (Meaning, companion, in Hindi) was built to cater to a population of 1.3 billion. This chatbot can handle up to 300,000 users a day and 20,000 concurrent users per minute. It is evident that chatbots were one of the 'frontline' workers in the Covid-19 pandemic with a substantial impact.**

When the pandemic struck, governments around world and organizations such as WHO and CDC suddenly felt an urgent need to roll out virtual assistants, which could help curb the spread of misinformation and provide an empathetic response to users on behalf of a human.

The growth in the chatbot market has left people wanting for more. Now, chatbots are not being treated as an automated system (IVR for example). Rather, they have transitioned into the role of a partner, a pseudo-human, who provide a similar experience as talking to a human being.

# The Taxing Triad

The advent of chatbots paved the way for their incorporation into different channels and platforms and henceforth, their exposure to end-customers. Chatbots are a key component in the conversational IVR ecosystem. According to a Gartner report, 40% of chatbots/virtual assistants which were built in 2018 would have been abandoned in 2020.

Why did this happen? What are the problems that lead to renunciation of a chatbot/ conversational IVR?

When looked at through the eyes of a customer / IT service leader, the problems that the current conversational IVR industry face, can be attributed into three broad categories – **Cost, Capability and Experience.** Let us look at these in detail.

## Cost

Any system's development and maintenance incur costs. The customer / IT service desk is no exception to this. The recurring cost a service desk incurs can be either due to many agents needed to cater to people; a Gartner report says **that on an average a cost per call around $8.0;** or it can be due to longer release cycles of conversational IVR.

## Capability

Capability is an important aspect of the service desk. Enabling the conversational IVR/voice bot to perform multiple functions is the key aspect to success. If customers are not getting what they expect and are regularly redirected to an agent, they would gradually start avoiding it and would like to jump to an agent directly. To avoid this, a periodic (and preferably in short span) feature/capability addition must be made to voice bot.

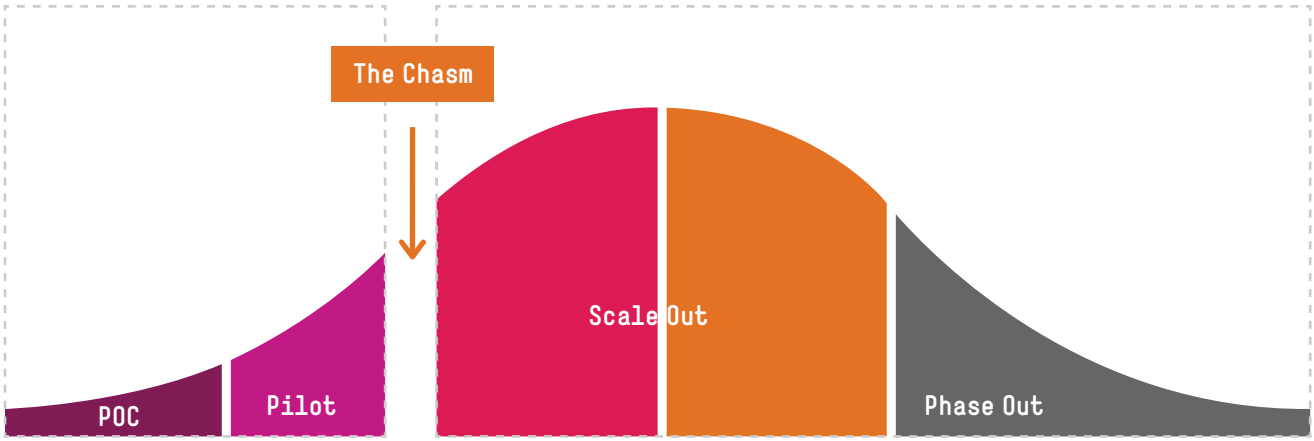
This, however, is not the end. The customer/IT service desk leader must take into account that every capability addition to the voice bot requires a certain degree of research and development. This leads to longer release cycles and not only increases the costs, but also impacts the user expectations.

## Experience

Experience is another key factor in the customer success journey. The significance of experience can be understood by the poll from **Publicis Groupe's DigitasLBi, where 73% of people said that they would not use a brand's chatbot if something goes wrong first time.**

Now that we have established the potential pitfalls in the journey to voice bot success, we would need to find out ways to counter them. So, the next question is how do we address the pitfalls in order to realize the true potential of voice bots while cutting down the recurrent costs?

## Crossing the Chasm




We saw the broad categories where the problems during the bot development phase arise. Let us ponder over the next question – **What is the most efficient way to resolve such problems?** Enter the accelerators.

**Accelerators can be defined as helper tools in a conversational AI ecosystem. An accelerator is a non-customer facing application, which helps the voice bot either through automation or bringing in a**

**process, were nonexistent.** For each of the triads that we saw in the former section, building accelerators can take the edge off the problems. Let us go ahead and look at these accelerators in detail

1. **Penny-wise accelerators**
    a. Conversational Test Studio
    b. Training Automation
2. **Capability accelerators**
    a. Conversational Data Generator
    b. Intent Analyzer and Model Benchmarking
    c. Voice bot DevOps Pipeline
3. **Experience accelerators**
    a. Conversation design – Refer to conversation design whitepaper.
    b. Conversational IVR Dashboard

Some of the accelerators were tried and tested with our customers, which resulted in improved performance and reduction in time needed for development, as documented below

# Conversational Test Studio

Conversational Test Studio is a dual-purpose accelerator; it can perform both volumetric testing for NLP and flow testing for conversation design. Note that this is built only for testing chatbots and not the client application. CTS (Conversational Test Studio) falls under the Penny-wise accelerator category. Let us understand both these in detail and see how an automated testing accelerator helps achieve this.

## 1. Volumetric testing

Volumetric testing takes its input from conversational data generator. It carries out three levels of testing to ensure your NLP model is performing in the desired fashion. Below are the three stages of volumetric testing.

### a. Smoke testing

Smoke testing is stage 1 of automated testing. It is done to ensure that at least the basics are working right before moving onto deeper testing strategies. This is done with 5-10 utterances / user query per intent. If the score in this testing is 100%, the tool moves to next stage.

### b. Amplified testing

Amplified testing is stage 2 of volumetric testing. Here, the number of queries per intent is bumped to 1000%. So, we would test approximately each intent on 50-100 queries each. This is a pre-requisite to in depth volumetric testing. For amplified testing, we select the queries to be tested using a combination of clustering using STS (Semantic Textual Similarity) and random choice. We form clusters in the query pool of each intent by STS technique. We aim to get around 50-100 clusters each, in the query pool, by the pattern they are following. Each cluster represents a unique pattern; a unique way of asking questions. From the clusters, we then pick one query each by random choice. This collection of queries becomes our test data for this stage.

### c. Scaled testing

In in-depth volumetric testing, each test phrase must be tested again for its corresponding intent. In other words, all the data that is available to test is utilized in this stage. It is a heavy process and often can take hours to finish, depending on the amount of test data available.

## 2. Flow Testing

Flow testing is done to test if the bot is performing exactly as per the conversation design. This testing compares the response of the bot at each step in a use case to that of expected outcome. The failures in this step indicate that there are problems in the bot's configuration. Flow testing covers both conversation design and the functional testing.

## 3. Voice Testing

A Conversational IVR also has one more layer that needs testing, which is the voice layer. CTS can perform voice testing as well and calculate the STT efficiency (WER calculation). CTS can also simulate different voices (based on noise, frequency and gender). It can also go beyond and do the end-to-end testing for your conversational IVR solution.

**We implemented a mini version of CTS for an insurance client, which has the capability of volumetric and flow testing, thus leading to a reduction of 20%-25% of testing efforts per release.**

# Training Automation

Training automation is another of our **penny-wise accelerators.** For a conversational IVR to stay on top of its game, it is imperative that it goes through periodic training cycles. The chatbot of the conversational IVR must be re-trained and tested with new training phrases and intents to accommodate other users' questions. On an average, a training cycle for a chatbot with 200+ intents and 100+ training phrases per intent can take up to 50 hours, depending on the complexity and annotation needed. This duration however can be cut down through a training automation accelerator. It can factor in all the training phrases for the intents and train the bot. The entire training process takes mere seconds to finish.

# Conversational data generator

Conversational data generator is a **capability accelerator** to provide test data for intents. Often, we run into unavailability of data when we are building new intents or sometimes, there is not enough test data available to perform volumetric testing.

This is where this accelerator comes into play; it has the capability to generate thousands of phrases for a given intent, which can then be utilized for training or volumetric testing. This accelerator is based on the 'End Focus' principle of linguistics. It can be integrated with the development pipeline either through an interactive UI or as a script/stage.

The conversational data generator also has a voice generation component; it can generate audio, simulating a variety of factors such as gender, noise level, frequency etc. As a capability type accelerator, this accelerator allows for faster rollouts of features.

**We have noted an improvement in the accuracy of chatbots, of a large North American Insurance provider, from 60% to a whopping 80%, post retraining and increasing the number of training phrases per intent to ~100+.**

# Intent Analyzer and Model Benchmarking

One of the key problem areas of a chatbot is accuracy. Enterprises and developers tend to spend a lot of time trying to get the accuracy of chatbot right when they should be trying to tend to F-mean. Let us understand this in more depth.

A voice bot is a Machine Learning system and should be treated like one. Accuracy is one parameter of ML systems but not the only one that should be used in judging the voice bot. The key-parameters for ML systems are Accuracy, Precision, Recall and F-mean. If we judge the voice bot only based on accuracy, we miss out on the true picture of how our bot is performing.

To solve this problem, we created another **capability accelerator,** which translates the output from the volumetric testing and puts it in a report that is easier to understand. This accelerator shows the details of voice bot's NLU model, intent-wise stats and the confusion matrix. The report from this tool helps us see the intents that are not performing well and the reason for the problem. Below is a snapshot of how the report from this tool looks like.

# Voice bot DevOps Pipeline

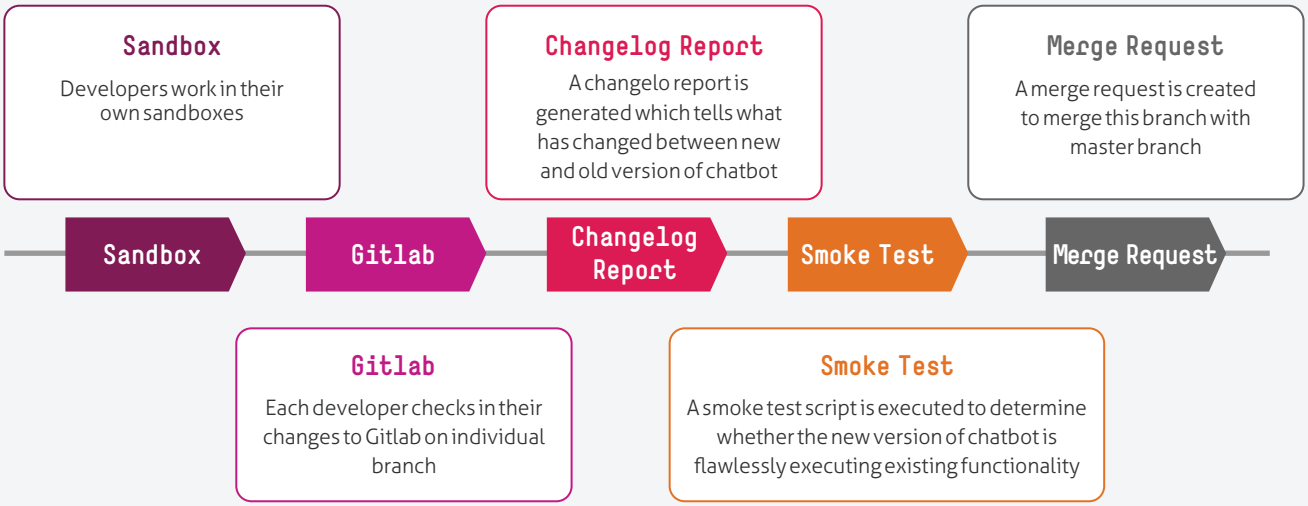
Voice bot DevOps pipeline is another developer-centric tool, a **capability accelerator** that helps reduce time-to-market. Without standard CI/CD rules in the development cycle of voice bot, there is always a risk of developers overwriting or creating conflicting configs with another developer. This tool ties your voice bot to standard repository platforms – Gitlab and Github and the changes to the voice bot. Promotions to higher environments also happens through this.

In addition, a changelog report is generated automatically, which shows the difference between new and existing versions of the voice bot on the current branch.

| Intent | Accuracy | Precision | Recall | F-mean |
|---|---|---|---|---|
| Intent 0 | 0.69 | 0.85 | 0.79 | 0.82 |
| Intent 1 | 0.23 | 0.26 | 0.71 | 0.38 |
| Intent 2 | 0.87 | 0.93 | 0.93 | 0.93 |
| Intent 3 | 0.53 | 0.67 | 0.71 | 0.69 |
| Intent 4 | 0.62 | 0.83 | 0.71 | 0.77 |
| Intent 5 | 0.87 | 0.93 | 0.93 | 0.93 |
| Intent 6 | 0.88 | 0.88 | 1 | 0.94 |
| Intent 7 | 1 | 1 | 1 | 1 |
| Intent 8 | 0.93 | 1 | 0.93 | 0.96 |
| Intent 9 | 0.8 | 0.92 | 0.86 | 0.89 |
| Intent 10 | 0.86 | 0.86 | 1 | 0.92 |

| | Intent 0 | Intent 1 | Intent 2 | Intent 3 | Intent 4 | Intent 5 | Intent 6 | Intent 7 | Intent 8 | Intent 9 | Intent 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Intent 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Intent 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Intent 2 | 0 | 0 | 0.714285714 | 0 | 0 | 0.285714 | 0 | 0 | 0 | 0 | 0 |
| Intent 3 | 0 | 0 | 0 | 0.64285714 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Intent 4 | 0 | 0 | 0 | 0 | 0.928571 | 0 | 0 | 0 | 0 | 0 | 0 |
| Intent 5 | 0 | 0 | 0 | 0 | 0 | 0.857143 | 0 | 0 | 0.071429 | 0 | 0 |
| Intent 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0.571429 | 0 | 0 | 0 | 0 |
| Intent 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Intent 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.928571 | 0 | 0 |
| Intent 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.857143 | 0 |
| Intent 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.928571 |

This whole process is depicted pictorially in the below diagram.




# Conversational IVR dashboard

Conversational IVR dashboard is the analytics tool to provide the KPI metrics around the IVR/IVA solution. It is an **experience accelerator;** it highlights the adoption brought about in a traditional IVR system by switching to a bolt-in conversational IVR system. A dashboard for KPI metrics is a fundamental requirement while reporting to contact center stakeholders about the performance of IVR/IVA.

Apart from the standard KPI metrics covered under this accelerator, it can also provide a more detailed analysis by initiating an analysis of customer sentiment (a function of customer experience) and suggest new topics / intents to be added to the voice bot for an improved containment. Another deciding factor in the conversational AI ecosystem is the end-to-end call analytics that provide information about each individual layer in your conversational solutions, highlight the user experience and expose the problematic areas along with suggestions to improve.

# Conclusion

Accelerators are game changers; they can bring about a rapid paradigm shift in the conversational AI landscape. However, we also need to consider whether we want to reinvent the wheel or take advantage of existing solutions. Accelerators also have a cost factor associated with them. Hence, a judicious decision must be made on whether investments are needed to build them, or if they can be bought from the market.

# References

https://www.gartner.com/en/newsroom/press-releases/2019-09-25-gartner-says-only-9--of-customers-report-solving-thei

https://www.gartner.com/en/newsroom/press-releases/2020-01-15-gartner-says-74--of--customer-experience-leaders-expe

# About the authors

## Bikartana Panigrahy

Practice owner

He is the practice owner for Conversational AI at Mindtree. He has over two decades of experience working for customers globally, and has been involved in multiple innovation and Digital Transformation programs to improve customer success. His core focus is enterprise IT automation, self-service for brands and contact center automation using cognitive service, AI/ML and deep learning models.

## Sushma Itagi

SME

She is the SME for conversation AI, leading the CPaaS CoE. She has implemented conversational IVR at scale for a leading retailer in UK. Her core expertise includes multiple chat and voice platforms, cognitive solutions, re-usable components and accelerators for a conversational AI ecosystem.

## Shashikant Jha

SME

He is the SME for conversation IVR, leading the Dialogflow CoE. He has implemented conversational IVR at scale for a fortune 500 North American insurance provider. His core expertise includes conversation design, AI/ML, deep learning, microservices, automation, accelerators and DevOps for a conversational AI ecosystem.

## *About Mindtree*

*Mindtree [NSE: MINDTREE] is a global technology consulting and services company, helping enterprises marry scale with agility to achieve competitive advantage. "Born digital," in 1999 and now a Larsen & Toubro Group Company, Mindtree applies its deep domain knowledge to 275+ enterprise client engagements to break down silos, make sense of digital complexity and bring new initiatives to market faster. We enable IT to move at the speed of business, leveraging emerging technologies and the efficiencies of Continuous Delivery to spur business innovation. Operating in more than 15 countries across the world, we're consistently regarded as one of the best places to work, embodied every day by our winning culture made up of over 22,000 entrepreneurial, collaborative and dedicated "Mindtree Minds."*