THE PATH TO SUCCESSFUL CONVERSATIONAL AI CAPABILITIES
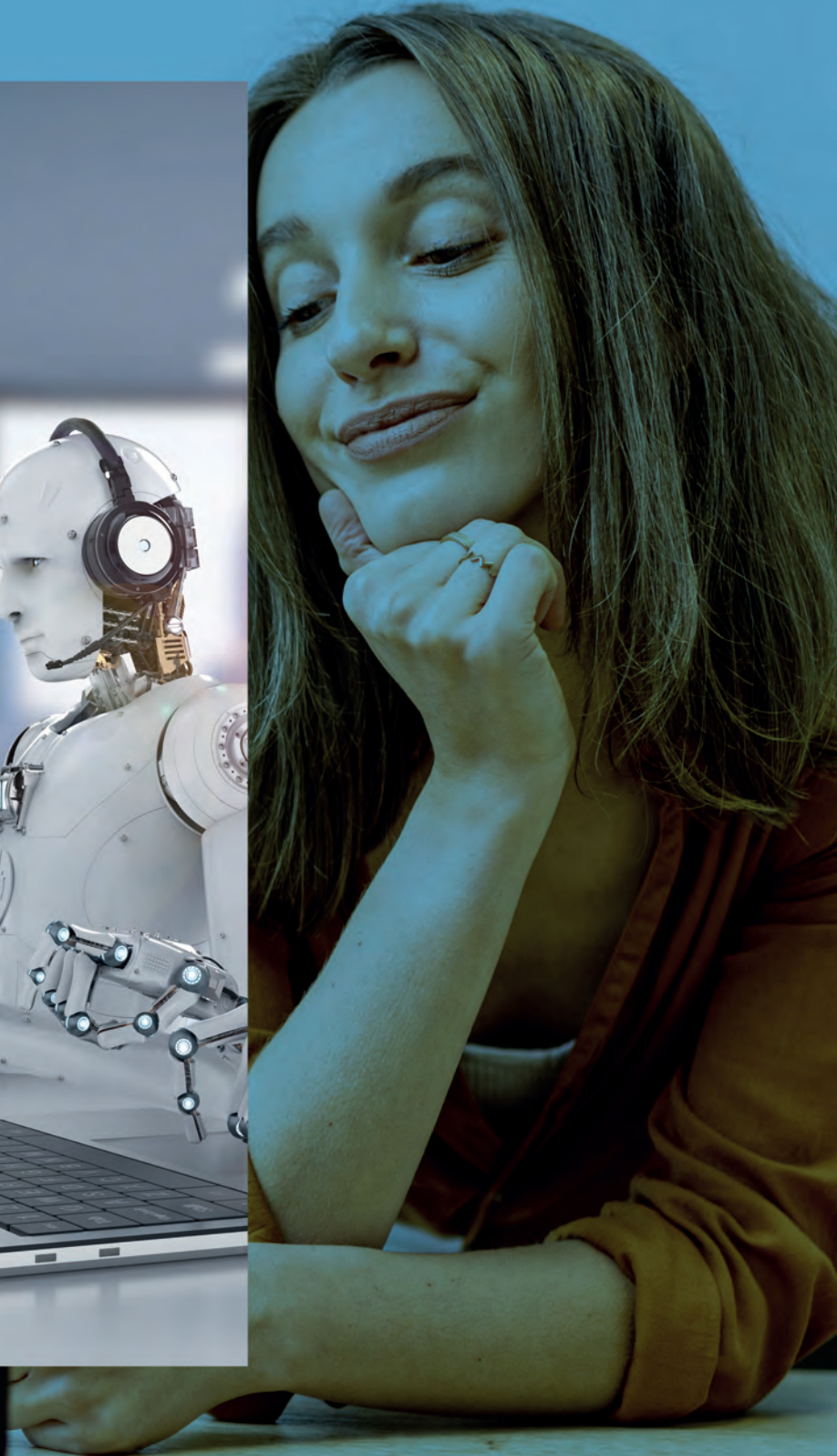
Infosys®
Navigate your next

Infosys® | Knowledge Institute

Contents

Conversational AI technologies are entering the era of hyperpersonalized, multimodal assistants that work seamlessly across channels, understand multiple languages, cater to specific industry domains, and are empathetic, inclusive, and immersive. Enterprises should take a gradual approach to conversational AI, increasingly moving toward complex features with continuous incremental advancements throughout the solution's lifecycle.

AI assistants are evolving rapidly

Around 56% of organizations worldwide have already implemented AI at scale, while another 32% are either experimenting with it across some business units or carrying out pilots, according to Infosys' Digital Radar 2022.[1] Gartner estimates global AI software market revenue to hit $62.5 billion in 2022, excluding hardware and services revenue from AI.[2]
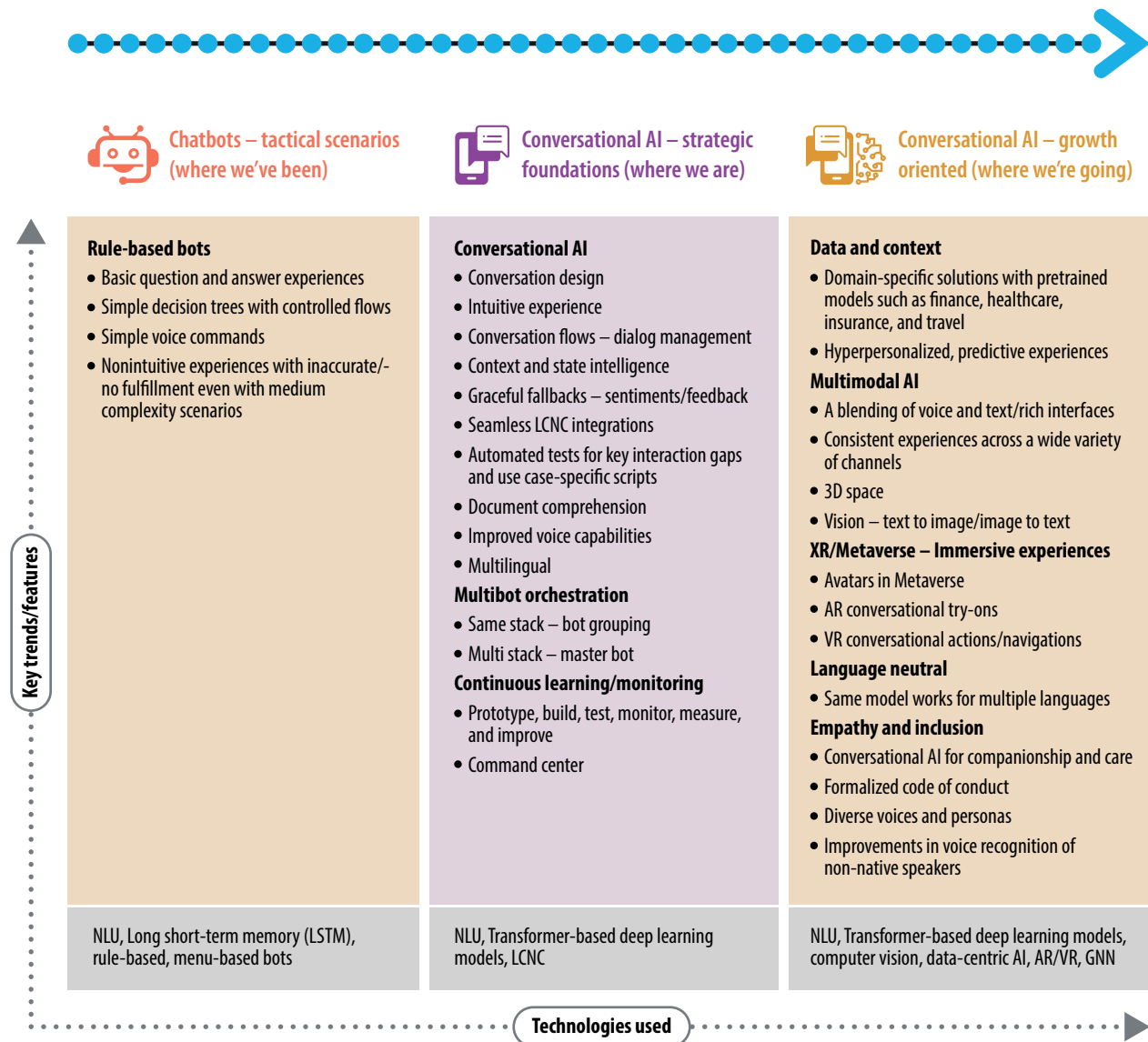
Conversational AI has been part of the enterprise landscape since 2016, but it came into focus during pandemic times. It supported customer service and provided content suggestions to get businesses and people through challenging times. As of 2021, more than 300 million households worldwide have smart speakers.[3] Conversational AI is among the top five categories of spending on AI software, with an estimated $7 billion in revenues tagged to virtual assistants this year.[4] Advancements in contextual

and cognitive conversational AI will add more value going ahead.

From simple rule-based bots, conversational AI evolved into intuitive chatbots that now serve customers/employees of online food delivery companies and banks. We are heading toward hyperpersonalized, immersive conversational assistants that can closely mimic human responses. This will help companies build more engaging and intuitive experiences. However, there's still much to be covered in this domain (Figure 1).

Technologies driving these developments include natural language understanding (NLU), deep learning, low code, no code (LCNC) tools, computer vision, augmented reality (AR), virtual reality (VR), and graph neural networks (GNNs). Currently,

Figure 1: Progression of conversational AI technologies and associated trends

Chatbots – tactical scenarios (where we've been)

Conversational AI – strategic foundations (where we are)

Conversational AI – growth oriented (where we're going)

Key trends/features

Rule-based bots
- Basic question and answer experiences
- Simple decision trees with controlled flows
- Simple voice commands
- Nonintuitive experiences with inaccurate/-no fulfillment even with medium complexity scenarios

Conversational AI
- Conversation design
- Intuitive experience
- Conversation flows – dialog management
- Context and state intelligence
- Graceful fallbacks – sentiments/feedback
- Seamless LCNC integrations
- Automated tests for key interaction gaps and use case-specific scripts
- Document comprehension
- Improved voice capabilities
- Multilingual

Multibot orchestration
- Same stack – bot grouping
- Multi stack – master bot

Continuous learning/monitoring
- Prototype, build, test, monitor, measure, and improve
- Command center

Data and context
- Domain-specific solutions with pretrained models such as finance, healthcare, insurance, and travel
- Hyperpersonalized, predictive experiences

Multimodal AI
- A blending of voice and text/rich interfaces
- Consistent experiences across a wide variety of channels
- 3D space
- Vision – text to image/image to text

XR/Metaverse – Immersive experiences
- Avatars in Metaverse
- AR conversational try-ons
- VR conversational actions/navigations

Language neutral
- Same model works for multiple languages

Empathy and inclusion
- Conversational AI for companionship and care
- Formalized code of conduct
- Diverse voices and personas
- Improvements in voice recognition of non-native speakers

NLU, Long short-term memory (LSTM), rule-based, menu-based bots

NLU, Transformer-based deep learning models, LCNC

NLU, Transformer-based deep learning models, computer vision, data-centric AI, AR/VR, GNN

Technologies used

Source: Infosys

these technologies focus on fulfilling strategic use cases such as context intelligence, document comprehension, and multilingual communication. Going forward, they will drive multimodal (blended voice/text/visual interfaces) 3D assistants with avatars and AR elements. They will also develop the ability to understand diverse voices, accents, and personas and take in text or images/videos as inputs.

We have identified the following seven major trends, which reflect the latest developments in the conversational AI space.

Trend 1: Conversational experiences to become hyperpersonalized

Standardized conversations based on predefined rules project behaviors typical of robots. During early developments, such conversations seemed amusing — imagine the humanized robot character "Vicky" from the 1980s TV show "Small Wonder". However, with smart technology, it isn't as amusing now. Investments in chatbots can easily go wrong and worsen customer experience. Conversing with chatbots can be frustrating with repetitive responses and limited coverage. People want an early resolution without unnecessary intervention, as in the earlier setup, where one could just call and speak to someone. Even if that person could not provide a satisfactory solution, one could walk away with some closure. But current customer service chatbots leave people exasperated when none of the predefined options solve their problems. It is even worse when there is no option to speak to an agent.

Now, organizations are building bots (chatbots, voice bots, and audio-visual immersive bots) that can mimic human responses with empathy and contextualization. These bots can analyze historical data to build context and understand the likes,

dislikes, personalities, and moods of the user to respond accordingly. Businesses intend to develop hyperpersonalized bots for all kinds of users, including human resources assistants for employees, and guided purchase and after-sales support assistants for customers.

Such hyperpersonalized bots provide customized menus, services, and resolutions based on a user's routine behavior and personality. They can automatically sense and switch vocabulary and language depending on user demographics such as age, gender, and region. These bots can respond to changes in customer voice tone, facial expressions, sentiments, and body language (in video calls) and fetch and share personalized user information in real time.

However, personalization requires personal data. Therefore, data collection and usage protocols need to be airtight and in alignment with jurisdictional rules. Organizations that utilize and protect data with utmost care while building good user experiences will see better outcomes.

Infosys has launched a hyperpersonalized learning assistant "Zoiee" for its employees. It is an avatar-based assistant that helps employees develop skills by navigating thousands of knowledge repositories. It utilizes users' inputs, previous conversations, and other job-related data to suggest certifications and courses that can help advance their careers and meet learning targets. It reminds users to complete their undertakings by stipulated deadlines and even changes the attire of the avatar as relevant (such as festive costumes during the Christmas season and masks during the pandemic).

Trend 2: Voice-enabled conversations to penetrate deeper in business and consumer spaces

Every smartphone now has voice AI, while smart speakers are in more than 300 million households worldwide.[5] Voice search and assistance features are increasingly getting integrated with interactive voice response (IVR) for customer support, and other digital devices and services such as smart TVs, online streaming, and drive-through food ordering. Increasing receptiveness toward voice assistants further boosts these initiatives. According to a study by Pew Research Centre, 55% of virtual assistant users favor speech recognition applications.[6]

Voice AI caters to a larger demographic, comprising the young, tech-savvy users of Alexa, Siri, and Google, the older generations, and the differently abled who might find typing difficult. The technology can integrate with smart speakers to serve as the first point of contact for employee/customer queries. In the contact center space, voice AI can significantly improve the first call resolution (FCR) rate and speed-up ticket closures. This allows agents to focus on more critical issues such as assisting in process improvements that reduce ticket generation in the first place.
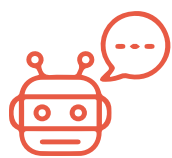
However, current voice AI applications lack the breadth in expression and the understanding of voice modulations, limiting their ability to process contextual information. Further, uncontrollable factors such as background noise, variations in accents, and the variety of languages people speak constrain voice AI.

Nonetheless, companies are finding innovative ways to solve these problems. For instance, Meta is working on a universal speech translator aimed at enabling people across the world to communicate without language barriers.[7] This single AI model-based translator will be deployed in applications supporting its wearable devices, and its augmented reality (AR) and virtual reality (VR) projects. Similarly, NVIDIA is working on reducing noise effects and making communications clearer with its NVIDIA Maxine software development kit (SDK) that works with audio, video, and AR effects.[8] In May 2022, Google also announced advancements that make its Assistant capable of processing signals such as proximity, head orientation, and gaze detection for Nest users to communicate with it more naturally, and without having to say "Hey Google".[9] It is also working on improving Assistant's responsiveness by incorporating understanding of natural speech elements such as pauses and words like "um".[10] Such developments make voice AI more evocative.

55% of virtual assistant users favor speech recognition applications.

Trend 3: Immersive conversational assistants to proliferate

As the metaverse ecosystem evolves, immersive technologies will become the norm for people, businesses, and machines to communicate. AR and VR elements will enhance all major spaces, from workplaces and factories to gaming, entertainment, and social media. Digital assistants are one of the early deployments of immersive technologies, as businesses strive to improve the customer experience while increasing automation. These assistants will increasingly become avatar-based audio-visual chatbots with gesturing capabilities to make the experience as human-like as possible. However, AI has a long way to reach there.

Customization features that allow enterprises to design their own avatars (reflecting companies' ethos) will also gain traction. If AR/VR headsets manage to penetrate consumer and enterprise spaces, users will be able to activate their own avatars and communicate with immersive virtual assistants to assign them tasks or ask queries.

Such experiences will be driven by a combination of multiple technologies on the user interface/user experience (UI/UX) front. These include 3D facial animation (allowing the avatars to reflect emotions through expressions), optical character recognition (OCR), computer vision (to analyze text/audio-visual information and respond accordingly), and gesture-based responses (like nodding while listening or raising eyebrows when an avatar is surprised).

Leading companies such as NVIDIA, Meta, and Amelia are driving innovation on this front. For instance, NVIDIA Maxine reduces noise to enhance voice AI communications. It can track face position, estimate body pose, maintain artificial eye contact with the camera in use, and will soon be able to swap a user's face with that of any animated character.[11] Similarly, Amelia (formerly IPsoft, an American tech company) facilitates the development of custom, no-code conversational AI applications that comprise elements of expression, emotion, and understanding of background information for more humanistic conversations.[12] Meta's recently announced Project CAIRaoke is also an end-to-end neural model for audio-visual conversational AI, which is currently being tested on its video calling service Portal and is later expected to be integrated with AR glasses.[13]

Immersive bots developed with these technologies have huge opportunities across industries. For instance, healthcare companies can develop bots that act as caretakers assisting patients by reminding them to take their medicines on time, answering queries related to their condition, monitoring their vitals and raising alarms (if needed), and providing emotional support by listening and responding empathetically. Corporates can use such bots to help with onboarding, HR and finance related queries, information technology (IT) issues, and others.

Trend 4: Context persistence to gain prominence for seamless conversations across devices

Users expect a seamless experience while they keep switching between multiple devices (phone, laptop, tablet, smart speakers, etc.) throughout the day. This compares poorly with impersonal and frustrating support experiences, where users explain the situation to a bot, then to a human agent, and then to another senior agent — pretty much every time from scratch.

Enterprises strive to add context persistence capabilities that enable customers to switch from social media to web or phone conversations without losing the context set previously. They also allow for a smoother payments experience, providing multiple options through customer-preferred channels.

Such AI solutions will transform contact centers in the upcoming years.

Infosys Cortex is an AI-based contact center solution that enables context persistence across channels. It focuses on enhancing agents' experiences by helping them with relevant information and suggestions in real time to effectively support customers.[14] It facilitates automated call transcriptions for smoother transfers between agents, and understands customers' tone and intent to suggest agents the best course of action. It allows companies to tap into previous call recordings, helping identify customers' core issues, agents' challenges, and training requirements.

Trend 5: Multimodal conversations to aid seamless interaction experiences

Real-life conversations involve subtle body language and gestures that humans parse effortlessly through face-to-face, audio calls, video conferencing, texts, virtual/augmented 3D spaces, and even gestures. Conversational AI is advancing to incorporate these aspects, as companies vie to deliver the best experiences and become increasingly inclusive. Multimodal AI solutions can gather information from multiple sources, including text, visual, and audio, to deliver more contextual and precise conversational experiences.

Conversational AI solutions already have or will soon have the capabilities to:

- Serve wider demographics, especially for people with disabilities.
- Integrate with AR glasses or other AR-enabled devices to enable meaningful conversations in metaverses.
- Better comprehend emotions through precise analysis of facial expressions, lip movements, and gestures.
- Advance to two-way continuity of conversations, from currently prevalent single question-answer type voice.
- Make reasonable predictions on customer needs, sale conversion possibilities, and overall interaction experience for continuous improvement.

Big players such as Google, Meta, and OpenAI, are adopting these multimodal AI developments at speed. Meta has launched multiple projects to enhance multimodal AI understanding. One of these projects aims at identifying hateful memes by deciphering both their image and text components. In 2020, Meta (then Facebook) released its comprehensive dataset on memes for external researchers (through a competition) to use and build models that can help prevent hate on social media. Another project from Meta, called data2vec, enables self-supervised learning for speech, vision, and text inputs so computers learn to develop an understanding of surroundings without much of labeled data currently needed.[15] Google also recently launched new image search features for Lens (its multimodal search tool), which enable users to scan products and check their reviews, as well as availability at nearby stores.[16] The search giant's Autodraw feature is another experiment to help users draw faster by predicting the possible shapes they are trying to create as they start doodling.[17] Similarly, OpenAI's Dall·E is a neural network that can create images from text inputs, such as writing down the name of an animal or an object can bring up its image.[18]

In the next few years, people will be able to use text, image, voice, video, gestures, and other inputs to converse with chatbots, which is expected to further improve query resolutions while reducing the time and resources required.

Trend 6: Increased adoption of LCNC platforms to boost productivity

AI solutions require data selection and preparation, feature extraction (focusing on the key data features that are of interest to the application in the question), model selection, fine-tuning, and training. These technologies need testing and debugging and, most importantly, time. In contrast, LCNC tools can enable non-technical people to develop solutions through intuitive interfaces that use familiar interactions such as clicking, dragging or dropping, without the skill to write code. This makes LCNC tools cost-effective, as they help cut product development time significantly. By 2025, 70% of new application development will happen via LCNC tools, compared to less than 25% in 2020, predicts Gartner.[19]

LCNC tools reduce organizations' reliance on system integrators to manage development and maintenance of applications. They allow companies to rely more on business or functional experts than technical experts/developers. These tools will soon become the preferred solution for business process automation. According to Gartner, by 2024, 80% of technology products and services will be developed by the nontechnical workforce.[20] This implies the tech workforce will be spend more time on challenging projects requiring their expertise.

By 2024, 80% of technology products and services will be developed by nontechnical workforce.

Infosys Conversational AI suite is one such LCNC platform that non-technical professionals can use to design, build, evaluate, host, and monitor the entire solution lifecycle. This suite was deployed at a U.S.-based manufacturing company that wanted to fully automate its internal query resolution process for employees in the legal and intellectual property departments. The solution was created, tested, and deployed without using a line of code, and could instantly provide answers to more than 700 queries.

Trend 7: Local language and domain-specific AI models become popular

Streaming platforms like Netflix, Amazon Prime Video, and Hotstar have well tapped into regional language content. In India, regional content is expected to claim a 54% share on streaming platforms by 2024.[21] However, most other enterprises are yet to incorporate customer experiences in local languages. Therefore, conversational AI applications will find more such use cases in the coming years.

Large language models (LLMs) — such as BERT and ALBERT by Google, RoBERTa by Meta, BART by Amazon, GPT3 by OpenAI, and GPTJ, GPT-NEO, and T5 from open-source communities — are widely used to develop conversational AI applications. These models are primarily trained in huge datasets but only in prominent languages. They are extremely good at understanding linguistics, syntax, and semantics of the languages they are trained in, but that leaves many people either poorly served or not served. These models also lack understanding of specific domains such as healthcare, finance, or telecoms. Due to these shortcomings, speed and scale in production deployment are still significant challenges for these LLMs.

However, the focus has already started shifting toward regional language and domain-specific models, which should drive wider use of conversational AI to deliver services to people who speak a range of languages and dialects. For instance, Infosys has developed a domain-specific model (based on BERT) for the telecom industry. It can understand industry lingo and the technical terms better than the generic BERT model. Some leading U.S.-based telecom companies have deployed this model to create their AI-based conversational agents.

Bottom line: Deploy conversational AI based on the maturity level with automated lifecycle management

Companies don't get the most out of their initiatives because they take on too much change at once. Around 80% of business transformation programs can't deliver their true potential for the same reason.[22] Businesses should design a gradual and suitable approach to implement conversational AI (Figure 2).

Infosys recommends the maturity model approach for maximum adoption among target segments and higher accuracy of AI responses. The model suggests that a firm with no prior experience should start with simpler rule-based FAQ chatbots, and then transition to more complex chatbots that can extract data from enterprise systems and provide additional automated support to users. The bigger the change, the higher the resistance to change from customers/employees.

This way, a company learns from the issues that arise during the step-by-step evolution so that accuracy doesn't suffer.

Most organizations are at the second stage of transaction-based bots. These bots share information and perform database updates and execute customers' orders. However, knowledge-based, multimodal chatbots that deliver highly contextual conversations, and immersive bots designed with augmented or virtual elements, are not that far away. Technologies such as contextual understanding, computer vision, OCR, and sentiment analysis will soon be seen in chatbots across major technology firms.

Figure 2: Conversational AI maturity model
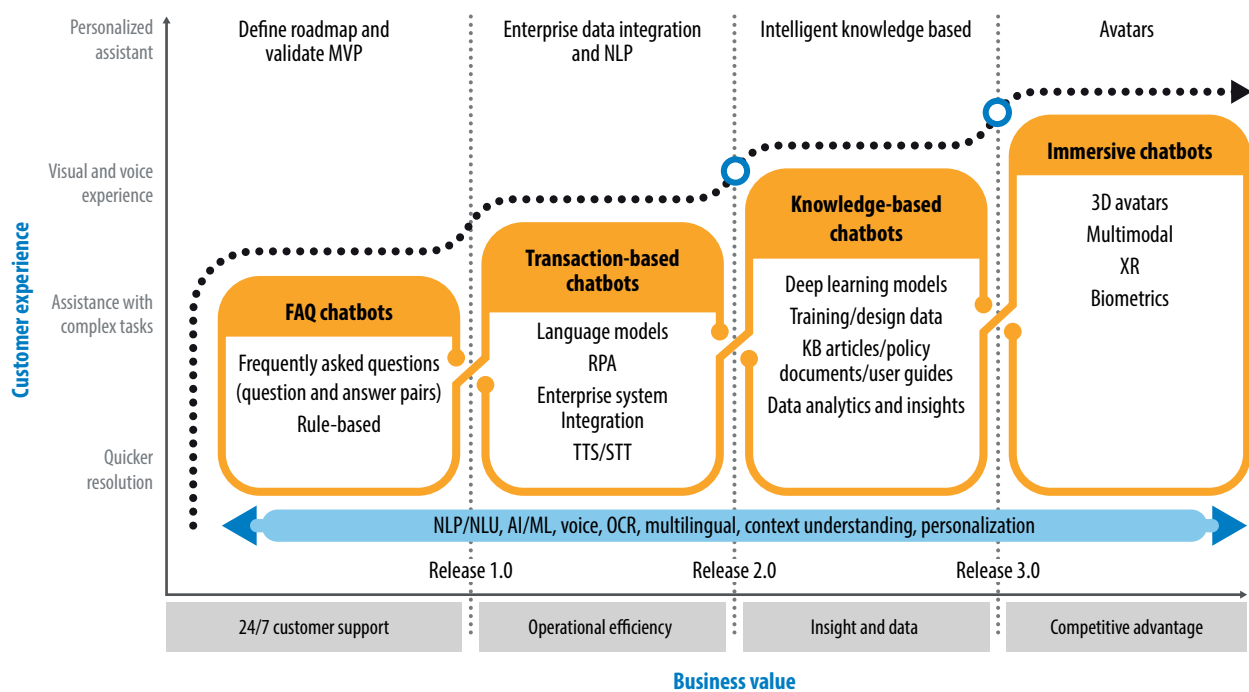
Source: Infosys

Figure 3: Conversational AI lifecycle

Prototype › Build › Test › Monitor › Measure › Improve

Feedback
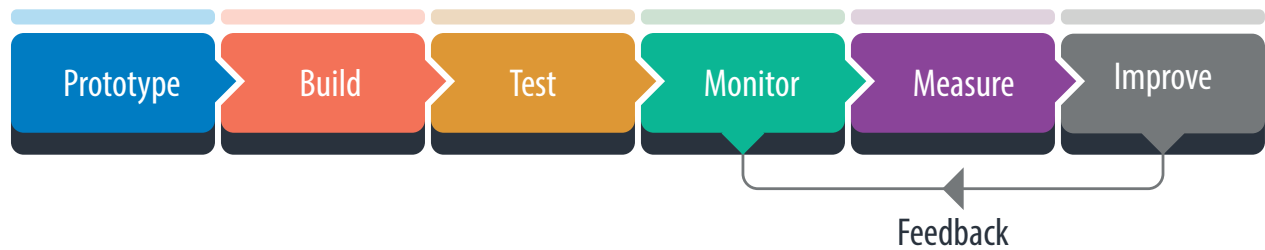
Source: Infosys

However, these advanced conversational AI solutions need to be deployed with a lifecycle management approach (Figure 3). Enterprises must look at a multistage process for their conversational AI journey.

This approach helps identify gaps and opportunities for automation at each stage. Most enterprises just focus on creating a feature-loaded multilingual solution that can be deployed quickly. However, they need to use dynamic and evolving data to incorporate changing customer preferences, continuously upgrade the AI model, and strictly emphasize model fairness to minimize bias. Firms

should also evaluate the AI model for accuracy and throughput; inference time, speed, and scale; prioritize privacy and security; and continuously monitor data for quality and representation.

These considerations cannot be managed manually. Enterprises must seek end-to-end automated solutions that not only help design chatbots but also test them, derive insights from data, and continuously improve their performance and accuracy. Ultimately, businesses can successfully venture into the conversational AI space through a maturity stage-wise deployment of a lifecycle approach.

References

1. Digital Radar 2022, February 2022, Infosys.

2. Gartner Forecasts Worldwide Artificial Intelligence Software Market to Reach $62 Billion in 2022, November 22, 2021, Gartner.

3. Number of households with smart home products and services in use worldwide from 2015 to 2025, March 17, 2022, Statista.

4. Gartner Forecasts Worldwide Artificial Intelligence Software Market to Reach $62 Billion in 2022, November 22, 2021, Gartner.

5. Number of households with smart home products and services in use worldwide from 2015 to 2025, March 17, 2022, Statista.

6. How Voice AI will drive the conversational AI space, Gaurav Singh, February 18, 2022, The Economic Times.

7. Meta Is Working On A Universal Speech Translator, March 10, 2022, Wersm.

8. What Is NVIDIA Maxine?, accessed on May 6, 2022, NVIDIA.

9. Google ups the AI ante with new features for Voice, Assistant, Maps and more, Taryn Plumb, May 11, 2022, VentureBeat.

10. Understanding the world through language, Zoubin Ghahramani, May 11, 2022, Google.

11. What Is NVIDIA Maxine?, accessed on May 6, 2022, NVIDIA.

12. Amelia Conversational AI, accessed on May 6, 2022, Amelia.

13. Project CAIRaoke: Building the assistants of the future with breakthroughs in conversational AI, Alborz Geramifard, February 23, 2022, Meta.

14. Enhancing Agent Experience With AI Contact Centers, Anand Santhanam, Sanjeev Bode, and Jitesh Gera, January 2022, Infosys.

15. Advances in multimodal understanding research at Meta AI, Laurens van der Maaten, Ishan Misra, Rohit Girdhar, et. al., March 16, 2022, Meta.

16. Google says its new image search features are like 'Ctrl+F for the world around you', Emma Roth, May 11, 2022, The Verge.

17. Autodraw, Google.

18. DALL·E: Creating Images from Text, January 5, 2021, Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, et.al., OpenAI.

19. Gartner Says Cloud Will Be the Centerpiece of New Digital Experiences, November 10, 2021, Gartner.

20. Gartner Says the Majority of Technology Products and Services Will Be Built by Professionals Outside of IT by 2024, June 14, 2021, Gartner.

21. Regional languages to account for 54 of OTT content by 2024, Lata Jha, March 28, 2022, Livemint.

22. The Curious Case of Transformations — Where's The Value?, Shyam Doddavula, Saibal Samaddar, Tanushree Halder, et. al., February 2022, Infosys.

Contributors

Venugopal Subbarao
AVP & Head, Digital Engagement CoE, Infosys Ltd.

Guruprasad NV
AVP, Digital Engagement CoE, Infosys Ltd.

Vishal Manchanda
Senior Principal Technology Architect, Infosys Ltd.

Samit Sawal
Product Technical Architect, Infosys Ltd.

Amit Kumar
Senior Technology Architect, Infosys Ltd.

Kavitha Sundararajan
Lead Consultant, Infosys Ltd.

Arpit Bhardwaj
Senior Associate Consultant, Infosys Ltd.

Producers

Jitesh Gera
Infosys Knowledge Institute

Kate Bevan
Infosys Knowledge Institute

About Infosys Knowledge Institute

The Infosys Knowledge Institute helps industry leaders develop a deeper understanding of business and technology trends through compelling thought leadership. Our researchers and subject matter experts provide a fact base that aids decision-making on critical business and technology issues.

To view our research, visit Infosys Knowledge Institute at infosys.com/IKI or email us at iki@infosys.com.