AISERA

WHITE PAPER SERIES

How AI Is Transforming IT Service Desk and Customer Service

Intelligent Knowledge Serving, Robotic Process Automation, Predictive Analytics, and Proactive Notifications

Antonio Nucci, Ph.D.

CTO, Aisera

www.aisera.com

AISERA

Introduction

In this fast-paced, competitive, and dynamic world, the need for speed is vital. Businesses want increased productivity with less resources, more cost-savings, and improved accuracy, to offer the ultimate customer experience. Customers increasingly expect an on-demand service with an instant response plus service—anywhere and anytime they want it. Many limitations to meeting such immediate, round-the-clock customer expectations relate to the people-centric way IT Service Desks and Customer Service are delivered nowadays. As such, there has been a surge of interest in conversational AI platforms that build on strengths of Artificial Intelligence (AI)—Natural Language Processing (NLP), Natural Language Understanding (NLU), and Robotic Process Automation (RPA).

The obvious question on everyone's mind is: will conversational platforms actually make IT Service Desk and Customer Service easier and more efficient?

Industry experts have strong opinions on this. Gartner predicts that by 2022, 40 percent of customer-facing employees and government workers will consult an AI virtual support agent daily for information or business process support and automation. Gartner adds that various AI technologies will help service agents respond faster and more efficiently. Additionally, AI virtual support agents will become the front line of service desks, performing actions that humans either can't do well, or would rather not do.

These actions fall into one of four categories: *intelligent knowledge serving*, *robotic process automation*, *predictive analytics*, and *proactive notifications*.

For example, conversational AI can crawl, normalize, and correlate today's structured and unstructured datasets that are scattered across multiple siloed knowledge base systems. They can augment such data with public data from companies, plus have it organized and readily accessible to employees and customers. Users can easily consume such knowledge by engaging in human-like interactions with conversational AI agents. Such agents are now powered by the latest Natural Language Understanding or NLU techniques to precisely extract user intents (known as ***knowledge intents***) from noisy, and often short sentences, and then serve the users with the right snippet of content.

By tapping into the disruptive technology of robotic process automation, conversational AI can enable Enterprises to quickly, more efficiently, and cost-effectively automate routinely tasks and processes that historically demanded the use of code or service agents. Employees and customers are finally equipped with a comprehensive, enterprise-grade process management and automation solution that streamlines and self-serves daily users' hurdles. These hurdles might be provisioning of an application (Zoom, Outlook, etc.), password management, and approval workflow automation, network troubleshooting (VPN, Wi-Fi, etc.), and more. Conversational AI makes it easy. Users can simply text, chat, or speak with a conversational AI agent, which understands their intents (known as ***action intents***) and automatically triggers the

Conversational AI helps with predictive analytics as well. For example, manually routing incoming service requests consumes a lot of time that a service agent could use for more important tasks. Till now, IT and Customer Service Desks have automated service request routing by defining rules that categorize requests based on preset conditions and parameters. However, these rules are static—meaning they won't adapt or improve with time. With the help of AI technology like machine learning (ML), service desks can now create a categorization model based on historic service desk data. They can then optimally assign service requests by taking into account service agents' past performance, workload, availability, expertise, and more relating to the specific type of request. Best of all, these ML models will become more accurate over time by taking live data into account. Such ML-based models are vastly more efficient than manual categorization or rule-based automations.

Last, Conversational AI helps with proactive notifications too! It can notify users about open service requests which may relate to their jobs, tasks, or locations, which translates into an immediate and comprehensive response to their inquiries. For example, let's say that a service request was opened for toner replacement on a printer located in a specific Enterprise building. Conversational AI can either proactively inform all employees located in the building about the temporary offline status of the printer needing toner replacement. Or, it can immediately inform users engaging with virtual agents that a service request was already created for that issue, providing them with the estimated time of resolution.

Why is Conversational AI happening now?

For starters, the infrastructure speed, availability, and sheer scale have enabled bolder algorithms to tackle more ambitious problems like Natural Language Understanding (NLU) and Natural Language Generation (NLG). Not only is the hardware faster, sometimes augmented by specialized arrays of processors (e.g., TPUs), it is also available in the shape of cloud services. What formerly ran in specialized labs with access to supercomputers can now be deployed to the cloud at a fraction of the cost and much more easily. Cheap storage, sometimes in the cloud, paired with intuition can collect every possible piece of data that will someday come in handy. This capability has brought about a high demand for solutions that go beyond simple statistical data analysis to promise new insights and intelligence. Because of unprecedented infrastructure speed and cheap storage, very sophisticated AI algorithms like Deep Learning (algorithms that mimic the human brain) have finally left their native academic lab environments and established an industry presence with a level of maturity that is well-understood and appreciated. By tapping into faster infrastructure and cheap storage, NLU models can now be scaled to several tens of hidden layer nodes, increasing their abstraction power and natural language understanding. They can be trained on tens of thousands of cores, speeding up the process of developing effective learning models.

As a result, much of the former operational resistance to enterprise AI adoption is now gone. More specifically, conversational AI has successfully demystified the following myths:

AI needs a lot of data and long training times. Historically, AI has imposed two stringent requirements to enterprises, i.e., the need of large datasets (hundreds of gigabytes) and long times (not hours but days!) to train models to the tasks of specific service desks. Today, novel techniques like transfer learning are available to tackle both challenges. They have proven to reduce training times to just a handful of minutes while requiring very little data from the enterprises. Transfer learning achieves this by splitting the model training into two phases: 1) data intensive and time consuming (feature training); and 2) light on data and super-fast in processing (classifier training). The first is executed on public data, pre-packaged and transferred as-is to the enterprise, while the second is used to customize the generic model to the specific service desk tasks.

AI needs a lot of operational supervision to produce good results. Over the last few years, there has been a big departure from supervised learning, which indeed required much human supervision—from data annotation to training to regular cycles for model calibration. The more advanced, but more complex, unsupervised learning has hence gained more and more adoption. With unsupervised learning, models can automatically learn the unknown by themselves and extract the meaning of it. Closed-loop automation can then be used to ensure models are regularly refreshed to assimilate new knowledge increments. This ability guarantees that models gets smarter over time and operate with fresh, current information.

AI cannot distinguish between good and bad learning. As the AI solution become less human- dependent (unsupervised learning), there is a common misbelief that learning could grow out of control, possibly leading over time to performance degradation. Reinforcement learning techniques have emerged to address this specific issue. Today, that learning is widely used in conjunction with unsupervised machine learning techniques. In reinforcement learning, an artificial intelligence agent employs trial-and-error to come up with a solution to any given model task. To make the model learn what the user would like to do, the agent gets either rewards or penalties for actions performed. The goal is to maximize the total reward, which translates to a model that adapts based on whatever the user acknowledges as a good action.

Choosing the right Conversational AI: basic versus advanced capabilities.

Nowadays, the market has seen an explosion of startups pushing virtual-agent solutions to serve industry use cases spanning IT service desk, customer service to HR, sales support, commerce, marketing, enterprise software front ends, advisory services, etc. But be aware: not all solutions providing virtual agent support are the same, and not all solutions deliver on their promises. The ability to understand and support complex interactions and transactions with customers and employees via virtual agents is largely determined by the sophistication of the artificial intelligence being employed in the underneath platform. To summarize the main differences between basic virtual agent support solutions and advanced conversational AI solutions, we focus next on some of the key building blocks: **dialogue management**, **user intent extraction** and **understanding,** and **continuous learning** and **self-adaptability**.

Dialogue Management. Most solutions in the market provide simple chatbot interfaces to users. These can only understand a limited number of user sentences, based on a set of pre-canned and static rules. The resultant user experience is a robotic interaction with end-users, with no ability to personalize either the served content nor the modality of interactions based on the user persona. Conversely, advanced conversational AI solutions can engage users in complex non-linear, multi-turn conversations. Here, the entire dialogue state—start to finish—is used to better contextualize user-system interactions and precisely identify when users are switching context and/or intents. Advanced conversational AI goes from one-shot single sentence understanding to multi-turn session dialogue management. With conversational-AI, users can open multiple simultaneous conversational flows while interacting with the virtual agent. They can pause, resume, or close any such flows at any time. For example, users can initiate a flow by requesting the status of his/her open tickets, then switch to request the provisioning of a zoom application. Later, they can return and add a comment to one of the open tickets previously listed. Conversational AI does this by embedding advanced natural language understanding techniques.

User intent extraction and understanding. A successful virtual agent must be able to hold a natural conversation with a human. Processing natural language is, however, not a piece a cake for Artificial Intelligence. When called to build an intelligent conversational AI system, one of the most challenging decisions is which Natural Language Processing algorithm to use (or more likely which bundle of algorithms) to correctly extract the user intents—those being knowledge or action intents. Most virtual agent technologies require users to speak, text, or chat precise sentences that the system can recognize. But, as the reader can see from Figure 1 (y-axis), the way users speak is not regular and robotic, but rather emotional and full of context. Advanced conversational AI solutions aim at correctly extracting and understanding user intents either from short, noisy user utterances (which are sentences in NLP terminology), or worse—hidden in long sentences, most of which contain meaningless details! By leveraging robust, complex neural networks powered by machine-learning models, advanced conversational AI solutions understand automatically the relevancy of certain parts of a user utterance. Thus, they can derive the correct user intent while filtering out the irrelevant portion. An example of such technology in action is shown in Figure 1, where the conversational AI agent was requested to correctly classify 60 user utterances pointing to 12 sample intent classes. As the reader can see, each user sentence was correctly classified into one, and only one, correct intent class. This offers the ability to deliver a smooth, fluid user-system dialogue flow without needing the virtual agent to repetitively ask the users clarification questions.

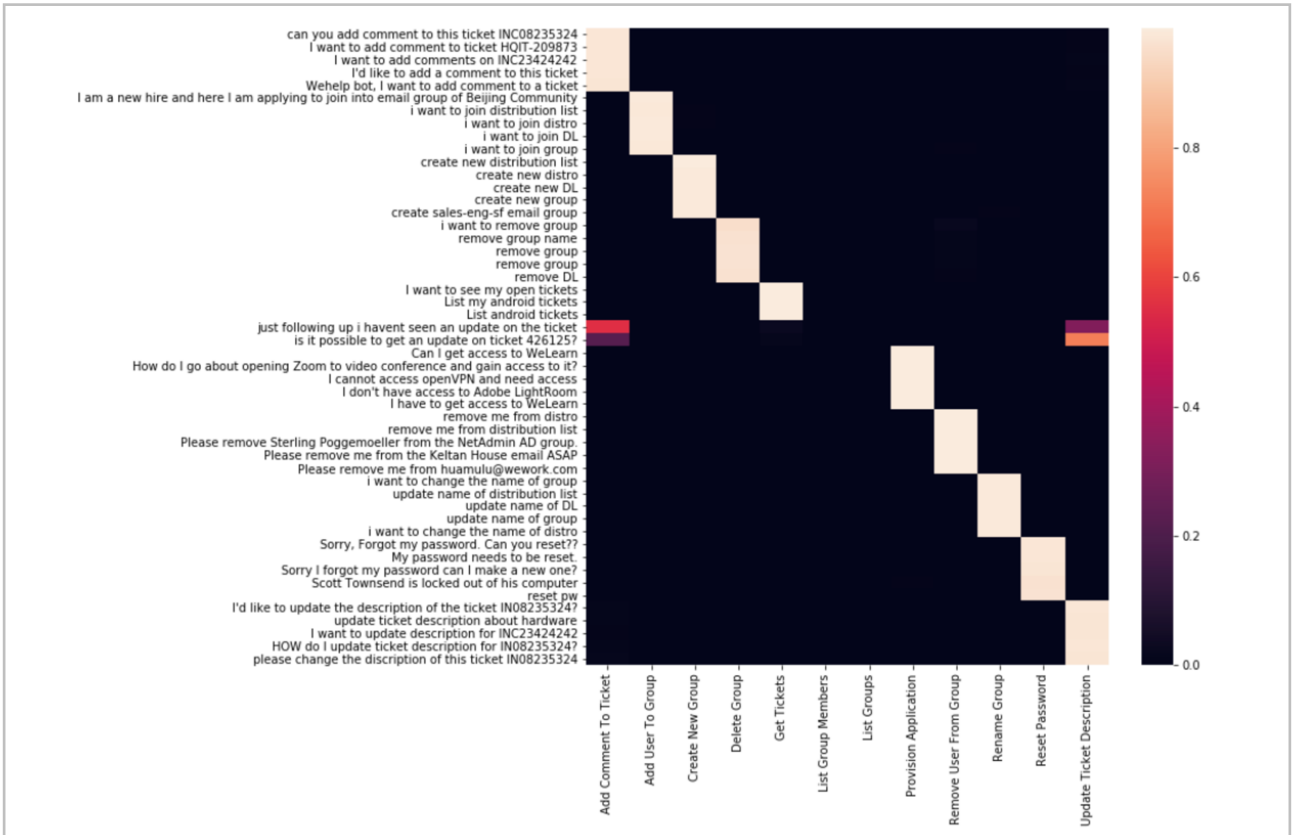

Figure 1. *An example of user intent classification executed using an advanced conversational AI solution. On the (y)-axis we show a sample of 60 user utterances inputted into a conversational AI agent. On the (x)-axis we show a sample of 12 intent classes that the conversational platform is capable of recognizing. The pixel color of the (x,y) entry correlates to the confidence score of a user-utterance (x) to belong to the intent class (y), with 0 (black) being the lowest confidence score and 1 (light brown) the highest confidence score. As the user can see, the advanced conversational AI solution is capable of precisely and correctly classifying each spoken user utterance (x) into one and only one intent class (y). Also, note how the solution is capable of detecting conditions of uncertainty when user utterances are not properly structured and could relate to more than one intent (confidence score around 0.5 from two distinct intent classifiers: "add comment to ticket" and "update ticket description").*

Continuous learning & self-adaptability. One major weakness of most available conversational AI solutions is their inability to understand when it is time to learn and how to learn new knowledge during the conversation process. Their knowledge is usually fixed beforehand and acquired only during the AI model training. As a result, it cannot be expanded or updated during the conversation and in an unsupervised fashion; thus requiring human supervision every time the models need to be retrained. Conversely, advanced conversational AI solutions are equipped both with sophisticated algorithms and jobs automation, which enable the system to continuously and interactively learn new knowledge during conversations.

As time goes by, they become progressively more knowledgeable and better at learning and interacting. This learning process resembles humans learning on the job—but it never stops. In Figure 2 we show an example of an intelligent conversational AI solution which automatically learns that it is time to learn, collects utterances that it cannot understand. It then clusters them together, based on similar intents, and automatically triggers a new round of AI model training without any human intervention, interaction, or supervision. The end result is a system which learns at every turn of every conversation with users, and grows continuously smarter.

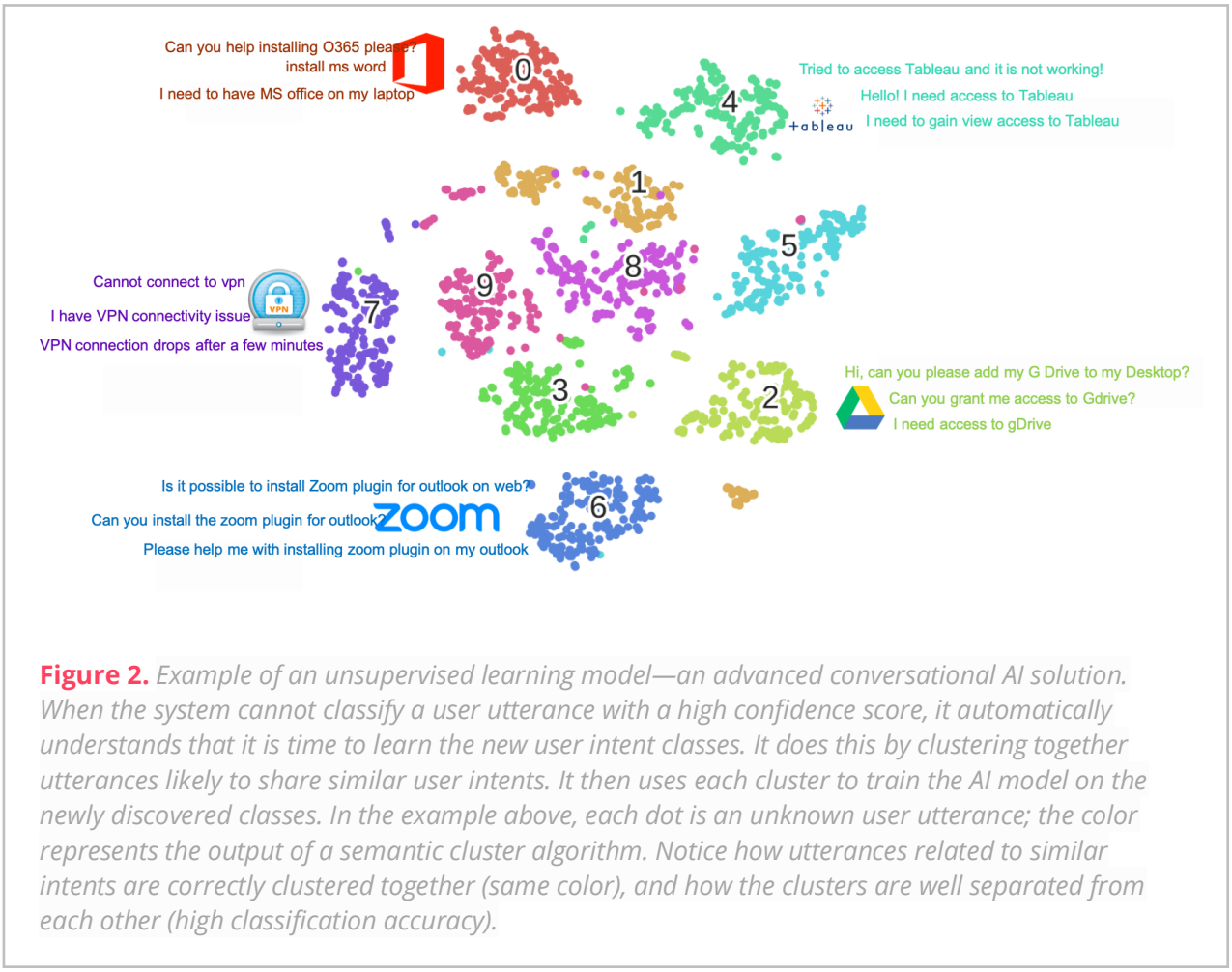

Figure 2. *Example of an unsupervised learning model—an advanced conversational AI solution. When the system cannot classify a user utterance with a high confidence score, it automatically understands that it is time to learn the new user intent classes. It does this by clustering together utterances likely to share similar user intents. It then uses each cluster to train the AI model on the newly discovered classes. In the example above, each dot is an unknown user utterance; the color represents the output of a semantic cluster algorithm. Notice how utterances related to similar intents are correctly clustered together (same color), and how the clusters are well separated from each other (high classification accuracy).*

Why AISERA?

AISERA, the World's First AI-Driven Service Desk, is enabling businesses and organizations to digitally transform yesterday's segregated and highly expensive service desks to today' AISM solution. AISERA offers a collaborative, all-in-one and cost-efficient enterprise- wise service desk. By building on modern technology like Robotic Process Automation, AI/ML and Conversational-AI, AISERA has repeatedly proven to deliver an unprecedented level of cost-efficiency and high-productivity. AISERA unleashes the true value of employee self-service and self-awareness, while facilitating cross-business domain collaborations and engagement.