CASE STUDY

Testing, Tuning and Monitoring for **CONVERSATIONAL IVR & ASSISTANTS**

Optimizing a Mission-critical Conversational Application

Bespoken

HIGHLIGHTS

6x

Scalability Increase

50%

Reduction ASR/ NLU Errors

90%

Reduction in Manual Testing Time

4

Weeks to implement

A Virtual Customer Agent for a critical use case

A POWER OUTAGE REPORTING

Bespoken recently worked with a major electrical utility company based in North America, which provides a variety of utilities across Canada and the US to more than 300,000 residences and 30,000 commercial businesses. As part of this service, they maintain a large call center that handles customer inquiries and responds to critical events impacting them, such as power outages and downed lines. The utility company recently decided to introduce a conversational agent as part of this call center. Initially rolled out to just a single, large market, this agent allows customers to report power outages as well as get information about repair times, all without requiring support from a customer service representative in their call center.

This workflow has strategic implications for the company – because power outages are intermittent in nature, having an automated agent to handle their reporting allows the company to ensure every customer can get through and make their reports in a timely manner. This means both not having to worry about call centers being flooded with calls when there is a surge in outages, and being able to leverage call center resources in a way that does not focus on low-frequency events. Simultaneously, it provides peace of mind that when those events do occur they can be handled gracefully. This simple virtual agent provides significant benefits to the utility company and its customers alike.

BESPOKEN, GENESYS AND TWILLIO

A POWERFUL COMBINATION

The team chose to leverage Twilio's state-of-the-art Autopilot platform for AI interactions with customers. Twilio Autopilot makes it easy to work across a variety of voice-based platforms, from IVR to Alexa to SMS and WhatsApp. It provides a consolidated, consistent way to interact with users across these channels. It ensures their natural speech patterns and language are understood. And it responds in a fast and friendly way to their inquiries. For this first version, the agent primarily handles customer reports of power outages. To capture these reports, the agent must gather from the caller their address and then confirm the details of their report. It's a fairly short workflow, but one that it is critical to get right.

ENSURING USERS ARE UNDERSTOOD, CONSISTENTLY, AND AT SCALE

Bespoken was brought in by the company and Genesys to test and tune the application prior to launch. They wanted to make sure that the assistant performed optimally and could handle the myriad types of queries coming from customers gracefully. Of particular concern was the handling of the over 400,000 addresses in their initial target market. When reporting a power outage, the most critical piece of the interaction is gathering a customer's address. So it essential these addresses are consistently understood. Additionally, Bespoken was tasked with creating automated functional tests and scalability tests. Because of the potential surges in reporting mentioned above, ensuring the system as a whole behaves well under heavy call volumes is critical.

OPTIMIZING UNDERSTANDING

The implementation is all based upon Bespoken's Virtual Devices for IVR. Using this solution, we tested tens of thousands of scenarios using a variety of speakers, accents, phrasings, and address variations, without anyone ever having to actually pick up a phone. Our virtual devices simulated real people calling the phone system and walking through real conversations.

We do this using a combination of generated audio, silence detection, and speech-to-text. The end result is effortless automated testing that delivers a higher level of understanding, quality, and throughput.
Using our virtual devices, we provided detailed, actionable findings for improving the Twilio Autopilot model. To improve the speech recognition accuracy of the conversational agent, we provided detailed, actionable findings for improving the utility company's Twilio Autopilot conversational model.

To improve the speech recognition accuracy of the conversational agent, we provided detailed, actionable findings for improving the utility company's Twilio Autopilot conversational model.

There were several challenges with handling addresses in the system, such as:

- Handling addresses with difficult to pronounce names
- Handling street addresses that number. For example, if I say "eighteen thirty-first street", does I mean 18 31st St or 1830 1st St?
- Handling unit numbers. This just further compounds the problem mentioned above, as the units can come before or after the address and are yet another number in the sequence. In the example above, it creates another possibility - 30 1st Street Unit 18.

THE THREE-PILLARS OF IVR QUALITY

Understanding

scalability

Monitoring

Scalability Testing

Equally important to ensuring users are understood is making sure that they can connect in the first place and that their queries receive prompt responses. To validate this, we simulated numerous customers calling simultaneously, closely monitoring and measuring the results.In our initial tests, we identified bottlenecks limiting the system to approximately 7 concurrent calls, with a maximum throughput of about 200 calls per hour.

This was far lower than the number expected and required during a large-scale outage. The team at the utility company took these findings and was able to remove these bottlenecks in the system. On subsequent tests, they were able to handle up to 30 calls simultaneously, with over 600 calls per hour. This increase in throughput means more calls get through and are handled automatically, saving time and money, and providing a faster and easier experience for customers.

> **Our initial tests showed that users were being understood approximately 85% of the time. By implementing the recommendations that came from our tests, the utility company was able to increase the level of understanding to 92% – that means we decreased the errors by nearly 50%. That optimization, which took only a few weeks, provides a significant, tangible reduction in the rate of incoming calls going to live customer agents, which is far more time-intensive and expensive.**

Ongoing Monitoring and regression testing

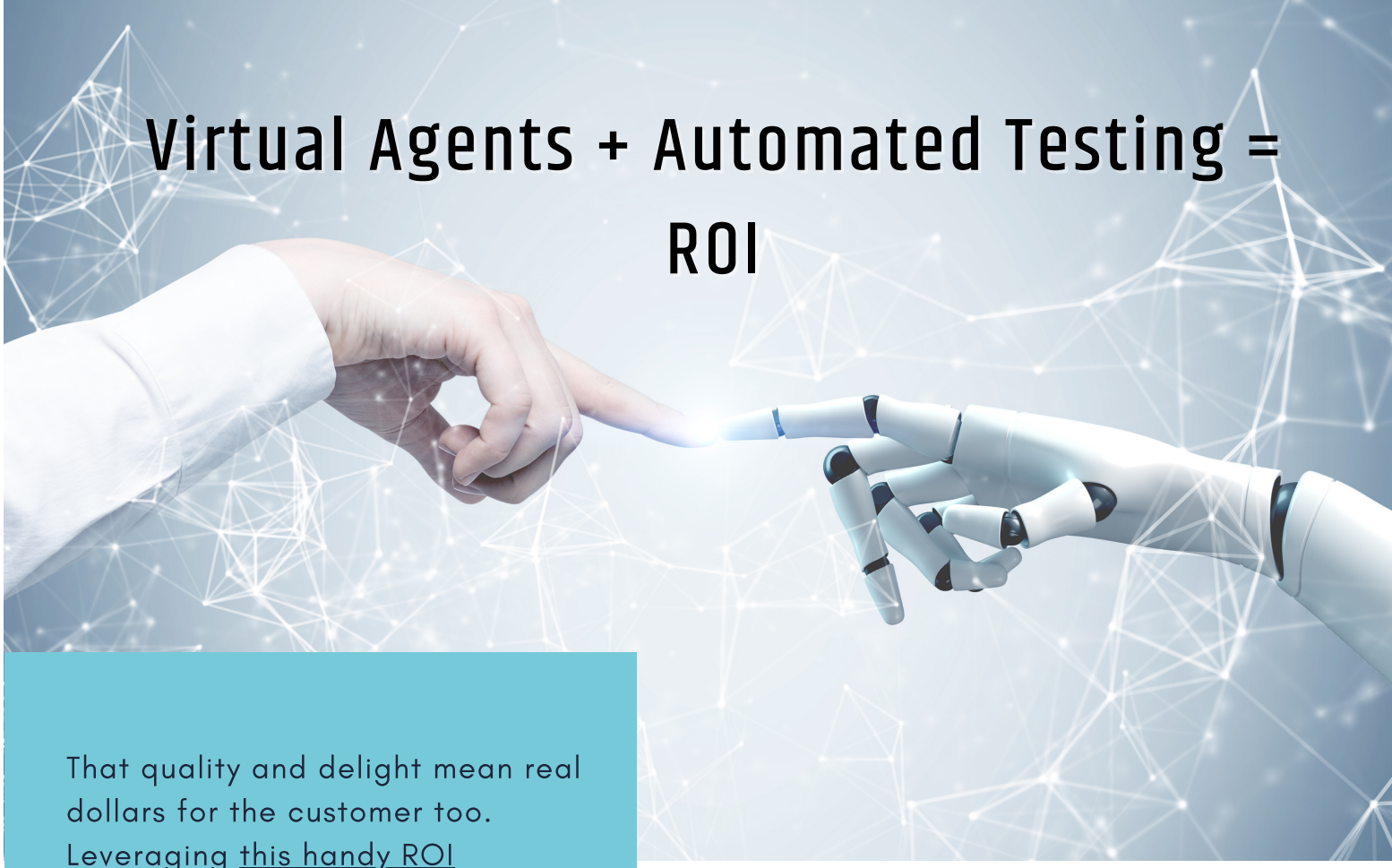
Tests were written initially to ensure the system behaved according to specification. These tests covered all the major functional paths in the system and were created by the Bespoken team working directly with key stakeholders from the power company's business team. The tests are easy to read and write and ensure all the essential interactions are working correctly. These tests were written and executed before the system went life, and as a result, several bugs and issues were identified. The development team worked to fix them, and automated regression tests verified the fixes prior to the application launching.

> Taken together, these three pillars are the foundation for long-term quality and customer delight.

THE ENVIRONMENT PROVIDES SEVERAL ADVANTAGES:

- Tests are run routinely to ensure the system is always working
- In-depth reports are created for every test run and automatically emailed to key stakeholders
- Historical data are automatically captured in a visual reporting environment to easily monitor and analyze long-term trends and issues

Virtual Agents + Automated Testing = ROI



That quality and delight mean real dollars for the customer too. Leveraging this handy ROI calculator provided by Twilio, we were able to estimate the potential financial benefits for the company over the next three years. According to the calculator, the return is over 100% for that period.

Overall, the potential savings from deploying a high-quality customer agent are immense, and we at Bespoken are proud of the work we did to enable it.

These numbers are based upon industry-standard costs related to call-center staffing - in this case, the system pays for itself in less than 12 months. What is more, these figures do not even include the potential benefit of increased conversion rates for customers - while that may not apply for this particular use case, for others it can more than quadruple the ROI.

To work with Bespoken, and gain these benefits for your own project, email us at contact@bespoken.io