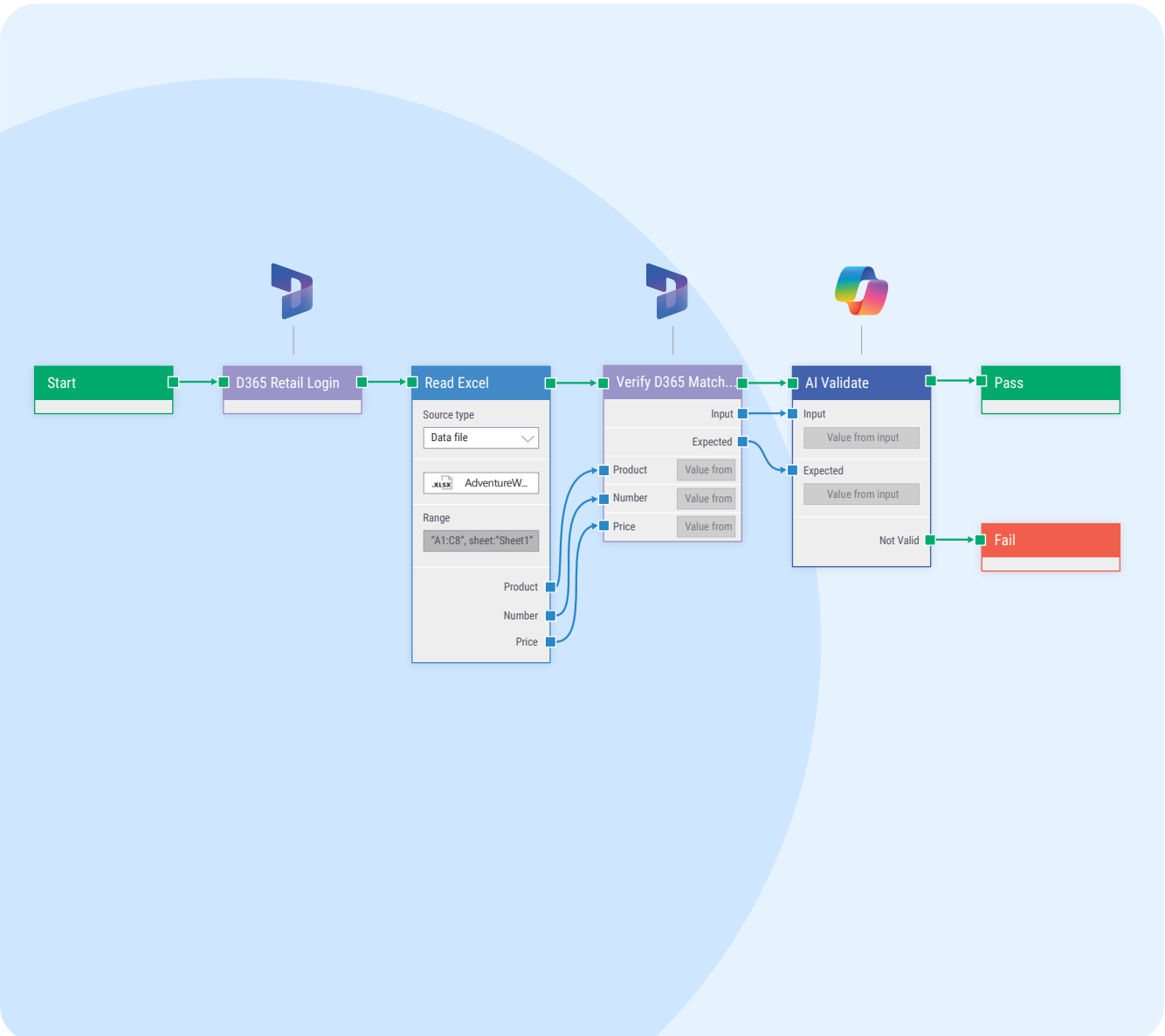
leapwork + Microsoft

Microsoft Copilot Automated Testing with Leapwork

Best practices, strategies, and key milestones to ensure a successful deployment process



This whitepaper is intended for businesses preparing to implement automated testing of Microsoft Copilot with Leapwork. It describes best practices, strategies, and key milestones to ensure a successful deployment process.

Key stakeholders, including Engineering, Application, and IT leaders, should use this whitepaper to evaluate their current operations and create a plan to integrate Copilot and Leapwork capabilities seamlessly.

Contents

Introduction

In today's fast-paced customer service landscape, the integration of advanced AI technologies like Microsoft Copilot can significantly enhance efficiency and accuracy. Automating the testing of these AI systems with tools such as Leapwork is crucial to maintaining high performance and reliability. This whitepaper aims to provide a comprehensive framework for implementing automated testing, ensuring that Copilot's deployment meets organizational goals and drives better customer service outcomes.

Responsible AI

At Microsoft, we adhere to three fundamental promises with our cloud services:

→ Your data is yours to own and control.

→ Your data is not used to train or enrich foundation AI models used by others.

→ Your data and AI models are protected by comprehensive enterprise compliance and security controls.

Ensure that your system is designed to keep data secure, adhering to Responsible AI principles and compliance with regulations such as GDPR.

Learn About Copilot Features

Copilot provides several key use cases for customer service:

→ **Ask a Question:** Contextual responses based on internal and external knowledge sources.

→ **Write an Email:** Assists in quickly creating email responses.

→ **Summarize a Case:** Offers a summary of case details.

→ **Summarize a Conversation:** Provides summaries at key points during the customer journey.

Prepare Your Organization

Identify Metrics to Measure Success

Set business goals to determine which metrics will help you measure success. Gather metrics through:

→ Out-of-the-box reports in Dynamics 365 Customer Service Historical Analytics.

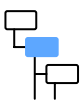
→ Feedback from Dataverse tables.

→ Custom Power BI reports.

→ Engagement with the Microsoft product team.

Adjust Your Knowledge Management Strategy

To leverage Copilot's capabilities, ensure that knowledge content is accurate and well-maintained. Use the 6Ds framework for deploying enterprise content and follow best practices for creating and maintaining your knowledge repositories. Learn more about the 6Ds framework here (The 6Ds).



Tag and Filter Knowledge Content

Create structured content tags for easy updates and retrieval. Use filtering to improve the quality of Copilot responses. Here is a learn doc with more information: Use Copilot filters | Microsoft Learn.

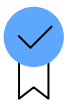
Set Up Your Tools and Infrastructure

Prepare your contact center environment by configuring the Customer Service admin center, mapping custom fields, and setting up security roles and integrated search providers.

Onboard Your Users

Create a Change Management Strategy

Develop a plan to manage the transition to using Copilot, involving all relevant teams in the initial pilot phase.



Review Your Corpus of Knowledge

Ensure that your knowledge articles are of high quality and aligned with Copilot's requirements.

Train Your Agents to Use Copilot Effectively

Provide specific training to help agents use Copilot efficiently. Use detailed prompts to get the best responses from Copilot.

Create a Testing Strategy

Develop a robust testing strategy to validate Copilot's performance. Use Leapwork for automated testing to ensure the accuracy and effectiveness of Copilot's responses.

Key Considerations for Automated Testing with Leapwork

→ **Default Data Refresh Interval:** Note that 15 minutes is the default interval for refreshing data, contrary to the real-time updates some may expect.

→ **Daily Testing Recommendation:** It is recommended to run the testing pack daily, ideally as a pre-start of day check, to ensure that the responses from Copilot are not outdated.

→ **Leapwork Reporting:** Leapwork reporting output can be linked to the tested prompts for an accuracy grade or rating, providing a clear measure of Copilot's effectiveness.

Optimize Your Implementation

Gather Feedback from Your Agents

Use the Helpful Response Rate (HRR) to gather feedback and identify areas for improvement.

Measure Ongoing Performance

Track the performance of your knowledge content and Copilot's effectiveness. Use analytics to monitor usage and performance.

Success Metrics for Copilot Implementation

To ensure the success of Copilot integration, it is essential to track and measure the following metrics:

→ **Time Efficiency:** Measure the amount of time Copilot saves agents in performing their tasks by comparing the time taken to complete tasks with and without Copilot.

→ **Relevance and Helpfulness of Responses:** Evaluate Copilot's responses based on their relevance and helpfulness. Categorize responses as totally irrelevant, partially helpful, mostly helpful, or completely helpful.

→ **Agent Satisfaction and Ease of Use:** Assess how user-friendly and intuitive Copilot is for customer service agents. Agent satisfaction can be a key indicator of its usability and effectiveness.

→ **Impact on Customer Satisfaction:** Monitor changes in customer satisfaction metrics through surveys or customer feedback to see if there is an improvement due to Copilot implementation.

→ **Return on Investment (ROI):** Evaluate the overall costs versus the benefits of implementing Copilot, considering the time saved and improvements in customer and agent satisfaction.

Summarization Problem

Summarization within Microsoft Copilot uses a combination of structured and unstructured data from a support case record. Structured data includes predefined fields such as case numbers, timestamps, and customer details, while unstructured data comprises case notes, email threads, and chat logs.

Key Challenges

→ **Data Integration:** Tests must verify that key facts from both structured and unstructured data are present in the generated summaries.

→ **Context Preservation:** Traditional tools may struggle as they often check for simple text strings, which can lead to false positives or negatives. The context in unstructured data is critical, making it challenging to ensure that the summarization captures the essence of the information accurately.

→ **Accuracy of Summarization:** Ensuring that the summary is both concise and comprehensive, accurately reflecting the essential details without omitting critical information.

Testing Approach

→ **Automated Validation:** Utilize Leapwork to automate the validation of summaries, comparing them against expected outputs.

→ **Context-Aware Testing:** Develop tests that consider the context of unstructured data to avoid false results.

→ **Metrics for Evaluation:** Implement evaluation metrics like ROUGE to measure the quality and accuracy of the summaries.

By addressing these challenges, organizations can ensure that the summaries generated by Copilot are reliable and useful for agents, thereby improving overall customer service efficiency.

Data Extraction Problem

Verifying accuracy for issue resolution, which relies on underlying knowledge articles, requires testing of the data extraction process and summarization of the answer. The accuracy of answers depends on both the large language model and the quality of the knowledge base articles. Questions and documents become inputs to the process, and regular testing of these use cases helps to identify accuracy drift as documents become out of date.

Key Challenges

→ **Prioritizing Common Questions:** Mapping expected answers to the most frequently asked questions to ensure accuracy.

→ **Unstructured Data Complexity:** Extracting accurate information from unstructured data such as emails and chat logs can be challenging due to the variability and context-dependence of the information.

→ **Validation of Extracted Data:** Ensuring that the data extracted is relevant and correct, and aligns with the structured data available.

Testing Approach

→ **Initial Test Packs:** Run initial test packs to validate the system's ability to extract and provide correct answers.

→ **Continuous Monitoring:** Regularly monitor ongoing performance to ensure consistency and accuracy in the responses.

→ **AI Validation Tools:** Utilize AI tools like Leapwork to automate the validation process, checking both the accuracy and context of the extracted data.

→ **Regular Updates:** Keep knowledge base articles up-to-date to prevent accuracy drift, ensuring that both the language model and the underlying data are current.

By focusing on these challenges, businesses can improve the reliability of Copilot's issue resolution capabilities, ensuring that customer queries are answered accurately and efficiently.

Q&A

Q: How is Copilot updated with new documentation?

A: Copilot can leverage plugins to pull data real-time status from another system. Updates to Dataverse Knowledge Articles are reflected in Copilot's responses after automatically refreshing this content every 15 minutes.

Q: How often should we run automated tests with Leapwork?

A: It is recommended to run automated tests on a continuous basis to match the Copilot refresh interval. Running at least once daily, as a pre-start of day check, will help to ensure that the responses from Copilot are accurate and up-to-date. Regular testing helps identify and resolve issues promptly, maintaining the quality and reliability of the system.

Q: What types of tests can we automate with Leapwork for Microsoft Copilot?

A: You can automate a variety of tests including functional tests, regression tests, performance tests, and accuracy tests of Copilot's responses. Leapwork allows you to create custom test scenarios tailored to your specific needs.

Q: How can Leapwork help in validating Copilot's responses?

A: Leapwork can automate the process of validating Copilot's responses by running predefined test scenarios and comparing the actual unstructured data outputs with expected results. The reporting output from Leapwork can be linked to the tested prompts for an accuracy grade or rating, providing a clear measure of Copilot's performance.

Q: What is the process for setting up Leapwork for automated testing with Microsoft Copilot?

A: Setting up Leapwork involves several steps:

1. Define the test cases and scenarios you want to automate.
2. Configure Leapwork to interact with Microsoft Copilot and other relevant systems.
3. Create automated test workflows using Leapwork's visual interface.
4. Schedule the tests to run at regular intervals (e.g., daily).
5. Monitor and analyze the test results to identify and address any issues.

Q: How do we handle data privacy and security when using Leapwork and Copilot?

A: Both Leapwork and Microsoft Copilot are designed with robust security and privacy features. Ensure that your data is handled according to your organization's data privacy policies and compliance requirements. Use secure connections, restrict access to sensitive data, and regularly audit the system for any potential vulnerabilities.

Q: Can we integrate Leapwork with other tools and platforms we are using?

A: Yes, Leapwork supports integration with various tools and platforms such as CI/CD (i.e. Microsoft Azure DevOps), bug tracking, Power BI, and other enterprise systems. This allows you to create a seamless automated testing environment and leverage existing tools for comprehensive analysis and reporting.

Q: What kind of support and resources are available for implementing Leapwork with Microsoft Copilot?

A: Both Microsoft and Leapwork provide extensive documentation, tutorials, and customer support to help you implement and optimize the integration. Additionally, you can engage with their professional services teams for personalized assistance and guidance.

Q: How do we measure the success of our automated testing strategy?

A: Success can be measured using various metrics such as test coverage, defect detection rate, test execution time, and the accuracy of Copilot's responses. Regularly review these metrics to assess the effectiveness of your automated testing strategy and make necessary adjustments.

Conclusion

By integrating Microsoft Copilot with Leapwork, organizations can automate testing processes, ensuring higher accuracy and efficiency in customer service operations. This whitepaper provides a comprehensive framework to implement and optimize this solution, driving better outcomes for businesses and their customers.

Learn more: leapwork.com

leapwork + Microsoft