Car Insurance Virtual Agent

1. Executive Summary

The "Car Insurance Virtual Agent" is a cutting-edge, AI-based solution that has the potential to transform customer support in the auto insurance industry. Developed on IBM watsonx Agent Lab, using strong Granite foundation models (with Granite 3-3-8b being the main model for the agent's core capabilities), this virtual agent will greatly alleviate call centre waiting times, enhance customer satisfaction, and give instant accurate answers to frequent customer queries. Through automating simple processes such as policy search, claim categorization, and general advice, the agent allows human capital to be used on more complex cases, making for a more efficient and sympathetic customer experience.

2. Problem Statement

Automobile insurance companies are presented with the major problem of handling the high number of daily customers enquires about policy information, claims, renewals, and complaints. Old-fashioned call centres tend to be overworked, resulting in:

- Long Hold Times for Customers: Customers get frustrated with prolonged delays in responding to straightforward requests, such as inquiring about policy information or clearing up claims.
- **Inconclusive Information:** Overworked human agents may give different degrees of detail or slightly divergent responses, creating inconsistencies.
- Excessive Load on Human Staff: Much agent time is consumed by repetitive, low-complexity interactions, taking them away from more important or understanding interactions.
- Customer Unhappiness: Delays and inefficiencies directly affect customer loyalty and trust.
- **Absence of Proactive Support:** Manual processes do not allow for timely follow-ups or proactive support for customers.

These problems in total result in customer frustration, elevated operational expenditure, and a bottleneck in service delivery, emphasizing the necessity for automation and smart assistance.

3. Solution Overview: The AI-Powered Virtual Agent

Our solution is a virtual agent powered by AI developed with IBM Watson Agent Lab. Its fundamental function is to automate and simplify the management of typical car insurance inquiry questions, providing an extensive and real-time support avenue.

Key Features:

- **Policy Number Extraction:** Reliably determines and extracts policy numbers from natural language customer messages (e.g., "My policy number is 7788XZ").
- Issue & Sentiment Classification: Categorizes customer questions into precise issues (e.g., accident, theft, glass breakage) and also identifies sentiment (e.g., "I am not satisfied with my previous claim" identified as negative sentiment).
- Natural Language Responses & Advice: Offers succinct, clear, and natural language answers, including advice for the next steps (e.g., how to retrieve policy information, whom to call for certain claims).
- Advanced AI Capabilities: Utilizes Granite foundation models (mainly Granite 3-3-8b for agent tasks, with llama-3-3-70b-instruct applied to prompt creation) for advanced natural language comprehension and creation.
- **Knowledge Retrieval (Google Search):** Includes a Google Search feature to retrieve more public data as needed, enriching the agent's knowledge database beyond coded answers.
- **Email Follow-up Composition:** Is capable of summarizing dialogues and composing draft follow-up emails to customers, raising post-interaction support.

Advantages:

• Rapider Responses: Offers immediate responses, cutting down greatly on customer wait times.

- Reduced Workload for Human Agents: Reduces workload for human agents by automating standard queries, enabling human staff to concentrate on intricate, high-priority cases.
- Same Responses: Guarantees consistency and accuracy in response given to customers.
- **Scalability:** Developed on IBM Watson foundation models, enabling seamless scalability to manage rising volumes of queries.
- **Customer Better Experience:** Facilitates greater customer satisfaction and confidence as a result of effective and accurate service.
- **Future Integration:** Built with the ability to connect to secure backend systems (such as claims databases) and live production APIs to provide enhanced functionality.
- **Empathetic Interactions:** Directed by explicit instructions to stay helpful and courteous, fostering more customer trust.

4. Technical Architecture

The Virtual Agent for Car Insurance is designed for intelligent, multi-turn dialogue and external knowledge retrieval.

- **Agent Framework:** Lang Graph
 - Selected for its capacity to describe flexible, multi-step conversational paths, enabling sophisticated decision-making and tool use.
- Architecture: ReAct (Reasoning and Acting)
 - This architecture supports the agent in reasoning dynamically about the ongoing conversation state, determining which tools to employ (if any), and producing corresponding responses, resulting in smarter and more adaptive conversations.
- **Foundation Model:** Granite 3-3-8b
 - The agent's central intelligence, tasked with comprehending user intent, extracting information (such as policy numbers), classifying problems/sentiment, and producing natural language responses. Note: Prompt testing during development was also performed using llama-3-3-70b-instruct.

• **Tools:** Google Search

O Used to bring outside, real-time knowledge into the agent. If the internal knowledge base of the agent is inadequate, it can use Google Search to locate pertinent public information and deliver a complete answer.

Flow Summary

- 1. User Message Received: A user starts a discussion with the agent through the agent preview interface.
- **2. Agent Processing:** The agent gets the user message.
- **3. Instruction & Prompt Application:** The agent applies its pre-defined instructions and context-dependent prompts.

4. Tool Utilization (ReAct Loop):

- Depending on the input from the user and its own reasoning, the agent determines whether to extract a policy number, categorize an issue (such as accident, theft, glass claim, sentiment), or look up information.
- If it requires external knowledge, the agent invokes the Google Search tool.
- The tool's results (if invoked) are input back into the agent's reasoning.
- **5. Response Generation:** The agent combines its knowledge, extracted data, and any data it has retrieved to create an overall and appropriate response.
- **6. Return to User:** The agent sends back its response to the user.

This structure allows the agent to receive a large variety of queries, ranging from elementary information retrieval to more sophisticated, knowledge-dependent interactions.

5. Agent Instructions and Features Exhibited

5.1. Agent Instructions

The behavior of the agent is instructed by a specific set of guidelines to maximize helpfulness,

consistency, and politeness:

You are an assistive helper that employs tools to provide detailed answers.

When a greeting is received, respond with "Hi, I am watsonx.ai agent. How can I help you?"

These guidelines define the persona and first interaction of the agent, creating a warm tone for the user.

5.2. Features Demonstrated

The virtual agent effectively illustrates a number of key capabilities:

• Policy Number Extraction:

- Ability to correctly identify and extract policy numbers contained within customer statements.
- Example: User input: "My policy number is 7788XZ" → Agent Outcome: Successfully extracts "7788XZ".

• Issue Classification:

- Ability to categorize customer problems into pre-defined categories, required for routing and giving appropriate advice.
- Example 1: User input: "My windshield is broken" → Agent Outcome: Categorized as "glass damage".
- Example 2: User input: "I had an accident" → Agent Outcome: Categorized as "accident".
- Example 3: User input: "My car was stolen" → Agent Outcome: Categorized as "theft".

• Sentiment Classification:

- o Capability to perceive the emotional sentiment of a customer's message.
- Example: User input: "I am unhappy with my last claim" → Agent Outcome:
 Identified as "negative sentiment".

• Guidance and Suggestions:

 Gives doable next steps and overall guidance pertaining to policy management and claims. Example: Provides examples of how to verify policy information or whom to contact for certain kinds of claims.

• Web Resource Retrieval (Google Search):

 Effectively employs the Google Search function to retrieve public information when a query calls for external information outside its immediate scope.

• Email Follow-up Generation:

o Illustrates the ability to paraphrase interactions and assist customers in writing follow-up emails, augmenting holistic support.

6. Testing & Results

The virtual agent was comprehensively tested inside the IBM watsonx Agent Lab preview environment through a series of conversational examples. The findings invariably proved the agent to be effective and reliable in its main features.

6.1. Test Examples Utilized in Preview

Some of the test cases utilized during the validation and development phase are given below:

• Policy Number Extraction:

- o User query: "My policy number is 7788XZ"
- o Expected Outcome: The agent should accurately pull out "7788XZ".

• Issue Classification (Accident):

- User query: "I had an accident"
- o Expected Outcome: The agent should classify the issue as "accident".

• Issue Classification (Glass Damage):

- User query: "My windshield is broken"
- o Expected Outcome: The agent should classify the issue as "glass damage".

• Issue Classification (Theft):

- User query: "My car was stolen"
- o Expected Outcome: The agent should classify the issue as "theft".

• Sentiment Classification:

O User query: "I am unhappy with my last claim"

 Expected Outcome: The agent should classify the sentiment as "negative sentiment".

6.2. Outcome

- The testing produced positive and uniform results in all illustrated features:
- The agent uniformly parsed policy numbers correctly from different user inputs.
- It consistently categorized insurance matters into their corresponding categories, demonstrating its familiarity with various kinds of claims.
- The agent successfully identified user sentiment, making it possible to respond with more varied and suitable answers.
- It generated courteous and pertinent responses, providing useful next steps and advice as intended by its commands.
- The Google Search tool integration worked properly, enabling the agent to increase its knowledge base dynamically whenever necessary.

These results verify that the Car Insurance Virtual Agent can support common customer inquiries effectively and accurately, functioning as intended to mitigate call center overload and enhance customer satisfaction.

7. Benefits and Advantages

Car Insurance Virtual Agent provides substantial benefits to customers and car insurance firms:

7.1. For Customers

- Faster Responses: Offers immediate replies to regular questions, ending long waiting times on the phone and instant access to data.
- 24/7 Support: Customers can have assistance at anytime, anywhere, improving convenience.
- Same Information Everywhere: Guarantees that every customer gets consistent, accurate, and current information, avoiding confusion.
- Better Experience: A smooth and effective support mechanism results in greater customer

- satisfaction and confidence.
- **Proactive Assistance:** Can walk users through steps or offer next steps, empowering users with knowledge.

7.2. For Businesses (Car Insurance Companies)

- Reduced Operational Costs: Automating repetitive queries decreases the burden on human agents, resulting in lower staffing requirements for routine tasks.
- Optimized Resource Allocation: Human agents can dedicate more time to complicated, sensitive, and high-value cases that genuinely benefit from human empathy and intelligence.
- Scalability: Using IBM watsonx foundation models enables the agent to process rising query volumes without corresponding increases in human resources.
- **Data-Driven Insights:** Agent interactions can yield useful data on frequent customer pain points, which can be used to make services better.
- **Improved Brand Image:** Providing advanced AI-fueled support establishes the company as forward-thinking and customer-focused.
- Future Integration Capability: The agent design enables easy integration with secure backend claims databases and other production APIs, enabling more automation and efficiency benefits.
- **Emotional Interactions:** With repetitive queries offloaded, human agents would have more time to provide empathetic support for intricate matters, further enhancing customer relationships.

8. Limitations

Though the Car Insurance Virtual Agent managed to showcase its main functionalities successfully, some limitations were faced, mainly because of the hackathon environment and the scope:

• **IBM Cloud deployment:** Complete deployment of the agent as a single AI service was beyond the direct scope of the hackathon. This was primarily because of environmental

limitations, namely the "catalog disabled" state of the IBM account that wouldn't allow for the creation of a Cloud Object Storage (COS) instance – a requirement for deploying AI services in IBM Cloud.

- **Demonstration Environment:** The agent's capabilities were mainly demonstrated and tested under the IBM watsonx Agent Lab preview environment. This configuration nicely demonstrates the agent's capabilities but is not a production implementation.
- **Backend Integration:** The version in hand leverages Google Search for external knowledge. It does not yet utilize secure, proprietary claims databases or other internal production APIs, which would be necessary for production policy lookups and claims processing.
- **Detailed Classification Models:** Although issue classification is shown, for actually sophisticated and subtle cases, personalized classification models would be needed, which fell outside the time horizon of the hackathon.

These constraints point to directions for future work and are the next natural steps towards a complete, full-featured virtual agent.

9. Future Work

To develop the Car Insurance Virtual Agent to production quality, the following essential steps are envisaged:

• Secure Backend Integration:

- Connect to a secure claims database: This is most important for real-time policy lookups, claims status inquiries, and tailored customer engagement beyond public data.
- Integrate with real production APIs: Allow the agent to take actions like opening claims, updating policy information, or submitting official documents straight through company systems.

• Advanced AI Capabilities:

o Insert custom classification models: Create and implement sophisticated machine

- learning models to classify customer issues into more detailed and nuanced categories, aside from the overall categories.
- Incorporate voice-based capabilities: Integrate with speech-to-text and text-to-speech capabilities to support phone support, so that customers can communicate verbally with the agent.
- o Insert multilingual support: Enhance the abilities of the agent to comprehend and answer in various languages to accommodate an international customer base.

• Deployment & Scalability:

- Complete IBM Cloud Deployment: Deploy the agent as an AI service on IBM Cloud, allowing it to act as a high-performance, scalable API endpoint that can be accessed by multiple front-end applications.
- o Front-end Integration: Integrate the deployed agent with multiple customer-facing channels, including website chat widgets, mobile apps, or IVR systems.

• Improved Personalization:

 Create features for extended personalization using customer history, preferences, and prior interactions.

• Continuous Learning & Improvement:

 Adopt systems for the agent to learn from ongoing interactions, feedback, and fresh data continuously in order to enhance its accuracy and usefulness over time.

These subsequent steps will turn the prototype into a full-fledged, enterprise-level virtual agent that can bring vast value to customers as well as the business.

11. Screenshots





