July 2023

INTELLIGENT AUTO-LEARNING GENERATIVE KNOWLEDGE FINDER FOR PROACTIVELY AIDING A VIRTUAL AGENT

Anupam Mukherjee

Nala Prasad Kamath

INTELLIGENT AUTO-LEARNING GENERATIVE KNOWLEDGE FINDER FOR PROACTIVELY AIDING A VIRTUAL AGENT

AUTHORS:
Anupam Mukherjee
Nala Prasad Kamath

ABSTRACT

Within a contact center, many caller and virtual agent (VA) interactions eventually reach a human agent (HA) for the proper resolution of a caller's issues. This may happen despite the presence of a state-of-the-art artificial intelligence (AI)-powered VA component (that can, for example, search an organization's knowledge bases (KBs) for the most relevant answers corresponding to a caller's queries) through, for example, outdated KBs. As a result, a caller's experience is negatively impacted. Techniques are presented herein that support an intelligent, proactive, auto-learning generative knowledge finder (KF) component which can be leveraged by a VA to automatically improve its competency level even if an organization's KBs are not current. Such a KF component may extract information from past caller-HA interactions and enrich itself with the help of an N-shot learning paradigm. Thus, without any manual intervention, a KF component, powered by generative AI, can enhance itself and, in turn, a VA's competency level beyond the VA's original intelligence (which may have been acquired through training and knowledge management system (KMS) access), exhibit a highly optimized reaction time, and dramatically lessen a HA's involvement.

DETAILED DESCRIPTION

Within the contact center arena, a new focus is on self-learning, whereby it is possible to learn from customer conversations and thus empower a contact center with new (e.g., self-service) capabilities. For example, a recent study found that such an approach is the only long-term solution that already meets customer expectations, with 70% of consumers reporting that they expect a self-service option will be available to them for resolving their questions and complaints.

However, many current caller and virtual agent (VA) interactions eventually reach a human agent (HA) for the proper resolution of an issue. While a state of the art VA that is powered by artificial intelligence (AI) may have the ability to search an organization's knowledge bases (KBs) for the most relevant answers corresponding to a caller's queries, on many occasions a VA fails to find an accurate resolution due to outdated KBs. Such KBs must be constantly updated for a VA to work efficiently, but that is a tedious and time-consuming effort that requires dedicated manual intervention. As a result, VA efficiency does not improve and a caller's experience suffers, eventually leading to more HA involvement through escalations and impeding the vision of true self service.

Currently, some potential mechanisms may seek to increase the competency of a VA in a VA-based self-service flow. Such mechanisms are primarily analytics- and observation-driven approaches, under which the entire journey through a flow is instrumented to capture various metrics and AI-based session characterization (such as sentiment analysis, intent recognition, topic modelling, etc.) can help an administrator detect any flaws in customer handling alongside other important attributes. Based on such findings, appropriate measures may be taken including enhancing an existing VA or introducing a more efficient VA, enriching a knowledge management system (KMS) with proper content, and so on.

A first potential mechanism employs a self-learning knowledge base (SLKB) in a self-service flow within a contact center. Such a facility learns often-repeated responses thus saving agents from needing to answer repetitive questions, empowers a team to focus on more complex issues, and gets a team out of a reactive mode and into a forward-thinking one. While aspects of the techniques presented herein (which will be described and illustrated in the below narrative) may employ a SLKB, the presented techniques are more advanced and algorithmically superior to this first existing mechanism.

In particular, the presented techniques ensure minimal HA involvement and focus on resolving complex issues which a VA cannot handle and for which HA intervention is required. The techniques work as a supplemental model to a VA (in addition to a SLKB if such a facility is present) which operate on a predesigned flow and often-repeated queries with the help of a client KMS (which may be instrumented with a SLKB).

Importantly, a SLKB is typically designed only for the resolution of routine and/or often-repeated queries and topics. Consequently, it cannot be used to sort out complex issues, as it does not encompass generative intelligence, and its intelligence relies on a historical data-driven process to enrich itself. Moreover, without the help of a knowledge engineer a SLKB cannot learn from complex issues. In contrast, the presented techniques incorporate generative intelligence, not only learning from history but also applying that learning to resolve unknown queries. Importantly, the presented techniques may supplement a SLKB to update a customer's KMS. In such a case the SLKB can also be used to resolve complex issues in addition to often-repeated queries.

A second potential mechanism may encompass a fully distributed, adaptive, and scalable system for learning the types of anomalies that are of interest to users of a distributed learning system. This information is then used for deciding which anomalies are forwarded to the user, improving bandwidth utilization, thus filtering anomalies that are considered a false positive. The mechanism also has a forwarding facility that effectively gathers user feedback and dynamically learns how to select the anomalies that are of interest. In contrast, the presented techniques are distinctly different as to both the algorithm (an N-shot learning paradigm) that is employed and the type of data (structured or unstructured, such as transcripts, KBs, etc.) that is considered.

A third potential mechanism may encompass a VA-like entity that is designed to store the answer of an unknown query and provide the answer in response to similar questions in the future. The mechanism works like a simple caching medium, responding from the cached data (by itself or through other similar entities) for the known questions and, in the event of a miss, caching a new answer back in its store for future use. In contrast, the presented techniques are more advanced and algorithmically superior to this third existing mechanism.

In particular, with the help of generative large language models (LLMs) a generative knowledge finder (KF) component of a VA, as supported by the presented techniques, may identify the flaws in a VA, extract the learnings from an HA's actions, and discover the most relevant set of KBs to incrementally train itself with contextual examples following an N-shot learning paradigm. In the case of unknown queries, the third existing mechanism is not directly able to find the right answer from its database while under the presented techniques a VA, along with a generative KF, may answer not only similar (including contextual) queries but will also provide contextually correct answers to less relevant topics based on its generative intelligence. It is important to note that the above is subject to a user's acceptance which, if denied, may result in a KF using auto-learning to enrich itself for future purposes.

To address the challenge that was described above, techniques are presented herein that support an intelligent, auto-learning, generative AI helper in a component-driven workflow (that is based on an N-shot learning paradigm) within a contact center ecosystem which can be leveraged by a VA to automatically improve its own competency level even if an organization's KBs are not current. Such an approach will help lessen HA involvement and promote VAs as the competent alternative even for complicated use cases.

Figure 1, below, presents elements of an exemplary solution that is possible according to the presented techniques and reflective of the above discussion.

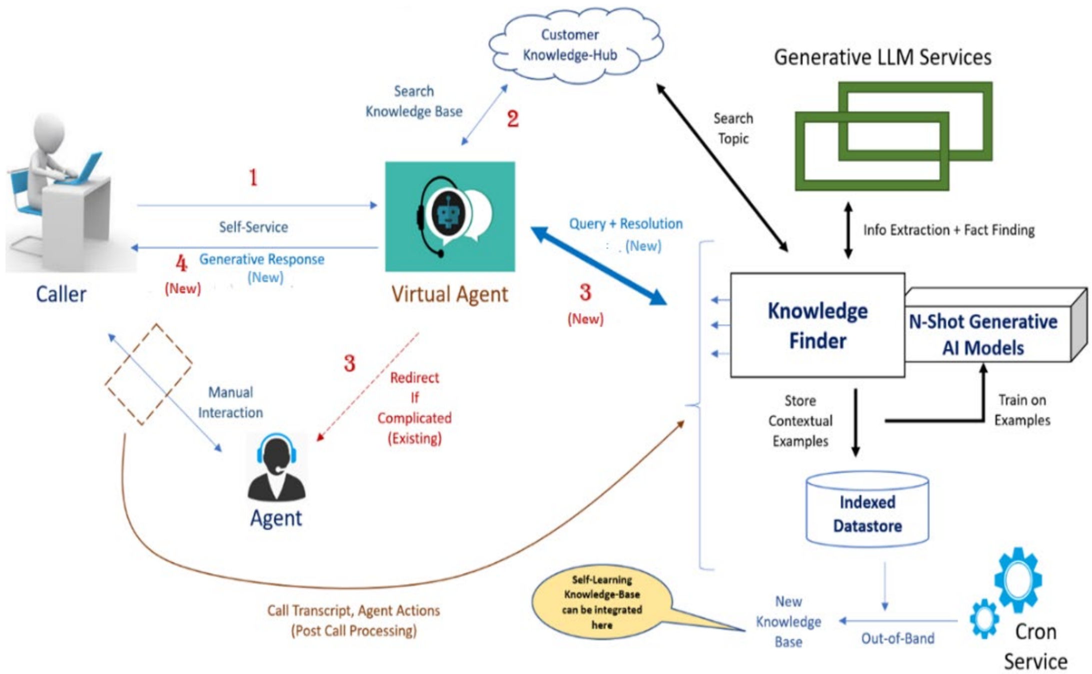

Figure 1: Exemplary Solution

As depicted in Figure 1, above, the presented techniques encompass an asynchronous proactive background learning flow. A new AI-based KF component may be introduced into a contact center to detect any flaws in a VA during a customer interaction and, through its generative intelligence, help the VA to rectify those identified shortcomings during future conversations.

A KF as described above may be backed up by suitable generative LLMs (which are efficient concerning information extraction and fact finding). Depending upon an LLM's capability, an appropriate set of LLMs may be selected to be integrated with a KF. Further, a KF may comprise three major components – a fact finder (which works with LLMs), a context generator (which also works with LLMs), and an answer generator (which may be backed up by a generative AI). Figure 2, below, visually depicts the different components that may reside within a KF.

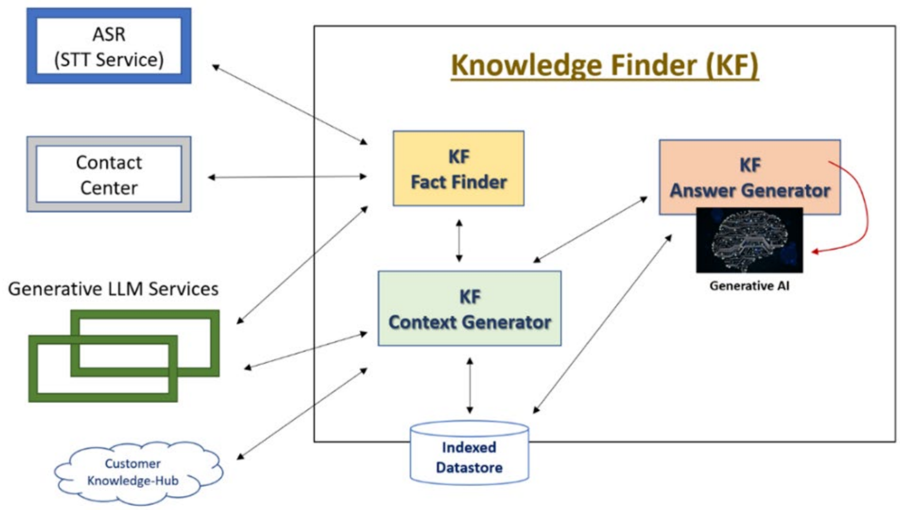

Figure 2: Knowledge Finder – Component View

For every case of HA involvement in a VA-based flow, the entire transcript of such an interaction may be examined by a KF's fact finder component, after a session is over, to discover the flaws or facts that led to the reason that the VA failed to handle the caller by itself. Such a fact finder component may also analyze an HA’s actions and extract the important learnings out of the same with the help of the LLMs.

A KF's context generator may search existing KBs and try to locate the relevant topics that are the closest match to the extracted knowledge and learnings. In the case where there is no matching KB topic, a KF may consider the HA's actions as the only source of knowledge. A context generator may, with the help of generative LLMs, also generate contextual examples based on new knowledge (such as an HA's actions) and the matched KBs (if any) and store the same in an indexed datastore. Further, a context generator may utilize such examples to train an answer generator (which, as indicated above, is a generative AI component within a KF) based on an N-shot algorithm, where N may represent a few or one, depending upon the circumstances.

Figure 3, below, presents elements of a process flow that captures various of the above-described activities.

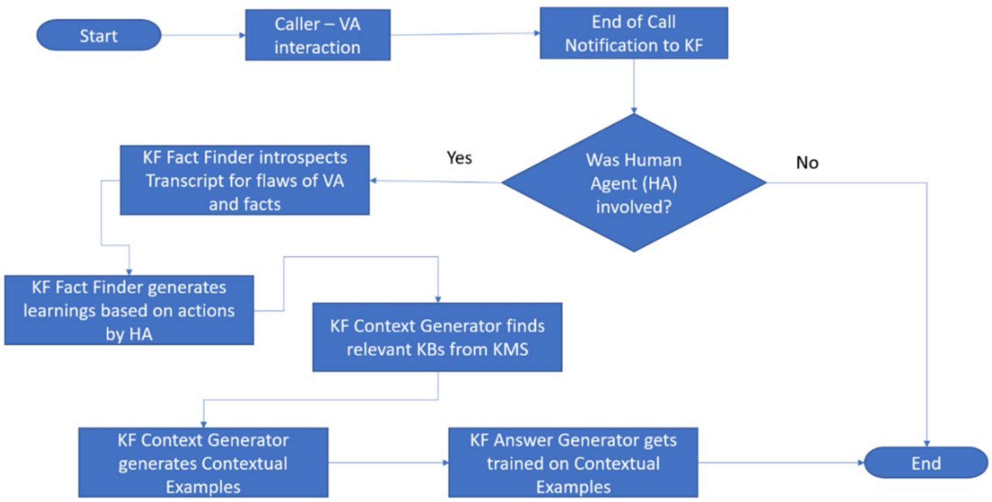

Figure 3: Process Flow – Background Learning

Through the above-described activities, a KF may continuously enhance its intelligence level as well as knowledge based on various contextual examples and store such examples in an indexed datastore so that it may aid a VA under a minimum reaction time approach.

In addition to the above-described elements, the techniques presented herein encompass a real-time VA-KF interaction flow. During any caller-VA interaction, if a VA cannot find a suitable resolution to the caller's query or problem from its own acquired knowledge or from an organization's KB, the VA may, in real time, contact a KF with the problem details.

A KF's answer generator may discover the best set of examples (e.g., a set containing X elements) from the indexed datastore and ‘prompt’ the internal generative AI component with those examples (as the helping context) to generate the appropriate answer to the query. The KF may then return the recommended resolution to a VA for its use provided that the answer generator's confidence score is above a predefined threshold.

The VA may then pass the same on to the caller and ask for their acceptance. If the caller accepts the resolution, then the VA may notify the KF and the KF may internally mark the same as a potential KB source (subject to verification for internal uniqueness) and store the results in an indexed datastore. If the caller rejects the resolution, or if the KF does not have a matching answer, then a post-call learning flow may be invoked to extract any embedded knowledge and enhance the KF's intelligence and knowledge level.

Figure 4, below, presents elements of a process flow that captures various of the above-described activities.

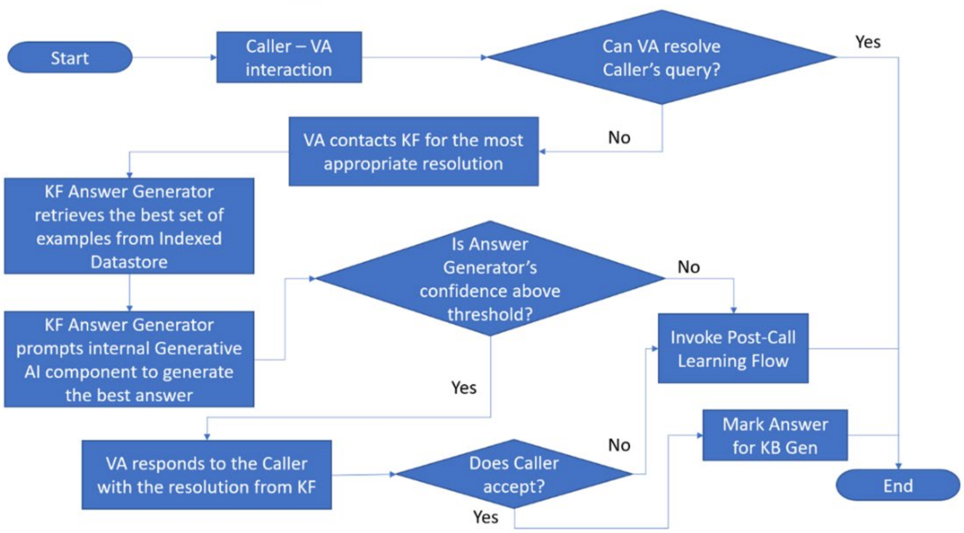

Figure 4: Process Flow – VA-KF Interaction

The techniques presented herein also encompass an asynchronous KB generation flow. Periodically, a subroutine may be executed (as, for example, a cron job) to process an indexed datastore's content and generate KBs that can be imported into a customer's existing KMS so that a VA may itself discover a best answer. Here, a SLKB may be leveraged to ingest the indexed datastore's content to enrich itself as to knowledge and capability, as noted above.

Any content which is not marked may be verified for contextual similarity against existing KMS entries. The content which is marked as a potential KB source may be verified for contextual similarity within the indexed datastore as well as against existing KMS entries.

Thus, the presented techniques will always ensure minimal HA involvement in a contact center workflow without compromising customer satisfaction.

As described and illustrated above, the techniques presented herein include a number of novel elements. A first element encompasses an intelligent, auto-learning generative KF in a component-driven VA workflow that supplements an existing VA to always ensure minimal HA involvement in a contact center.

Under a second element, without any manual intervention a KF component may enhance itself and, in turn, a VA's competency level beyond the VA's original intelligence (which may have been acquired through training and KMS access) with the help of a state of the art N-shot learning methodology that is powered by generative AI.

Under a third element, immediately following an unsatisfactory caller-VA interaction a KF may proactively engage its auto-learning mechanism (that is powered by generative AI) to ensure faster and more enhanced VA performance, thus guaranteeing a highly optimized future reaction time. A fourth element encompasses a KF's generative intelligence-driven answer prediction capability, even in an unknown context, and confidence score-driven suggestion mechanism that, together, contribute to maximizing customer satisfaction.

It is important to note that the techniques presented herein may supplement a SLKB, if the techniques are integrated with a SLKB, for updating a customer's KMS. In such a case, the SLKB may also be used to resolve complex issues in addition to often-repeated queries. It is also important to note that the presented techniques may be used in any kind of customer-VA interaction, including call and digital channels.

In summary, techniques have been presented herein that support an intelligent, proactive, auto-learning generative KF component which can be leveraged by a VA to automatically improve its competency level even if an organization's KBs are not current. Such a KF component may extract information from past caller-HA interactions and enrich itself with the help of an N-shot learning paradigm. Thus, without any manual intervention, a KF component, powered by generative AI, can enhance itself and, in turn, a VA's competency level beyond the VA's original intelligence (which may have been acquired through training and KMS access), exhibit a highly optimized reaction time, and dramatically lessen a HA's involvement.