Webex AI Quality Management

AI Transparency Technical Note

March 2026

webex by CISCO

Contents

At Cisco, we believe that Artificial Intelligence (AI) can be leveraged to power an inclusive future for all. We also recognize that by applying this technology, we have a responsibility to mitigate potential harm. That is why Cisco adheres to our Responsible AI Framework (the "Framework"), which is based on six principles of Transparency, Fairness, Accountability, Privacy, Security and Reliability. Cisco translates these principles into product development requirements, which ultimately form part of the product development lifecycle alongside our Security by Design, Privacy by Design, and Human Rights by Design processes.

Accordingly, Webex by Cisco (Webex) features that leverage AI are built with transparency, fairness, accountability, privacy, security, and reliability at their core. Each feature powered by AI undergoes an AI Impact (AII) Assessment — a best-in-class review of how the technical underpinnings of the functionality measure against the Framework Principles.

Webex AI Quality Management (QM) solution is built with the Framework Principles at the center of how we deliver AI-powered technology. This Technical Note describes more information about this feature and how Cisco responsibly leverages AI deliver the functionality.

Feature Overview

Webex AI Quality Management offers a native, intuitive feature set that seamlessly unified the evaluation and quality improvement of both human and AI agent interactions. By using a common framework to evaluate all interactions, customers will get consistent experiences for all interactions, AI or human, resulting in smoother and faster resolutions driving stronger loyalty and retention.

The AI QM feature set is designed to assist contact center supervisors and quality managers by providing AI-assisted evaluations of interactions, coaching, and sentiment analysis at scale. This feature set is available in Webex Contact Center supervisor desktop to users with a supervisor license and the appropriate Roles Based Access Control (RBAC) permissions as defined in Control Hub. The capabilities are tailored to the supervisor and quality manager persona. Throughout this document these roles may be referenced individually or collectively; however, all described functionality applies equally to both.

Key functionalities include:

- **Evaluation Forms:** Supervisors and Quality Managers can define evaluation criteria relevant to their business. Evaluation forms are comprised of a set of questions, typically ordered in sections, and assigned scores for each answer. Evaluation forms can be assigned to queues, teams, and individuals.
- **Automatic AI Evaluation:** Automatically evaluates interactions against criteria established by the supervisor. Evaluation scores are provided along with accompanying evidence and the ability for supervisors to adjust if necessary. Evaluation models are specific to customers and can never be used by other customers. Supervisors can calibrate these models to their own quality criteria and can view and override AI evaluation scores. The system will automatically retrain models to align with these supervisor-labeled scores.
- **Coaching:** Generates coaching guidance against criteria established by the supervisor.
- **Sentiment:** Evaluates customer sentiment on interactions. Webex AI QM automatically analyzes the context of a conversation based on transcripts and generates a sentiment score between -100 and 100. Supervisors can see the score and verify if the customer sentiment was deemed positive, neutral, or negative based on the score.

The AI QM feature set aims to improve contact center quality management by leveraging AI to give supervisors and quality managers visibility into the quality of all interactions, not just the limited set of interactions they have time to directly view.

Model Overview

Model Architecture

Webex AI QM uses models to evaluate interactions, evaluate sentiment, and summarize coaching tips.

The models used are fine-tuned BERT models, statistical classifiers, and third-party Large Language Models (LLM) from Microsoft Azure OpenAI Service. Reference the Microsoft Azure OpenAI Service Transparency Note for additional information on Azure OpenAI Service and how data is handled.

Often with AI, there is not a single model that performs best for all problems. AI QM dynamically assesses the accuracy of our models against the quality questions contained in evaluation forms. The best performing model is chosen for each question. Cisco models are fine-tuned to align with each customer's definition(s) of quality, using supervisor evaluations collected during a calibration phase and reinforced through routine human assessments. Cisco fine-tuned training models are specific to each organization and training data is never shared across organizations, nor does it leave the Webex Contact Center cloud.

Before models are fully trained, AI QM will default to LLMs provided by Microsoft's Azure OpenAI Service and uses generative AI models to summarize pre-calculated coaching tips. AI QM provides additional instructions, guardrails, and context to the Microsoft Azure OpenAI models to ensure a well-orchestrated experience for the end users. The feature currently uses the OpenAI GPT models offered by the Azure OpenAI Service, and based on research and benchmarking, these models provide the best experience for the use case in terms of cost, latency, and quality. Cisco is constantly evaluating available options, and in the event of any material changes the document will be updated accordingly.

Cisco employs telemetry to monitor LLM performance metrics and regional availability in real time. Cisco has established stringent data usage and privacy agreements with its LLM providers, and these agreements explicitly prohibit the providers from utilizing customer data for their own model, training, or improvement purposes. This helps ensure that all customer data processed through the LLMs is handled with the utmost confidentiality and is not stored or retained by the providers for any purpose beyond the immediate transaction.

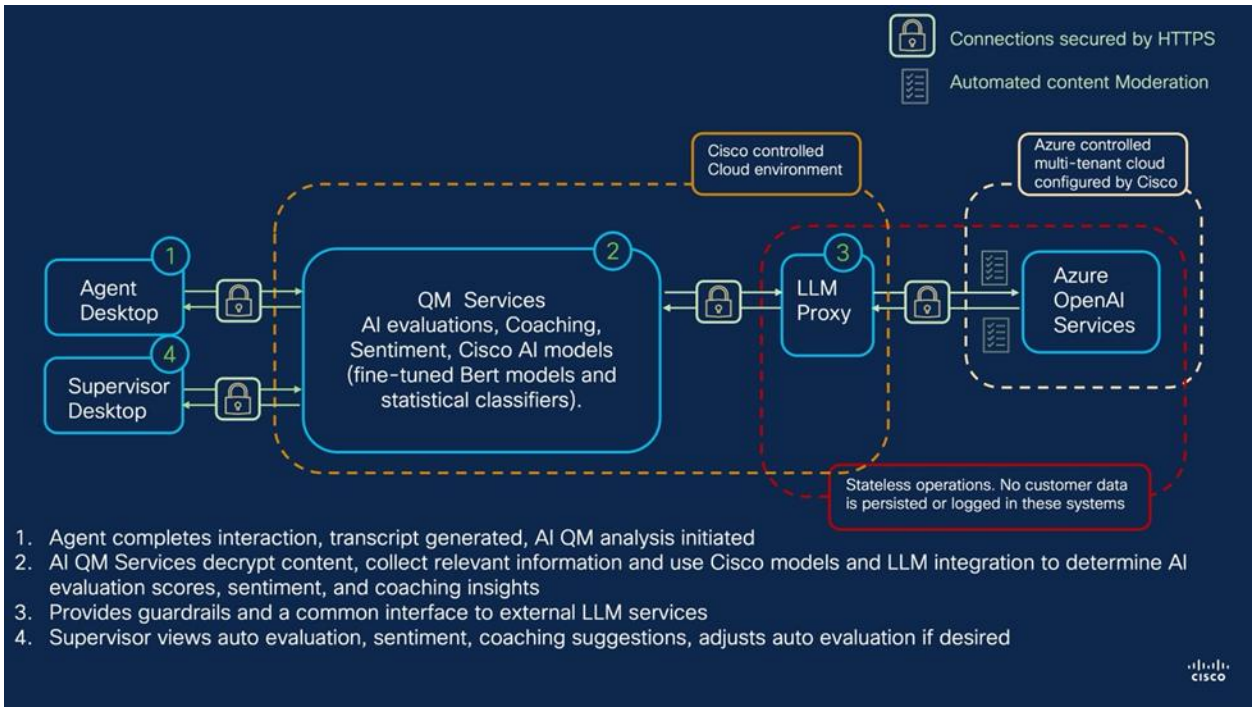

Figure 1: Webex AI Quality Management

Usage Guidelines

The AI QM feature is designed to support quality management of interactions by supervisors and quality managers within the Webex Contact Center environment. The following guidelines apply to the use of the feature:

- Intended Use: This feature is specifically designed for use within the Webex Contact Center supervisor desktop for quality management of human, digital, and AI customer interactions and is not intended for use outside of these defined channels.
- When a supervisor or quality manager views an AI evaluation, the system will provide a justification to explain why it scored a question the way it did. Supervisors or quality managers will be able to adjust these scores. Such overrides will be considered in the next training run so that over time the AI QM system will converge on the view of quality that is specific to their team and organization.
- Sentiment analysis is based on text transcripts, and audio is not used to derive sentiment. AI QM automatically analyzes the content of a conversation based on transcripts and generates a sentiment score between -100 and 100. Supervisors and quality managers see the score, and are told whether the interaction was positive, neutral, or negative. The information is used in combination with operational and quality data to determine the overall quality of an interaction.
- Evaluation questions, as defined in evaluation forms, are the responsibility of the supervisor or quality manager who created the question. Questions, especially those affecting agents, should be carefully verified by the supervisor or quality manager.
- Confidentiality and Ethics: All information presented by AI QM should be handled with strict confidentiality and ethical consideration and used solely to help improve the quality of the contact center business.
- Potential for Inaccuracies or Errors: Outputs generated by AI and machine learning technologies may contain inaccuracies or errors. Depending on how these technologies are used, such errors can lead to incorrect information or unintended actions. Users are strongly advised to validate critical outputs and take appropriate precautions when incorporating AI into their applications or decision-making processes.

Model Inputs

- **Evaluation questions:** A natural language question from the supervisor/quality manager, designed to measure an aspect of interaction quality they wish to measure.
- **Interactions transcript:** The ongoing conversation history, including customer utterances and agent responses
- **Model Weights:** Models are trained on synthetically generated data that is specific to an organization, as well as on the results of evaluations that supervisors/quality managers adjust. The resulting trained model weights influence the model outputs.
- **Metrics such as evaluation scores, call durations, and hold time, together with metrics targets:** Metric targets are defined by supervisor/quality managers in the supervisor desktop; where supervisors/quality managers can define lower and upper limits for each metric so they can have control over aspects such as “good” or “needs improvement” and what that looks like for their organizations.

Model Outputs

- **Generated response:** Evaluation score per question, where the system of scoring is defined by the supervisor/quality manager when creating a form; however, it will often be y/n or on a scale such as 1-3 or fair-ok-best. The supervisor/quality manager may accept or override the generated response.
- **AI Justification:** An explanation of why the AI system came up with the answer it did, together with references to places in the transcript that support the justification.
- **Sentiment:** An estimation of customer sentiment of the interaction with generated scores normalized between -100 and 100. Sentiment is deemed to be negative if score is between -100 and -45, neutral if score is between -45 and 45, and positive if score is between 45 and 100.
- **Coaching:** A summary of where metrics significantly exceed targets of fail significantly below targets.

Data Sources for Training and Evaluation

AI QM is trained using synthetic data that is generated specifically for each question and is specific to a supervisor's organization. AI QM is also trained on the actual evaluations that supervisors perform, and this training data is specific to a supervisor's organization and is never shared with other organizations. The resulting trained model is specific to a supervisor's organization and is never shared with other organizations.

Where questions are evaluated against Azure OpenAI models, Microsoft represents that it does not use customer data to improve Azure OpenAI models and does not store or retain Cisco customer data in the Azure infrastructure. Reference Microsoft's Data, privacy, and security for Azure OpenAI Service for more information about Microsoft's data privacy and security practices associated with the Azure OpenAI Service.

Model Evaluation and Performance

The system continuously monitors model performance and automatically improves accuracy through supervisor feedback. When supervisors correct or override AI-generated scores, that correction is captured as training data. Once enough corrections accumulate, the models automatically retrain and update without requiring manual intervention. This creates a feedback mechanism where models progressively align with an organization's quality standards.

Cisco frequently assesses and evaluates third-party models to maintain and/or improve both performance and accuracy of output. Humans are involved with review, testing, and quality assurance with the deployment of the underlying models.

Safety and Ethical Considerations

Cisco takes steps to mitigate safety risks and ensure ethical use. For example, third-party vendors, including Microsoft, undergo rigorous vendor reviews, which include security and safety assessments. Please contact Cisco if you have concerns or feedback about your experience with Webex AI Quality Management.

Fairness

For more information on how the third-party models underlying Webex AI QM are trained for fairness, please reference Microsoft Azure OpenAI Service Transparency Note. Large Language Models (LLMs) can reflect societal biases present in the training data. These biases, related to factors like race, gender, and religion can lead to unfair or discriminatory outputs.

Privacy and Security

Webex addresses processing and security around personal data, including retention periods, in its Offer Disclosures, found on the Cisco Trust Portal. Humans with the appropriate roles and permissions are involved in the review, testing, and quality assurance processes and may sample the inputs to and outputs from those models periodically.

Microsoft represents that it does not access, monitor, or store Cisco customer data. Regarding Cisco access to customer personal data, access rights are limited, and encryption is standard, as indicated in the relevant Offer Disclosures.

Updates and Maintenance

Cisco communicates ongoing significant product updates and improvements in user experience via updates to the product release documents, customer email notifications, and Help Center documentation to inform customers of product changes.

References

RAI Principles

RAI Framework

Cisco Trust Portal

Transparency Note for Azure OpenAI Service

Microsoft Data, Privacy, Security for Azure OpenAI Service