

CRESTA

Cresta Responsible AI Principles



Introduction

In today's rapidly evolving technological landscape, the implementation of generative artificial intelligence (AI) has become increasingly prevalent. While it's undeniable that generative AI has the potential to deliver transformative value, it can also introduce new risks and ethical concerns that can be daunting to navigate without a trusted technology partner. At Cresta, we recognize the importance of the use of responsible and ethical generative AI, particularly in the sensitive context of customer and prospect conversations. This document outlines Cresta's commitment to ensuring fairness, transparency, privacy, quality, and risk mitigation as we develop and deploy generative AI for the world's leading contact centers.



At Cresta, we are dedicated to pioneering a future where artificial intelligence not only drives transformative business outcomes but does so ethically and safely. Our commitment to responsible AI is formally recognized through our ISO 42001 certification, which defines the requirements for an Artificial Intelligence Management System (AIMS). This certification promotes effective governance through the Plan-Do-Check-Act methodology and addresses key AI challenges, including ethical considerations and transparency.

By adhering to ISO 42001 standards, we are empowered to manage risks while seizing opportunities in the ever-evolving AI landscape. This enhances trust for our customers and provides clear guidance for the responsible utilization of AI. Our dedication to trust, transparency, and ethical practices in AI development and deployment shapes a better, more human-centric future.

Fairness

Models. Cresta's base models are trained on large, diverse, and representative datasets and can be further customized using each individual customer's proprietary data. Our highly experienced AI delivery team is composed of individuals with diverse backgrounds and points of view, and our labeling processes consist of multiple levels of redundancy to mitigate the risk of individual bias during training tasks. We continuously monitor and evaluate each of our bespoke, task-specific contact center models for potential bias and have developed underlying systems designed to facilitate quick, corrective action if necessary.

Applications. At the application level, Cresta can help contact centers avoid bias and promote inclusivity. By helping reinforce context-specific best practices in real-time, facilitating the quick and accurate use of knowledge resources, and automating a variety of typing-related tasks like chat responses, note-taking, and interaction summaries, Cresta can empower agent populations with more diverse backgrounds and skill sets to be successful. Our agent-facing desktop application can further support the needs and preferences of different individual users by making the components and layout of their workspace customizable. Cresta's behavioral adherence and auto-QA solutions provide a consistent framework for assessing agent performance that isn't limited to a small sample of work or subject to whims of human evaluators, reducing the potential for bias that exists within manual performance management and quality evaluation programs.





Transparency

At Cresta, we believe that transparency and interpretability are foundational requirements for building trusted AI systems. We clearly document the purpose, capabilities, limitations, and potential risks of our AI systems as they continue to evolve, and are committed to making this information available to our customers as a matter of policy. Cresta also invests heavily in R&D designed to make it easier for humans to interpret and understand the outputs of our AI models, including cutting-edge techniques like [Chain-of-Thought Reasoning \(CoT\)](#) and [Model-based Critique](#). We've developed a variety of internal tools designed to observe and benchmark our models, and plan to continue devoting resources and roadmap to exposing more of these directly to customers in ways that can add value and increase trust.

Privacy & Ethics

Cresta is committed to adhering to strict ethical standards. Our generative AI models are designed to provide assistance to agents and managers that improve the efficiency and quality of customer and prospect conversations, but our platform does not make recommendations or predictions based on sensitive personal information, such as medical or financial data. Cresta does not use any protected explicit signals, such as metadata related to medical or financial attributes, in the process of training our models. We use industry-leading personally identifiable information (PII) redaction policies and algorithms as well as Payment Card Industry Data Security Standard (PCI-DSS) and ISO27701 compliance. In order to build responsible data governance, we also follow General Data Protection Regulation ([GDPR](#)) which is the most up-to-date policy regime for handling data-driven systems. These practices help ensure that our AI maintains a focus on enhancing agent performance and customer experience without compromising sensitive personal information. Please refer to [Cresta's Trust page](#) for more details on Cresta's security and data privacy programs.



Quality Optimization & Risk Mitigation

Cresta takes a comprehensive approach to maximizing quality and mitigating risk that spans our technology, people, and operating processes.

Data-driven Design. Cresta's AI models are developed using a two-pronged approach. First, we analyze the KPI performance data and behaviors of top agents to identify best practices and strategies that yield successful outcomes. This helps to create a north-star for our AI models, ensuring that they are designed to target optimal performance in each conversation. Second, we fine-tune our AI models using chat and call conversation data from those differentiated agents and behaviors. This enables our generative AI to quickly adapt to a customer's specific use cases and tailor its responses accordingly. By combining KPI data with conversation data, Cresta's AI models can provide valuable assistance to agents, supervisors, managers, and executive leaders - helping them enhance their productivity and deliver exceptional customer experiences.

Deployment in Stages. We partner with our customers to deploy generative AI solutions in stages, typically starting with transcription and post-conversation analysis that is used to thoughtfully design and implement real-time assistance and automation. Implementation of each core use case involves rigorous solution design, testing methodologies backed by our in-

house data science team, user acceptance testing, and a hypercare period to ensure quality and minimize risk. Once a Cresta solution is implemented, our customers can further refine the quality of their models and orchestrate them across the Cresta platform using an intuitive no-code platform called Cresta Opera. Cresta Opera provides robust backtesting and simulation capabilities designed to enable users to clearly understand how rules and workflows will behave before pushing them into production.

Human-in-the-loop Quality Assurance & Optimization.

We believe in the vital role of human expertise in ensuring the effectiveness and ethical use of AI. Our human-in-the-loop quality assurance measures involve continuous monitoring and evaluation of our AI models by experienced professionals. These experts oversee AI-generated assistance and provide feedback to fine-tune the models further, including implementing guardrails such as blocklisting potential responses. Cresta's underlying systems have also been designed to monitor usage trends from end users like agents and supervisors to identify deviations from generated responses, edits to summaries and notes, and other forms of implicit feedback that can be used to retrain models for better performance. Lastly, Cresta regularly collects explicit feedback from our customers through user surveys, onsite monitoring sessions, and regular performance read-outs to inform on-going optimization and alignment.



Generative AI Guardrails. Cresta puts extensive guardrails in place for all products and services that leverage Generative AI outputs, such as suggestions, smart compose, auto-summarization, and knowledge assist. Our LLMs are tuned to the context and use case of a particular enterprise domain. For knowledge-driven capabilities like Knowledge Assist and Generative Virtual Agents, we leverage metadata filtering and retrieval-augmented generation (RAG) that incorporates each customer's business-specific knowledge articles to ground generated responses to sources of truth. Cresta's AI pipeline can trace all knowledge chunks used in each generated response when evaluating and tuning our models, empowering our customers to better identify where their knowledge may be conflicting or require updates. Our domain-adaptive LLMs such as [Ocean](#) are further fine-tuned to reduce hallucination and improve steerability. We do extensive post-processing, leveraging techniques such as filter models, re-ranking and [self-critique](#), to double-check the outputs of the LLMs. Our stage-based deployment and continuous monitoring make sure humans always verify these outputs before they reach the customers.

Future-Proof Development. Cresta is committed to ensuring that our customers always have access to the latest breakthroughs in AI technology. We employ a modular approach to AI development, enabling individual components like integrations, transcription, LLMs, and the retrieval stack to be easily updated as new technology and techniques provide opportunities for improvement. Cresta's sophisticated approach to benchmarking and performance evaluation enables our customers to confidently deploy new state-of-the-art base models and rapidly adapt them to their customer domain.

CRESTA

Conclusion

At Cresta, we view the responsible use of artificial intelligence as both a moral imperative and a mission critical success factor for our business. As the underlying capabilities of artificial intelligence continue to advance, we are committed to ensuring that our research, policies, supporting technology, and operating models evolve as well. While the details of our approach may change over time, we will continuously evaluate them through the lens of fairness, transparency, privacy, quality and risk mitigation, serving as a trusted partner to our customers as they transform their contact centers with AI.