# The Role of Quality Training Data in AI-Enhanced CX

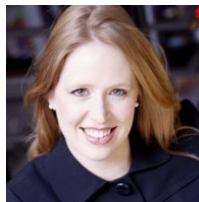WRITTEN BY: BROOKE LYNCH

This report is provided by TaskUs

# Featuring Insights From:

**Cedric Wagrez**

VP, Japan AI & ML Technology

Taskus

**Kris Crichton**

VP Client Services, Account Management and Operations

Visa

**Anthony Drummond**

VP Client Experience

Cox Automotive Inc

At its core, AI is intended to display human-like capabilities. It boasts reasoning, learning, planning and creative functions and creates a framework for technology to solve problems and achieve a specific goal.

It's these capabilities that make AI essential as we look to improve customer experiences and facilitate human-centered digital interactions. With AI-powered chatbots and virtual assistants, organizations can provide customers with instant, seamless and personalized support. However, to create a flawless system, companies must focus on training their AI technology with the most accurate, high-quality data possible.

Only then can their AI models continuously learn and improve to establish long-term success. Because, without effective, diverse and accurate data, this sophisticated AI technology will underperform and ultimately diminish experiences.

This report will uncover the importance of quality data in powering exceptional AI experiences. It will also touch on the leading challenges preventing companies from analyzing and interpreting data and provide solutions to deliver automated experiences that exceed customer expectations.

# The Challenge of Modern Data

Customer data has become critical; to properly tailor interactions and meet rising expectations, companies leverage the abundance of data they collect every second. By measuring items like purchasing behaviors, social media interactions and email communications, they can establish a quantified look into their customers' daily lives.

While it may seem like all of this data is seemingly collected and analyzed in real-time, prompting instant offers and targeted ads, not all companies are equipped to interpret intelligent insights from the data they are curating. According to CCW Digital research, more than a third of companies currently recognize the need to improve their customer data collection and analysis in 2021

To begin improving, companies must focus on unstructured data, which makes up 80-90% of all enterprise intelligence and is the most difficult to categorize. Unstructured data does not fit into a predefined model; because it incorporates a wide variety of formats like email, photos, video and text files, it is challenging to store and manage in a traditional database. Further, in the past, it often existed in siloed systems and was chronically underutilized, even being labelled 'dark data.'

But, because the majority of new data takes form as unstructured, we must work to prioritize it as a means of obtaining critical insights for sustained success.

Companies now have the opportunity to leverage sophisticated AI technology to effectively categorize and store these critical digital resources. Once properly assessed, this data can be used to identify key patterns and actionable insights to power better, faster and more personalized customer interactions.

"Bots, AI, whatever it takes in order to get fast responses, the right answers and access to data. So, if you don't have it in place today, you're going to see some downturn in regard to what customers are really expecting, because they're going to go somewhere else."

**- Kris Crichton, VP Client Services, Account Management and Operations, Visa**

## Quality Data Powers High Quality Interactions

When companies can properly analyze their data, they will come to recognize the value each data set holds and the level of quality they represent — because, at the end of the day, not all data should be treated equal. Further, low-quality data poses just as much risk as unused and under-analyzed data resources.

"You can have the best engineering team in the world, but if your model is trained on poor quality data, it will identify the wrong patterns, jump to the wrong conclusions, and not work as you or the end consumer intend it to. Low quality can lead to biased results, low accuracy and poor performance overall, making training data the most critical element of any machine learning strategy." **- Cedric Wagrez, VP, Japan AI & ML Technology, Taskus**

According to recent research, poor data quality costs organizations an average of $12.9 million annually due to increasingly complex data ecosystems leading to poor decision-making. Beyond poor decisions, low-quality data impairs AI technology. Quality data is ultimately the foundation for building Artificial Intelligence and Machine Learning tools. Without proper, high quality data, sophisticated AI models are trained to identify inaccurate patterns, leading to diminished experiences and biased results.

CCW Digital research confirms that currently, digital experiences are not overwhelmingly improving — only 16% of customers stated that digital experiences have actually gotten better. Therefore, we must work to power AI experiences with the highest quality data if we want to continuously enhance interactions over the long-term.

To establish exceptional digital self-service options, organizations must train their datasets to encompass an extensive range of user queries and historical interaction data to ensure accuracy and inform strategic intents. Additionally, companies must ensure that their data is diverse — including a range of user demographics, geographic locations and languages to understand an accurate picture of all interactions and uncover deeper, well-rounded insights. Then, organizations can optimally develop their AI-powered self-service technology to reflect the nuanced, complex and personal interactions customers are having each day.



"Companies that marry data to experience intention are going to far outpace their rivals in driving customer loyalty behaviors. When we can capture customer sentiment data, customer behavioral data and related operational metrics at a level that allows for a data overlay on a depiction of the customer journey, we put ourselves in position to understand customer needs at a very tangible level and to establish a baseline of our performance at delivering appropriate experiences." **- Anthony Drummond, VP Client Experience, Cox Automotive Inc**

# AI Operations for Optimized CX
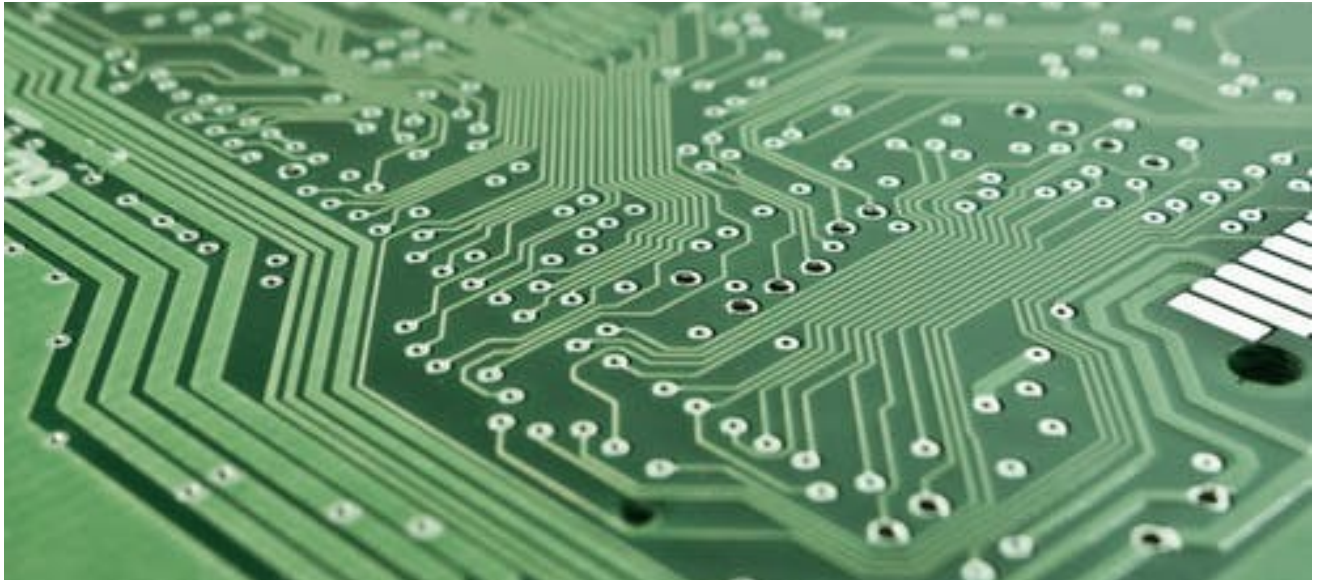


## Data Labelling for Accuracy

To optimize AI operations, organizations must prioritize both the quality and diversity of their data to see long-term improvements. But, before they can actually improve their data quality, they must interpret, analyze and store the exponential amount of data they are currently collecting.

To assure the success of their training data, companies must prioritize data labeling. By annotating complex unstructured data, like images, video and text, organizations can ascribe value to each piece of data and provide context to inform deeper AI learning. For example, data labelling can help identify pieces of text for sentiment analysis and natural language processing to power chatbots or voice assistants. It can also be used to monitor social media to identify patterns and customer sentiment trends.

With an accurately labelled dataset, companies will establish a foundation to further train and improve their AI technology. Once put in place, companies will secure an efficient and accurate training model to power better, more relevant customer interactions.



"These days, companies are collecting more data than ever before. However, according to multiple analyst estimates, 80-90% of enterprise data is unstructured…This is where large scale human annotation comes into play. Most machine learning models are trained using supervised learning, which relies on a large foundation of data tagged and validated by humans. However, due to the sheer scale and complexity of data required to train algorithms, many companies choose to outsource data annotation to third parties that already have the human capital and process expertise." **- Cedric Wagrez, VP, Japan AI & ML Technology, Taskus**

However, to establish a successful AI model, humans must guide the technology and annotate each dataset for optimal accuracy. Experts can identify areas of improvement and retrain the technology for continued development. A human-centered feedback mechanism gives companies the ability to increase predictability and confidently support customers.

## AI-Powered 'Human Approved' Chatbots

Once companies can successfully organize and interpret their data, they can facilitate better self-service interactions. With trained datasets, chatbots can take on a more human approach and provide relevant and intuitive interactions.

CCW Digital research found that 51% of companies are now focused on elevating their quality standard in digital channels to improve customer satisfaction and trust — characterizing the advancement of chatbot technology as critical for enhancing digital experiences.

"High-performing AI chatbots are built with the end user in mind. Training datasets should cover a wide range of user queries, and are ideally developed using historical conversation logs and plenty of user feedback. By doing so, you'll be able to extract valuable insights into queries made by your users, which will further inform strategic intents for your chatbot." **- Cedric Wagrez, VP, Japan AI & ML Technology, Taskus**

Through collecting and analyzing speech data from text and audio, organizations can better identify commands and seamlessly communicate with users.

With Natural Language Processing, chatbots can be trained to understand varying languages and utterances, and identify sentiment even in environments with heavy noise. Through proper tagging and transcription, companies can bolster their self-service offerings to support customers effectively in every context.

Once organizations are finally able to comprehend their customer's needs, they can prioritize feedback to further enhance experiences. Human-in-the-loop mechanisms ensure that humans can quickly step in to support the chatbot once required but continue to focus on the most complex and meaningful cases.

"Offering blended customer support can also help balance chatbot technology with human empathy. Many chatbots are able to recognize the need for human intervention, transferring customers to an agent if a human-in-the-loop is needed. This workflow increases overall productivity, and leaves agents to focus on complex cases that require complex human interaction." **- Cedric Wagrez, VP, Japan AI & ML Technology, Taskus**

## Content Moderation for Digital Success

With effective data management and AI technology, organizations can also enhance the safety and relevance of their online environments. On a daily basis, individuals engage and access a wide variety of content. Through social media alone, users share 95 million photos and videos on Instagram and connect with 2.8 billion daily active users on Facebook. While this outpouring of content is often a positive reminder of our greater global community, not all content is fit for every user. This places greater responsibility on brands to protect customers from potentially harmful subject matter online.

"With the ever-increasing amount of digital content, comes the need for continuous monitoring to safeguard online spaces. AI-powered content moderation models work hand-in-hand with human moderation teams to scale and optimize much of this process." -

**Cedric Wagrez, VP, Japan AI & ML Technology, Taskus**

Quality training data allows companies to improve search relevance by training AI technology with categorized search queries to distinguish between unique searches. Rather than provide content based on frequency of keywords, companies can empower search relevance to identify the most accurate results. Organizations can then further tailor experiences by evaluating ads, ensuring that customers are only receiving acceptable and useful advertisements.

AI also empowers companies to optimize content and detect sensitive discussions online. Content moderation services give companies the tools to instantly identify harmful contents — with effective training, AI can recognize text, images and videos for hate speech, fraudulent news or spam.

This is all made possible by human evaluators; once they assess each search result for relevancy, they are able to feed the search engine's algorithm to move the most applicable results to the top of the page. With precise labelling and categorization, and continuous feedback from expert evaluators, companies can ensure that their customers are receiving safe and relevant recommendations and optimizing digital experiences overall.

"AI doesn't eliminate the need for human moderators in the loop—in many cases, humans still need to actively monitor output for accuracy as well as handle the more contextual, nuanced content concerns. As human moderators label content as harmful or not, they feed freshly labeled training data back to the algorithm to continually improve its accuracy." **- Cedric Wagrez, VP, Japan AI & ML Technology, Taskus**

# About the Author

**Brooke Lynch, Digital Writer & Analyst, Customer Management Practice**

Brooke Lynch is an analyst and staff writer for CCW Digital. With a background in television news and production, she's worked across industries covering B2B marketing, procurement and finance events.

Her current work highlights challenges and opportunities for customer experience and contact center leaders, with a recent focus on e-commerce, retail, and technology.

# Meet        Our Analysts
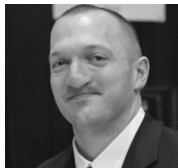
**Brian Cantor**

Principal Analyst &
CCW Digital Director

**Michael DeJager**

Principal Analyst & Divisional Director
Experience Design Series

**Max Ribitzky**

Principal Analyst,
Go-to-market Research

**Nadia Chaity**

Senior Analyst, CCO Series

**Sandy Ko**

Senior Analyst

**Amanda Caparelli**

Analyst

# Get Involved

**Ben McClymont**

Business Development Director

**E:** Ben.McClymont@customermanagementpractice.com

**P:** 212 885 2662
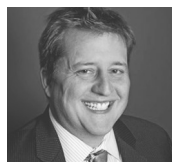
**Simon Copcutt**

Head of Strategic Accounts

**E:** Simon.Copcutt@customermanagementpractice.com

**P:** 212 885 2771