

FLIP AI-CSAT: Predicting Customer Satisfaction

I. WHAT IS CSAT?

CUSTOMER Satisfaction (CSAT) surveys have long been the industry standard for measuring customer satisfaction. They typically prompt the customer to score their experience along a 5-point¹ Likert Scale, with 0 meaning highly dissatisfied, and 5 meaning highly satisfied. These CSAT scores can be analyzed per-customer or aggregated across various scopes (such as time, intent, or channel), and — with further investigation — can indicate a variety of important factors such as brand loyalty, repurchase, customer retention, and churn.

II. CSAT DISADVANTAGES

Whilst CSAT scores are a crucial tool that can quickly identify areas for improvement, they tend to suffer from several pitfalls.

Insufficient Sample Size. Typically, only 5%-10% of all surveyed customers actually submit a response. This can lead to results that do not have sufficient statistical significance, and may be interpreted incorrectly. Moreover, any customers who have a negative experience and don't respond to the survey will not be identifiable within a timely manner, potentially leading to missed CX improvement opportunities.

Skewed Distribution. Customers who do take the time to respond are far more likely to express strong negative (or, to a lesser extent, positive) experiences, thereby skewing the CSAT scores — resulting in a respondent population that is unlikely to match the overall population.

Respondant Bias. Customers can have a range of socio-demographic and cultural biases. This can result in varying degrees of subjectivity; for instance, a (hypothetically) *identical* interaction could elicit different CSAT scores from different customers.

Unidimensionality. CSAT scores are solely derived from responses to the single question "How satisfied are you?". As such, they are unable to capture the plethora of attributes that define a customer experience (such as sentiment, speed, clarity, effort, and resolution).

Agent Partiality. For conversations with a human customer service agent, a customer might respond with a specific CSAT score based entirely on whether they liked the agent or not, irrespective of whether or not their intent was satisfactorily resolved.

Costly and Time-Consuming. Sending CSAT surveys, waiting for responses, and interpreting the results can be a laborious and expensive process, requiring additional resources with separate workflows.

In order to overcome these drawbacks, Flip has created **AI-CSAT**: an AI-based CSAT score for every customer interaction.

III. AI-CSAT

AI-CSAT is a score between 1.0 and 5.0 that represents the satisfaction of a customer during a customer service phone call. It is generated from the call transcript using a Large Language Model (LLM), and evaluates the conversation across three core dimensions: Speed of Resolution, Completeness of Resolution, and Emotional Polarity of the Customer. No survey is required, and results are available in real-time. AI-CSAT is calculated for each call as follows:

- 1) The call ends, and the call audio is sent to Flip's transcription service, based on Faster Whisper's AI model. The call audio has separate input and output channels (recorded on the left and right channel respectively), and each channel is transcribed individually. Thus, each call transcript is the result of *two* transcriptions, and a re-segmentation algorithm combines the two transcriptions together, creating a cohesive transcript with

appropriately-defined turns. Hold time and agent pickup time are determined with some natural language processing, and timestamps are formatted.

- 2) The formatted transcript is embedded into a prompt that is submitted to OpenAI's LLM, currently GPT-3.5 Turbo. The LLM generates a score from 1-5 with an accompanying description for each of the core dimensions (see examples below), in addition to several other metrics such as the agent name, save opportunities, successful sales, and marketing attempts.
- 3) The final AI-CSAT score is the average of these three dimensional scores, with the Emotional Polarity score weighted slightly with a *barge-in* score that represents how often the caller spoke over the agent during the call (which is able to be calculated as a result of the dual-channel transcription).

A score representation of 1-5 was chosen instead of a percentage as it enables a more intuitive comparison with existing CSAT scores, and, with the inclusion of explanatory summaries, allows for a simple interpretation. Here is an example of a high AI-CSAT score:

Speed of resolution: 4.5

The representative quickly identified the reason for the call and efficiently resolved the issue by updating the delivery date.

Completeness of resolution: 5

The representative fully addressed the customer's request by successfully changing the delivery date to Tuesday.

Emotional polarity of caller: 5

The caller expressed gratitude and satisfaction with the resolution, indicating a positive emotional experience.

Each dimension that comprises an AI-CSAT score has an easy-to-understand summary, clearly explaining the reasoning for each dimension's calculated score. Here is another example, this time illustrating a low CSAT score:

Speed of resolution: 1.5

The call took a long time to resolve the customer's issue. The customer had to wait multiple times and was not able to get a direct phone number for the location they needed to contact.

Completeness of resolution: 1.5

The customer's issue was not fully resolved. They were not able to get a direct phone number for the location and were told to wait for a call back. The customer expressed frustration and dissatisfaction with the service.

Emotional polarity of caller: 2.5

The customer started off with a neutral tone but became increasingly frustrated and upset as the call progressed. They expressed their dissatisfaction with the service and the lack of assistance provided.

Each score is based on the customer's intent on the call, and the subsequent handling of that intent; in other words, AI-CSAT is contextual, and doesn't rely on learnt conceptions of — or sentiment towards — specific phrases, actions, and intents. AI-CSAT is entirely objective, and is *immediately* and *consistently* applied to 100% of customer calls, eliminating the need to extrapolate sparsely-collected CSAT survey results, removing the human bias element, and illuminating areas that might otherwise have stayed hidden from analysis.

IV. FUTURE WORK

Flip AI-CSAT has been released in BETA, and is included at no extra charge as part of the Flip platform, now live for all customers. Over the next couple of months we will be improving its algorithms and models for increased accuracy, in addition to exploring further AI-CSAT dimensions that may offer increased nuance in reasoning. We will also be benchmarking Flip AI-CSAT with traditional CSAT responses across our client-base. To see your AI-CSAT score, along with additional new metrics, visit the Analytics section in the Portal².

To whom correspondence should be addressed: dev@flipcx.com.

Manuscript creation January 2024.

All information copyright © 2015-2024 Flip CX, Inc. Do not disseminate.

¹Sometimes, a 7-point or 10-point scale is used.

²<https://portal.flipcx.com/analytics/performance>.