frame.ai

CSAT Under the Microscope

Introduction

Most CX organizations share a primary goal to improve customer satisfaction, drive up retention and growth of customer relationships, and ultimately reduce service costs. However, while "satisfaction" is an intuitive concept, reliably measuring satisfaction to inform CX strategy is deceptively challenging.

The most commonly adopted tool for measuring satisfaction is the Customer Satisfaction ("CSAT") survey. CSAT surveys are delivered following specific types of user interactions such as a chat or email, and typically ask recipients to rate the support they received as "good" or "bad" or choose a number on a numerical scale. Advanced CSAT programs may target surveys at specific types of interactions or moments in the customer journey and may use framing language to better target results — but the overall construct is a voluntary response to a specific interaction.

Frame AI has developed another approach designed to meet our goal of measuring satisfaction anywhere without requiring additional customer effort. Frame AI's sentiment engine labels specific sentence fragments as Wins, Issues, and Risks (see Table (1)) across the customer voice, in any channel, and in real-time. Frame AI applies labels to every interaction, giving us a lens to evaluate the impact of sampling bias across CSAT survey results.

Table (1) Frame AI's Definition of Wins, Issues, and Risks

Frame AI Sentiment Label	Definition	Examples
 Wins	• Non-perfunctory expressions of praise and appreciation • Evidence of a strengthened relationship • Demonstrates exceptional service, product experience, etc.	• ***"Honestly love your service..."*** • ***"I appreciate the extra attention to our sites on Saturday."*** • ***"Thank you so much for going the extra mile with this."***
 Issues	• Problem statements that describe the motivation for the communication • Summarizes issues described in exchange • Demonstrates relevant process and product feedback	• ***I'm an electronic resources librarian and am experiencing connection issues with some of my databases."*** • ***"I reported my card lost on 10/9 and STILL don't have my replacement card."*** • ***"Again, excellent product but your doc could use some love..."***
 Risks	• Demonstrates frustration with the service, product, business • Evidence of weakened relationship • Demonstrates that major problems requiring intervention have occurred	• ***"Guys... this is unacceptable."*** • ***"The number of issues that affect us and that your engineering apparently isn't planning to fix keeps growing."*** • ***"This is horrible user experience in that the UI tells you you're doing everything."***

To explore how CSAT surveys and our sentiment engine complement each other, Frame AI conducted a study on the overlap between CSAT ratings and our automatically labeled sentiment annotations across millions of support conversations. Our analysis included all conversations with survey recipients, including non-respondents, and any free-form comments on the surveys themselves.

Table (2) summarizes the overall relationship between our detected sentiment and CSAT scores across the dataset, further confirming that both methods detect a common notion of satisfaction but also feature important differences. We observed high rates of cross-pollination, with risks featuring prominently in conversations with good CSAT ratings. Conversely, conversations with bad CSAT ratings also featured wins. However, our study's most compelling finding is the breadth and depth of high-fidelity CX feedback we surfaced across interactions that received no rating at all.

Table (2) Distribution of CSAT Ratings for Wins, Issues, and Risks

	Wins	Issues	Risks
Bad CSAT	2.5%	5.1%	21.8%
Good CSAT	25.4%	22.7%	16.4%
No Response	72.1%	72.2%	61.8%

In the following sections, we use data on these discrepancies to highlight some key themes regarding where CSAT succeeds and fails at accurately measuring the customer experience.

CSAT Surveys Reflect Many Types of Feedback

While CSAT surveys were originally intended to measure agent performance, they easily become feedback catch-alls — no matter the phrasing of survey questions, many customers inevitably feature other aspects of their experience in their rating. For example, our study showed that the risks we captured in interactions with bad CSAT ratings run the gamut from poor service quality to frustration with product performance or company policy.

Similarly, interactions that received good CSAT ratings covered a broad range of subjects and complexity. While both a simple password reset assist and a days-long troubleshooting effort may have received a good CSAT rating, the two scenarios are likely to have vastly different impacts on operational costs and value-added to the customer. A third positive rating for getting the desired answer on a billing policy might have nothing to do with the service but with the product or policy being queried.

To evaluate the factors driving CSAT scores, we evaluated[1] 500 "Win" free response comments in positive CSAT reviews and 500 "Risk" free responses by relation to service, product, or business issues, such as billing policies. We found that comments explicitly focused on service comprised less than 35% of our sample. Furthermore, product appreciation represented a nearly equal number of positive scores, and billing frustration represented a disproportionate number of negative scores.

Table (3) Distribution of Causes for Wins and Risks

Frame AI Sentiment Label	Service-related	Product-related	Business-related
Win	33.9%	39.2%	26.9%
Risk	23.7%	31.9%	44.4%

Most service teams will not be surprised by the fact that non-service factors complicate CSAT surveys — however, the extent of variation is increasingly problematic for organizations tracking simple averages on a large set of surveys.

[1] *Frame AI's FSAT module, introduced later in this report, provides automatic categorization of drivers for positive and negative statements. We hand-verified these to confirm a ground truth for this comparative analysis.*

Good CSAT Ratings Can Mask Serious CX Flaws

In our study, 85% of customers who responded to the CSAT survey gave a positive rating. Industry sources like Customer Thermometer and Zendesk often cite figures in a similar range.

In many organizations, good CSAT is a means unto itself rather than a means to an end. CSAT is often used to evaluate and compensate Support personnel, from frontline agents to leadership. High CSAT scores are touted in internal QBRs and on company websites. Unfortunately, this means that many organizations accept good CSAT ratings at face value.

Forrester says that 60% of companies don't regularly track operational data from interactions to help explain why customers felt the way they reported. This approach is dangerously reductionist in today's competitive CX environment. Our analysis revealed a striking proportion of negative feedback underneath good CSAT ratings — among rated conversations, 43% of negative feedback came from conversations with good CSAT ratings.

A purist interpretation of this co-occurrence would suggest high agent performance across consequential tickets, but stopping the analysis there can prove risky. Even though agents may have adequately resolved issues, accepting good CSAT ratings as an indication of good CX can let many potentially serious CX flaws go unchecked.

Let's review some examples:

Table (4) Risks in Interactions with Good CSAT Ratings

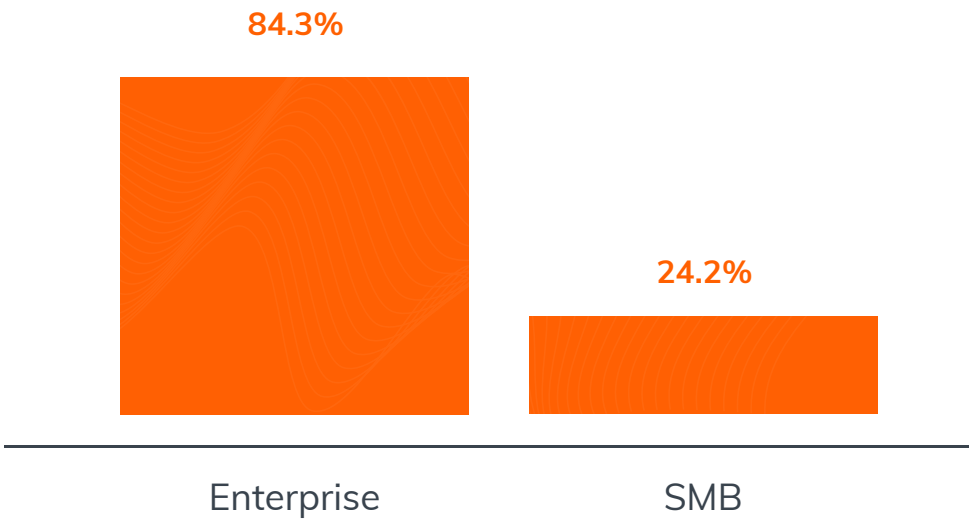
Service-related Risks	Product-related Risks	Business-related Risks
"I'm disappointed with the support we've gotten on this issue so far, please do not close it out yet."	*"This shouldn't take 11 seconds to load, and it's consistently taking 11 seconds to load."*	*"I was told I would never have this issue with your system."*
"I've attempted to export via your API, but you said before that this particular bug won't allow me to do that, and I can't see how to export any other way on your docs."	*"This might be documented behavior, but I don't think it's acceptable in a world-class product."*	*"This is nothing short of a rip-off."*
"This is so hard only getting one email per day to resolve this."	*"I have now sent 3 or 4 proposals and my customers are not receiving them."*	*"The only thing I don't like about ***** is that I get sent numerous requests to answer the same survey each time I contact *****. I'm going to have to mark the emails as spam if you cannot turn this off. Stop."*

Because the interactions in Table (4) received a good CSAT rating, the Product team may not be aware of a latency issue, and the Sales team may not be aware of incorrect expectation setting.

Similarly, by failing to look beyond good CSAT ratings, Support teams can miss valuable opportunities to improve service, from coaching agents to tightening processes. A conversation that ended well enough to deter the customer from leaving a bad rating is not necessarily indicative of a great customer experience. Several excerpts from interactions with good CSAT ratings reveal that the resolution process leaves a lot to be desired.

Relationships that face multiple stakeholders at each customer (“enterprise relationships” in this paper) pose unique challenges in using CSAT to diagnose improvement areas. When we divided our data between enterprise and SMB relationships, we found that the prevalence of risks in good CSAT interactions was 3.4x more pronounced in enterprise relationships. For SMB relationships, an average of 24% of negative feedback came from conversations with good CSAT ratings, and for enterprise relationships, that number skyrocketed to 84%.

Figure (1) Risk Prevalence With Good CSAT Across Enterprise and SMB



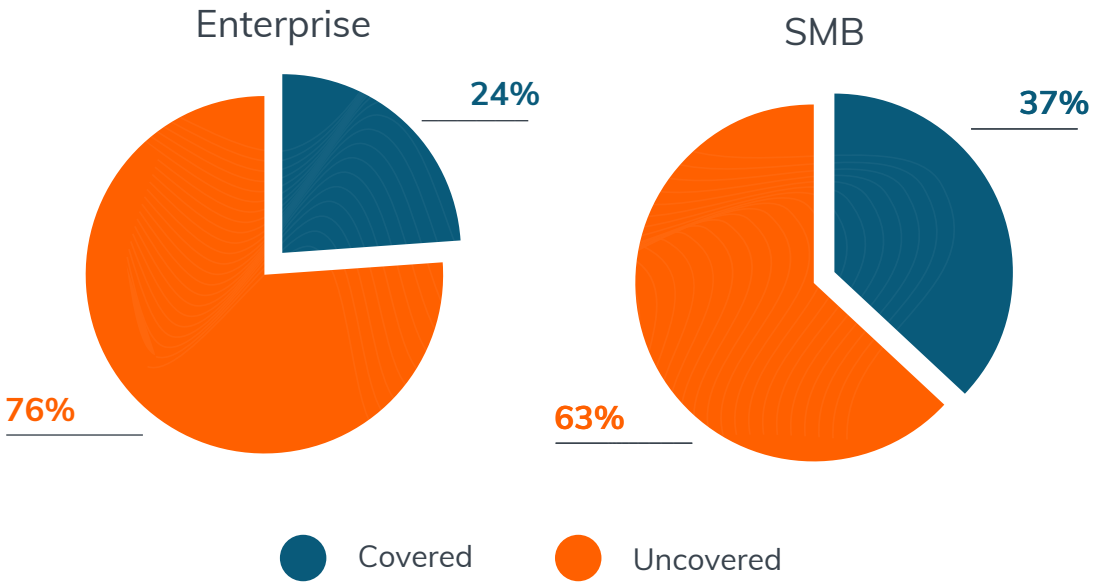
One reason for the disparity is that because enterprise relationships feature more stakeholders, there are more potential frustration points across the customer journey, even if individual interactions receive good ratings. Teams supporting enterprise relationships may also be more inclined to experiment with survey distribution considerations such as recipients and timing. For example, a delay between ticket close and survey distribution may promote better ratings if the customer has sufficient time to cool-off from a heated exchange.

CSAT’s familiarity makes it an attractive choice for reporting, but it underserves today’s CX leaders in recognizing specific issues and prioritizing actions. Rich textual analysis can help isolate different reasons for CSAT ratings, helping to target and track efforts that meaningfully improve the customer experience.

Bad CSAT Ratings Fail to Capture the Majority of CX Risks

Many organizations rely on bad CSAT ratings to alert them to situations that require attention. However, our study showed that interactions with bad CSAT ratings captured an average of less than 30% of risks. This discrepancy indicates that 70% of the relationship-weakening frustrations that customers are motivated to express easily accumulate under the radar.

Figure (2) Risk Coverage of Bad CSAT Across Enterprise and SMB



As we noted earlier, the complexity of enterprise relationships can make CSAT ratings even harder to interpret and more dangerous to accept as an indicator of CX quality.

While most companies have a process for reviewing conversations with bad CSAT ratings, they are difficult to interpret at any meaningful scale. We noted earlier that customers often fail to decouple agent performance from other parts of their experience. Despite good agent performance, customers often give bad CSAT ratings, even when they react to factors outside agents' control. Many organizations struggle to attribute CSAT ratings to discrete aspects of customer experience, and as a result, bad CSAT ratings can lead to resource-draining internal finger-pointing.

Table (5) highlights some examples of risks we captured from conversations with bad CSAT ratings that feature disparate potential motivations:

Table (5) Risks in Interactions with Bad CSAT Ratings

Service-related Risks	Product-related Risks	Business-related Risks
"The 'I'm sorry I'm not trying to give you the runaround' answer is simply bogus apology."	*"This has cost me time that I don't have, and I have had to foot the bill for your app's continuous fails and setbacks."*	*And now, even after paying for my current subscription, I'm bombarded with ads to upgrade to a newer version of Pro."*
"And deflecting blame on the vendor of your choosing is not the best customer service practice."	*"There is nothing you guys offer to protect me from people scamming me."*	*"This company continues to be ridiculously obtuse to the needs of those it claims to be providing a service to."*
"I understand she's not a developer, but she did not offer any additional insight or workaround option for an extremely detrimental issue to our operations."	*"A hung screen is unacceptable -- no matter what the reason behind it."*	*"It would be a real shame for you to lose a paying customer because of your poor company policies."*

Relying on CSAT ratings alone, many organizations cannot differentiate between the scenarios reflected above without spending a great deal of time and resources.

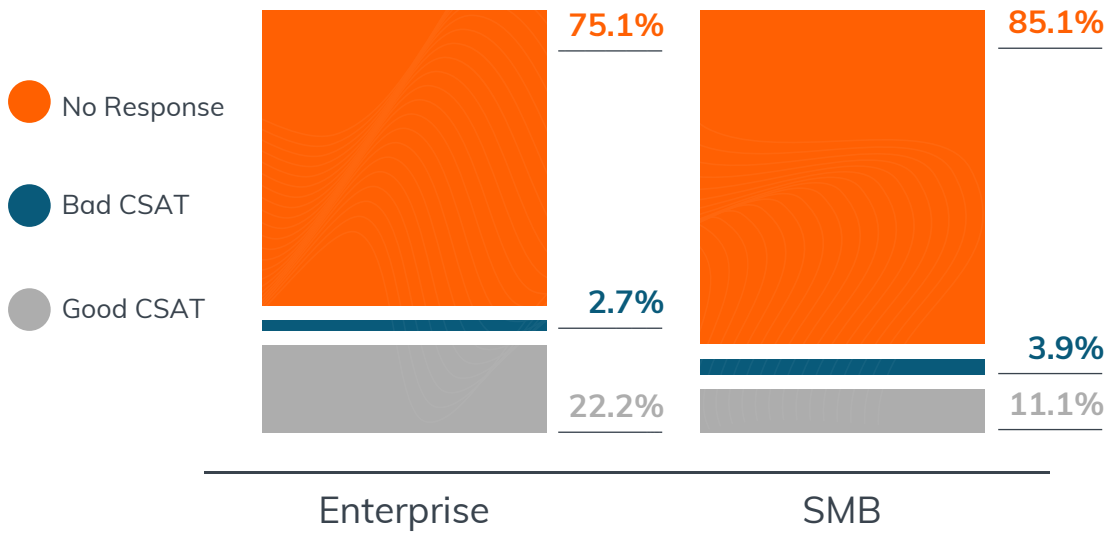
The impact of the loss aversion principle further magnifies the interpretability challenge. The loss aversion principle holds that losses loom larger than gains — in a CX context, bad CSAT ratings garner the most attention. As a result, especially given the underlying data sparsity, many organizations are prone to over-indexing on problems that catch their attention via a bad CSAT rating. However, the awareness that bad CSAT ratings draw infrequently translates to action, as most companies lack an efficient way to determine whether problems raised affect a significant customer contingent. In fact, according to an Oracle study, 79% of customers who shared negative feedback about their experience online had their complaints ignored. Similarly, according to a 2017 survey by CCMC, 79% of customers who complained were dissatisfied with how companies responded.

Failure to identify specific CX drivers and their impact inhibits organizations' ability to address them appropriately and thus improve CX.

The Silent Majority Holds Strong Opinions

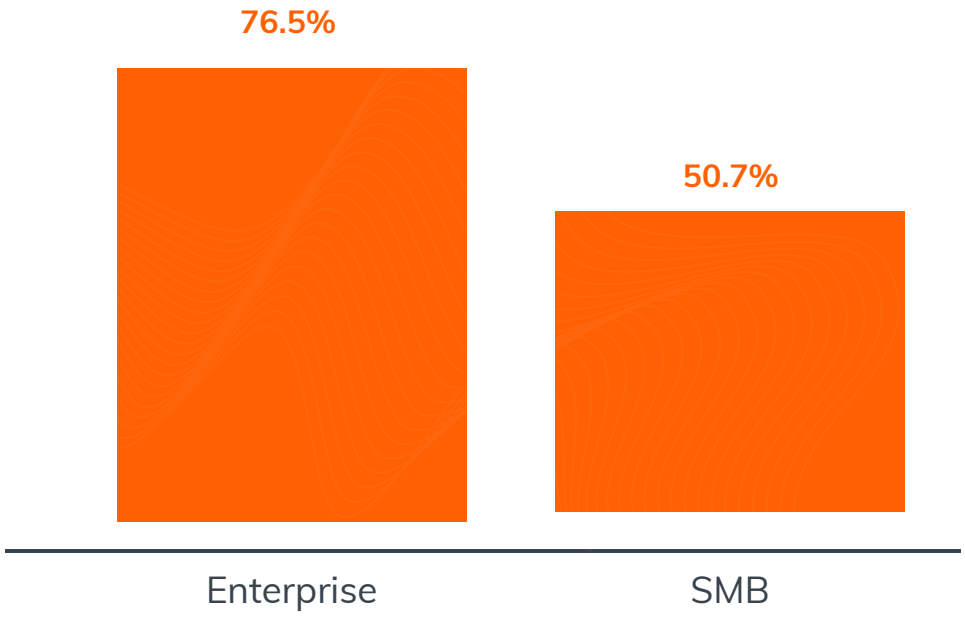
Earlier, we noted that our study found that 85% of customers who responded to a CSAT survey gave a positive rating. On the surface, this looks great, and many CX leaders would proudly tout that statistic. However, our study also found that across enterprise and SMB relationships, 75% and 85% of customers, respectively, did not respond to CSAT surveys. When we consider the non-responsive majority, the group of customers who received a survey and responded with a good rating plummets to 22% across enterprise and 11% across SMB.

Figure (3) Distribution of CSAT Ratings Across CSAT Survey Recipients



Unfortunately, in the context of CSAT, no news cannot be taken for granted as good news. Figure (4) shows that more than three-quarters of unrated enterprise interactions and more than half of SMB interactions contained risks.

Figure (4) Risk Prevalence Across Enterprise and SMB Interactions with No CSAT Rating



Let's take a look at some examples:

Table (6) Risks in Interactions with No CSAT Rating

Service-related Risks	Product-related Risks	Business-related Risks
"I don't have the time to wait around for them to figure out how to address the situation."	*"This issue is getting more serious since some scheduled tasks are failing."*	*"If additional fees of $7,200 are due for this, then a gross misrepresentation has occurred."*
"Please continue to try and resolve this issue as it's now gone over a week without resolution."	*"This is too an annoying behaviour that we don't know how to cope with."*	*"since it's unfair for us to pay more, you need to develop the algorithm to make up the cases."*
"You are obviously not qualified to handle this."	*"We need this as a matter of urgency – the raw site is performing poorly and we need to test on top of it to see if that stabilizes it."*	*"HATE your method of customer communication; hate it enough to be looking at other similar apps now."*

Relying on CSAT as a singular CX barometer can bring unfortunate consequences for both enterprise and SMB relationships. Table (6) shows that Support can miss valuable opportunities to accelerate resolution times and train agents to handle difficult inquiries, Product can stay oblivious to critical performance issues, and Operations is blind to pricing and communications frustrations.

Bain Capital's Delivery Gap study famously notes that while 80% of companies believe they are delivering a superior customer experience, only 8% of customers agree. Our study showed that only 15% of customers gave a bad CSAT rating — with this in mind, it's not surprising that over three-quarters of companies think their customer experience is excellent.

One reason for the stark disconnect is a lack of legible, normalized customer feedback that accurately represents the authentic customer experience. CSAT responses are typically too sparse to provide actionable insight, and surface-level takeaways can prove misleading to companies who incorrectly assume that their customer experience is outstanding.

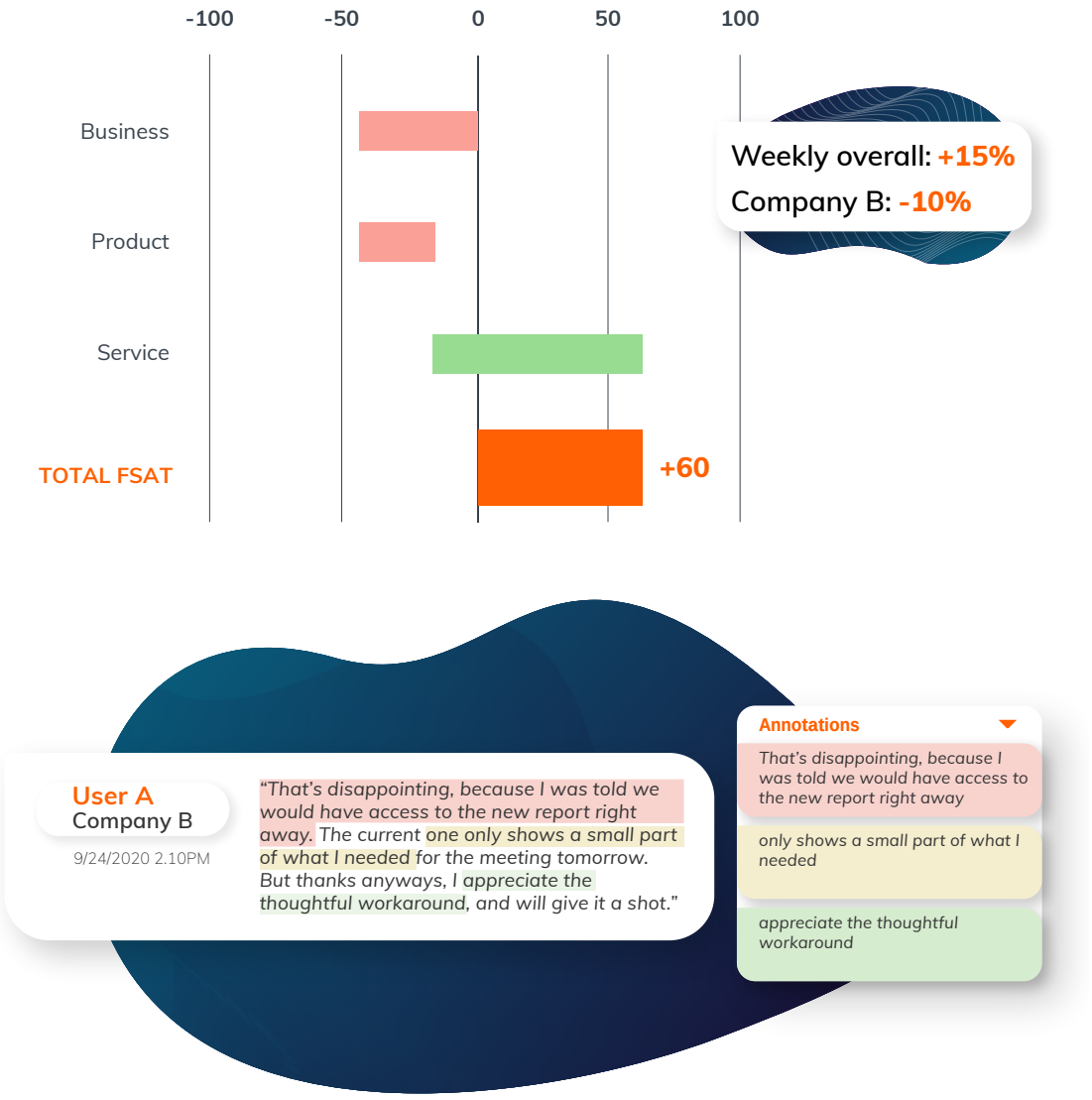
Support tickets are one of the best and highest volume representations of the customer voice — yet the vast majority of tickets go entirely unobserved once they have been closed by an agent. Many companies try to solve the data gap with tagging processes, though they often prove too rigid and fragile to capture meaningful insight. This approach has two critical flaws —first, tags cannot capture the unknown unknowns that represent some of the most impactful CX drivers. Secondly, the goals of the individual or team doing the tagging will always influence the tags that are applied.

CX leaders need a reliable metric based on normalized data that tells a rich, holistic cross-functional story — without one, they will fail to accurately measure customer experience and limit their capacity to improve.

Towards FSAT: Actionable Insight From 100% of Customer Interactions

In this paper, we have used Frame AI's sentiment module as a lens to illustrate some of CSAT's strengths and weaknesses. Wins, Issues, and Risks represent one component of Frame AI's approach to measuring satisfaction. The FSAT (Frame AI Satisfaction) metric is a satisfaction score passively applied to 100% of interactions in every channel that captures actionable insights by labeling not just positive and negative moments but specific service, product, and business factors embedded in each sentence.

Figure (5) Frame AI's Proprietary FSAT score



FSAT is one pillar of a well-executed Voice of Customer Ops strategy, an action-oriented approach that leverages the customer voice to drive continuous improvement across the customer experience. By covering 100% of interactions, FSAT delivers high-quality data that enables CX teams to make better decisions faster, with confidence. By attributing scores to specific parts of the customer experience, FSAT helps CX teams clarify priorities, reducing the time between information and insight and between insight and action. Most importantly, FSAT captures customer sentiment based on what customers took the time to put in their own words about issues they were motivated to express — bringing organizations' view of the customer experience closer to the customer's.

Conclusion

Organizations that intend to compete on customer experience cannot afford to operate without a robust measurement framework. CSAT surveys serve an important purpose, but they are insufficient as a standalone method of extracting actionable insights. The most compelling opportunity for CX organizations is to seize the wealth of insight already at their fingertips. Advances in Natural Language Understanding have made this information more accessible than ever, enabling next-generation measurement mechanisms like FSAT. In fact, in 2018, Accenture reported that 31% of organizations have already invested in artificial intelligence to outpace the competition.

Proper measurement multiplies the value of other CX investments by amplifying successes and targeting new efforts more effectively. With more clarity, depth, and automation across CX measurement, CX leaders can prioritize more effectively, act faster, and iterate efficiently to deliver industry-leading experiences.

About Frame AI

Frame AI is the engine that powers Voice of Customer Ops. We unify your customer voice from every channel and measure sentiment and effort on 100% of conversations, so you can activate more feedback when it matters. Learn how leading companies like Fastly, NVIDIA, and Oscar Health use Frame AI to get a complete picture of their customer voice for best-in-class customer experience at **www.frame.ai**.

Get in touch:

contact@frame.ai

blog.frame.ai

Frame AI

@frame_ai

frame.ai