servicenow

ServiceNow Inferred CSAT & Factors LLM

ServiceNow Inferred CSAT & Factors LLM

Intended Use and Functionality

Purpose of the Model

The Inferred Customer Satisfaction (CSAT) & Factors model is designed to ingest a conversation and predict a CSAT score as well as factors that explain the predicted score.

It provides insight into the user experience in virtual agent, AI agent conversations and human agent conversations beyond simple metrics like deflection rate or mean time to resolution (MTTR), aligning with the business intent of providing UX observability, driving operational excellence, and improving explainability for AI-based metrics.

Autonomy Level

The Model operates at an autonomous level and as part of a batch processing pipeline. The CSAT & factors rubrics have been predefined.

The model ingests a conversation and generates a JSON string containing CSAT & factors. Currently, no customization is available.

Optimization Scope and Limitations

The Model is optimized for conversations in ITSM, HR, and CSM, ensuring high performance in its intended use cases.

It has been fine-tuned on English data but is expected to be generally serviceable in ServiceNow P1 languages with some degradation in prediction quality.

Model Name

ServiceNow Inferred CSAT & Factors LLM

Model Version

v_0.2.0

Model Release Date (GPAI)

July 24, 2025

Model Distribution Method (GPAI)

The model is accessible using APIs to internal ServiceNow components; not available for public consumption or for direct consumption to ServiceNow customers.

Model License (GPAI)

Please refer to your agreement with ServiceNow for all license information.

Products Using this Model

Now Assist AI Agent and Virtual Agent

Now Assist feature availability by region

User Benefits

The Model delivers the following benefits for our users:

- Augment conversation insights with deeper AI based Inferred CSAT on 100% of conversations and underlying factors - resolution, effort, empathy, next steps, confusion, transfers and escalations.
- Pinpoint low CSAT interactions with both human agents and AI agents to adapt workflows and agent behavior.
- Provide close-to-real-time feedback on customer interactions
- Inform customer decision making and highlight areas for improvement by delivering automated conversational insights across AI Agent / AI engagement layer analytics.

Risks

- LLMs, in particular smaller LLMs (7B - 12B), are susceptible to prompt injection attacks. Customers are not going to interact with our model directly, so the risk of deliberate prompt injection attacks is low. We have mitigated the risk further through reformatting the prompt in our instruction fine-tuning. But such a risk cannot be entirely removed.

- Similarly, the model can be influenced by the presence of CSAT surveys or other types of input noise, for instance, in a speech-to-text transcript, resulting in potentially inaccurate predictions. It is recommended that the CSAT surveys be removed from the main conversations. ServiceNow is aware of the issue and continues to reduce the impact through preprocessing and model retraining.

Factors and limitations

The Model is designed to perform CSAT & factors prediction effectively but has specific factors and limitations that influence its performance and applicability. These include:

Input Requirements:

- The Model relies on **a specific conversational input format** to generate accurate outputs. In particular, the conversation needs to have clearly designated speaker roles - AGENT, USER, and BOT. Without clear speaker roles, the model may predict inaccurate results. The Model does not currently handle multi-agent input and only outputs one CSAT score with factors for the entire conversational input.
- Existing CSAT surveys with user-provided scores should be preprocessed out of the conversational input. If a CSAT survey with a user score is part of the conversational input, the Model is likely to capture the score given by the user rather than predicting CSAT based on the main conversation.

Domain-Specific Challenges:

- The Model has been fine-tuned on conversational data in the ITSM, CSM (retail), and HR verticals, and attempting to use it in other domains or other data formats without additional fine-tuning may result in **decreased accuracy.**

Language-Specific Challenges:

- The Model has been fine-tuned on English data only, although the base model has been pretrained on multilingual data. Our multilingual eval has been limited to translated datasets with a delta in accuracy up to -8 points for Japanese (75% EN vs. 67% JA).

Ethical considerations

The Model runs in the background and the user does not interact directly with the model but only sees the model outputs in a dashboard. Hence, it is unlikely to produce harmful content in terms of biased, toxic, or hallucinated texts.

Supported Languages

Primary Language:

- English

Multilingual Capabilities:

- The model will generate an output in the proper format in other languages but will result in accuracy hits.

Model Architecture

Base model: mistralai/Mistral-7B-Instruct-v0.2

Fine-Tuning: LoRA fine-tuning with cross-entropy loss on next token prediction

Number of Parameters

How many parameters does the model have? (GPAI): 7B

Maximum Input and Output Size

- **Maximum Input Size:** 32k tokens (capped at 6k in the triton service)
- **Shared Context Window:** 32k

Input and Output Modalities

The Inferred Customer Satisfaction (CSAT) & Factors model ("the Model") is designed to handle the following modalities of inputs and outputs:

Modality Type

Single-modality: The Model processes and generates text to text.

Inputs

Input Type: The Model accepts plain text input formatted with the chat template that comes with the Model tokenizer.

Processing Method: Inputs are processed through a custom preprocessor in the batch processing pipeline. **If querying the model directly, the conversational input has to be formatted with a predefined template consisting of a system prompt and rubrics.**

Input and Output Modalities

Input Constraints:

- The Model expects clearly defined speaker roles (BOT, AGENT, USER) in the conversation and may perform sub-optimally with no speaker tags or different speaker tags like names.
- The Model expects each speaker turn of the conversation separated by double newlines. Failure to follow the formatting may result in small variation in predicted scores & factors.

Outputs

Output Type: The Model generates **JSON formatted strings.**
Output Characteristics: Outputs are **JSON strings - a dictionary with CSAT & factors as keys and predictions as values.**
Output Constraints: The Model is optimized to produce **JSON formatted strings. Occasionally, it may generate misformatted JSON (< 1%).**

Input and Output Formats

The Model requires specific formats for its inputs and outputs to ensure seamless interaction and reliable performance.

Input Format

Data Structures:

- Inputs must be formatted with the predefined input template and then by the model tokenizer with the predefined chat template.

Encoding Types:

- Inputs must adhere to UTF-8 encoding standards for accurate processing.

Context Handling:

- The input template contains a system prompt and the predefined rubrics for CSAT & factors

Output Format

Data Structures:

- Outputs are generated in **JSON schema.**

Encoding Types:

- Outputs follow UTF-8 encoding standards for compatibility.

Templates or Schema:

- Outputs include fields like resolution, confusion, frustration, transfers and escalations, effort score, empathy, next steps, and csat.

Training Data

ServiceNow did not conduct further pre-training.

Fine-Tuning Data

- The Model has undergone instruction fine-tuning for the CSAT prediction task using proprietary datasets consisting of anonymized conversations in the ITSM, CSM, and HR domains.
- There is a small set of synthetic conversations.
- Data anonymization was conducted before use.
- Labeling was provided by our in-house expert and team, as well as the rubrics.

Evaluation Data

- The model has been evaluated by a proprietary dataset (separate from train) in the same domains.
- The multilingual capability has been evaluated by translating the eval data from English to ServiceNow P1 languages while preserving the labels.

Metrics

Evaluation Metrics

What evaluation metrics were used?

- Key Evaluation Metrics: Micro and Macro F1 for factors, Micro, Macro F1, and MSE for CSAT.