

# Mitigating Privacy Risks When Analyzing Conversational Chat Data

With Privitar Data Privacy for Chat.

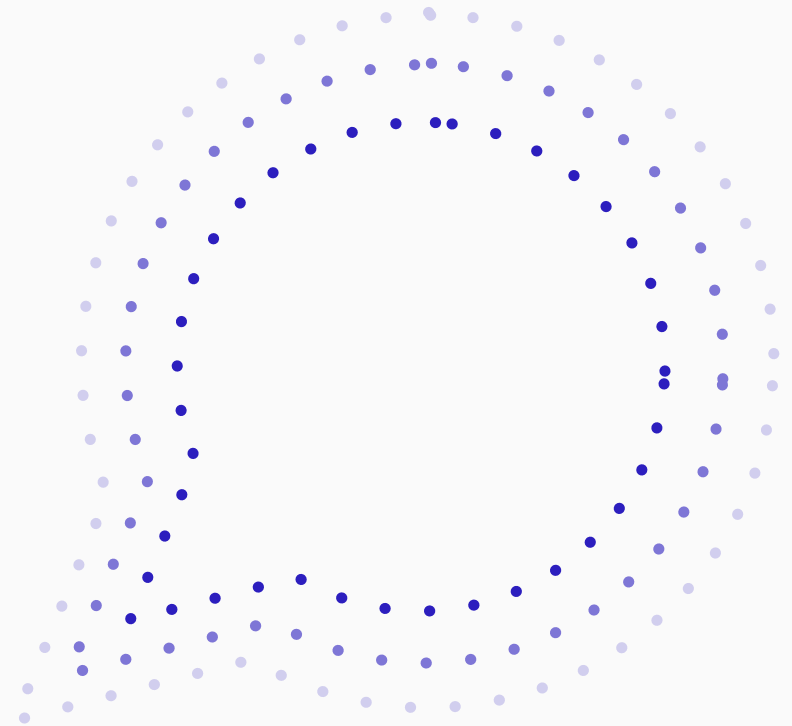


## The Untapped Value of Conversational Chat Data

Customer interactions on communication channels like social media, SMS messaging, and chatbots generate large amounts of valuable conversational text, also known as *utterances*. This data contains valuable information about product or service issues, sentiment, feedback, and more. The data can also contain personally identifiable information that is subject to regulatory control. Therefore, access to this conversational text for analysis is often limited or completely restricted in its raw form.

**“ Compliant analysis of customer conversational text is the most pressing thing for my team just because we work so much with free text data, and we haven’t even started using all of it. ”**

Group Data Science Manager for Operations and Servicing Analytics at a Global Retail Bank





## Uses for Chat Text in Data Initiatives

De-identifying this sensitive data allows you to provision it to a wider group of analysts while upholding customer privacy and ensuring regulatory compliance. Responsible analysis using this valuable unstructured (free text) data source starts with protecting sensitive data that can reveal an identity—such as name, address or location, organization, phone number, email address, credit card number, date of birth, and so on. This safe data can now create opportunities in use cases like:

- Analyzing customer service effectiveness and efficiency
- Determining new product and service features
- Gauging customer, citizen, supplier, and employee sentiment

## Privitar Data Privacy for Chat

Using natural language processing techniques, Privitar Data Privacy for Chat recognizes and classifies text, then applies policies to de-identify sensitive data. De-identification permits analysis without exposing sensitive, protected information.

A policy is a collection of rules that determine:

- How sensitive data is masked
- How sensitive data can be used
- Who can see and use the data in raw form or in masked form

### Benefits:

- Derive value of customer interactions for analysis

---

- Reduce regulatory risk by de-identifying sensitive data attributes

---

- Deploy into existing conversational channels and workflows with minimal disruption of operations

---

- Improve the performance and utility of common NLP analytics on conversational data

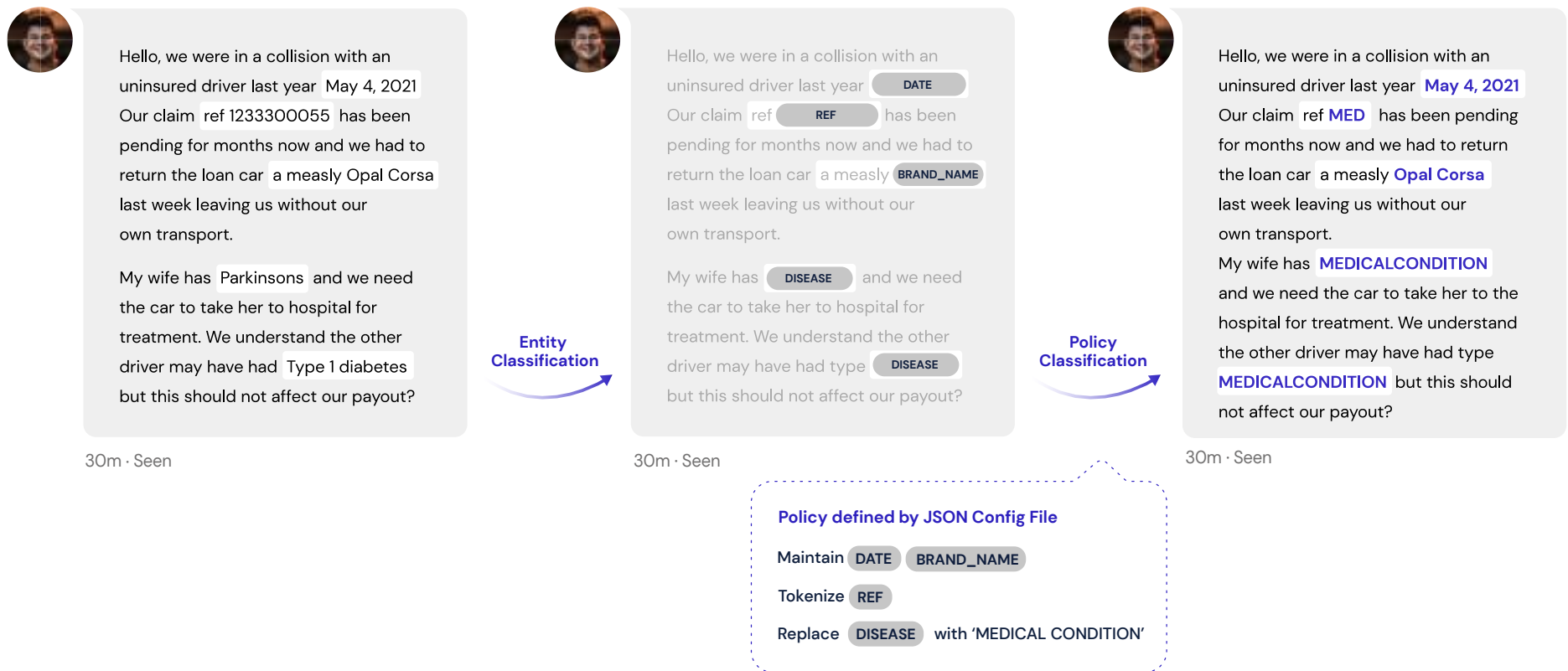
---



# How Does it Work?

Privitar combines a set of de-identification transformations with deep learning models for natural language processing (NLP) to allow users to apply privacy policies on English-language<sup>1</sup> conversational text data.

Figure 1. Before and after applying a privacy policy to sample conversational chat data.



1. Dutch support expected in 2H2022.



## Classifying sensitive text data with natural language processing

The NLP model classifies free-form text, character by character.<sup>2</sup> The Privitar transformation engine uses the classifications to apply a privacy policy. A policy consists of rules determining how to transform the text, allowing for consistent tokenization of protected information.

Privitar trains the NLP model using a proprietary active learning process that enables higher accuracy than competing offerings from cloud vendors. You can tailor the model to your organization's data.

**“ Unstructured text is a great thing—the PoC was incredibly useful. Analyses performed better after de-id. ”**

Systems Architect at a Global Insurance Company

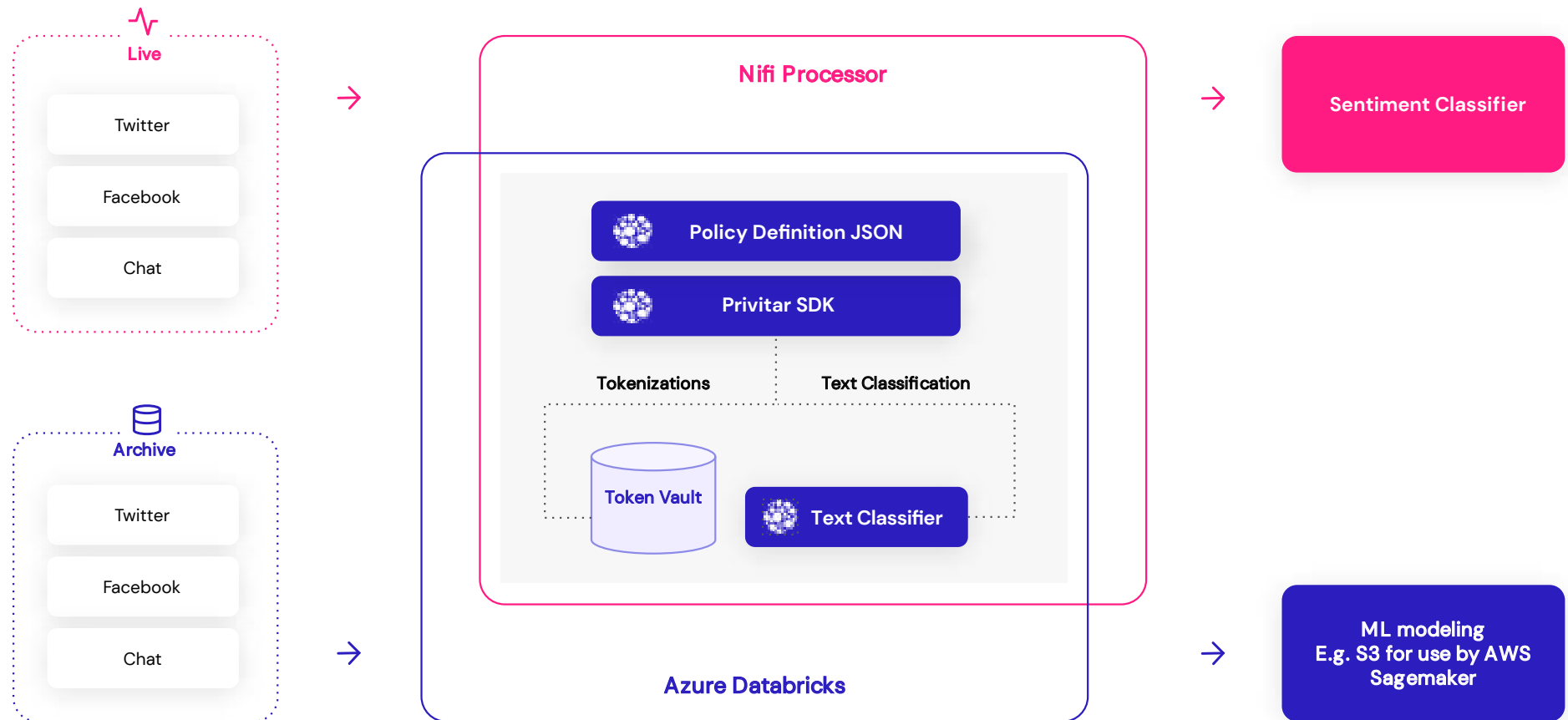
2. Recommended configuration is 480 characters per message.





## Integrating with your Existing System Landscape

You can integrate Data Privacy for Chat into your existing data pipelines like Databricks, NiFi®, Kafka®, and many more. Kubernetes® deployment of the NLP model allows you to scale demanding text classification workloads. By deploying the model to CUDA® accelerated hardware you can support real-time de-identification.





## Using batch mode to de-identify large volumes of unstructured text data

Some use cases require that you transfer large volumes of unstructured text messages—perhaps from an archive—across a network boundary to a different part of your organization to be used for a different purpose. To reduce the privacy risk in the data in this scenario, Data Privacy for Chat supports batch mode de-identification use cases such as:

- Developing a chatbot to automate parts of your customer service infrastructure
- Analyzing customer, supplier, or employee complaints so that you can prevent churn of your high-value clients
- Building a predictive model to help your procurement system get ahead of new customer trends

## Technology best practices for batch mode

You may have, in total, hundreds of millions of characters of text sitting in your archive of chat messages. The best way to process this data is to use the parallel processing power of an Apache Hadoop® cluster. Data Privacy for Chat can be seamlessly integrated into the reducer of a Databricks Spark™ application. You can deploy a cluster of Privitar text classifiers to Kubernetes for the duration of the job and scale the Databricks job as needed to process the archive of chat messages. Dividing the archive into batches allows you to process the archive during quiet times on your cluster.

## Using streaming mode for production analytics pipelines

The NLP model classifies free-form text, character by character.<sup>2</sup> The Privitar transformation engine uses the classifications to apply a privacy policy. A policy consists of rules determining how to transform the text, allowing for consistent tokenization of protected information.

## How do I sign up or get more information?

Visit [www.privitar.com/DP4C](http://www.privitar.com/DP4C)

## About Privitar

Privitar empowers organizations to use their data safely and ethically. Our modern data provisioning solution builds collaborative workflows and policy-based data protection into data operations. Only Privitar has the right combination of technology, domain expertise, and best practices to support data-driven innovation while navigating regulations and protecting customer trust.

For more information, visit [www.privitar.com](http://www.privitar.com).